立 心

——华中师范大学心理学院成立10年论文选辑

◇ 名誉主编　刘华山

◇ 主　　编　周宗奎

◇ 副 主 编　郭永玉　江光荣　马红宇

华中师范大学心理学院

SCHOOL OF PSYCHOLOGY，CENTRAL CHINA NORMAL UNIVERSITY

华中师范大学心理学科的历史可以追溯到 20 世纪 30 年代。1931 年华中大学教育学院设立教育心理系、并开设系列心理学课程，同时中华大学教育系也开出多门心理学课程。1934 年美国哥伦比亚大学心理学博士胡毅受聘华中大学教授并创办心理学系，第二年芝加哥大学心理学博士骆传芳也来此任教。1953 年调整后的华中师范学院教育系继续开设心理学课程，期间著名心理学家朱希亮长期在此任教。1978 年改革开放以来，我校心理学事业持续发展。周镐、王启康等老一辈心理学家为恢复心理学的教学和科研工作做出了重要贡献。1984 年在本科教育学专业分出心理学方向，招收第一届心理学硕士研究生，并成立"华中师范大学心理咨询保健所"。1985 年恢复招收心理学专业本科生。1993 年学校组建教育科学学院，下设心理学系，刘华山教授任系主任。2003 年获得发展与教育心理学博士学位授予权。2005 年，学校成立心理学院，从此进入跨越式发展时期。

心理学院教职工（2015 年 11 月）

立　心

——华中师范大学心理学院成立 10 年论文选辑

名誉主编：刘华山

主　　编：周宗奎

副 主 编：郭永玉　江光荣　马红宇

编　　委（按姓氏笔画排序）：

马红宇　刘华山　江光荣　佐　斌

周治金　周宗奎　范翠英　郭永玉　龚少英

目　录

脑与认知

人格与社会心理

发展心理

心理咨询与健康

管理心理

序

　　十年光阴之于天地宇宙，转瞬即逝，不足一瞥；十年时间之于一所大学，以动辄百年计的校史烟云，亦不过弹指一挥，匆匆而过；而十年历史之于一个学院，则似乎可轻可重：其轻者，若学院书脉兴旺，学子繁盛，源远流长，则十年亦可谓白驹过隙，轻忽飘渺；其重者，譬如十年树木，塑其形，成其荫，十年足矣。

　　适逢心理学院立院十年，同事诸君，劳心费神，精选专文，终成此辑。文集选题广泛，恰如心理学大树之分枝繁多。似乎天马行空，各行其是，而审视渊源，则各项研究自有缘起，或横向联展，或纵向相依。细细读来，虽风格各异，方法有别，然而篇篇凝聚心血，探寻真知灼见，实乃文同此境，贯穿始终。求真务实之心路，历历在目。

　　文章千古事，得失寸心知。虽不免敝帚自珍，实向往珠玑之制。

　　北宋大家张载名言："为天地立心，为生民立命，为往圣继绝学，为万世开太平"。千余年来，多少名儒学子奉此为圭臬。吾辈可谓身居太平之世，治学历年，钻研心理，更当引此为鉴，磨砺心志，治顶天立地之学，求为天地立心之境！

　　是为序。

周宗奎

公元二〇一五年十一月三十日

于武昌桂子山

教育心理研究所团队成员

刘华山，教授，主要研究领域为学习与教学心理、学校心理辅导、中小学生心理健康教育、网络学习。

龚少英，博士，教授，教育心理研究所所长，主要研究领域为学习与教学心理、网络学习、教师心理。

张微：博士，副教授，主要研究领域为特殊儿童学习与促进、学校心理辅导、网络心理健康。

王福兴：博士，副教授，主要研究领域为网络学习、教学心理、媒体与儿童、眼动技术。

熊俊梅：博士，讲师，主要研究领域为学习与教学心理、网络学习、心理健康。

唐云：博士，讲师，主要研究领域为心理统计与测量、网络学习。

彭明：博士，讲师，主要研究领域为社会认知，包括情绪识别和调节的神经机制，进化角度下情绪的功能与发展，情绪对社会认知的影响，网络情境中的社会认知特点。

教育心理学：沟通心理学与教育的桥梁

刘华山，龚少英，熊俊梅

（华中师范大学心理学院）

摘 要 教育心理学是研究人的心理与行为改变规律的科学，是心理科学的一个重要应用分支，学校情境中学与教的基本心理学规律是其研究的主要对象。教育心理学是心理学与教育长期结合的产物，当其作为一门独立学科在 20 世纪初诞生后，曾给当时教育研究走向科学化以强有力的推动。近一二十年来教育心理学研究在学习与认知，动机、情绪与信念，教学与学科学习，学习的个体差异等主题上有了较大进展。研究发展趋势表现在：一般认知过程的研究转变为课堂里具体学科学习中的认知研究；教学心理研究受到重视；研究视角多元化；开始从对个体的关注转到对社会文化情境中个体的关注；计算机与网络等技术因素对学习的影响成为一个新的研究领域。最后，作者从发挥教育心理学作为沟通心理学与教育桥梁的功能出发，对我国教育心理学发展提出了若干建议。

关键词 教育心理学；发展脉络；研究进展；发展趋势

1 教育心理学的诞生

教育心理学是研究人的心理与行为改变规律的科学。人的心理与行为的改变，或曰人的学习，可以在自然和社会环境影响下自发地产生，也可以在学校教育影响下有目的、有计划地进行。由于后一种学习形式的重要性，当今大多数教育心理学家都认同把教育心理学定义为"研究学校情境中学与教的基本心理学规律的科学"。

以冯特 1879 年在莱比锡建立世界上第一个心理学实验室为标志，心理学从哲学母体中脱离出来，实现了从哲学心理学向科学心理学的转变；稍后，教育心理学也从教育哲学中脱离出来，成为一门独立的学科，其标志是 1903 年美国心理学家桑代克（E.L.Thorndike）《教育心理学》一书的问世。

从起源上说，教育心理学是心理学与教育结合的产物。心理学与教育的结合并非始自科学心理学诞生之后，而是经历了哲学取向下的结合与科学取向下的结合两个阶段的漫长历史过程。因此，虽然从学科体系来说，教育心理学是心理科学的一个分支学科，但这并不意味着教育心理学是从普通心理学中分化出来，它也不单纯是科学心理学的衍生物。

直到实验心理学成为一门独立的学科之后，欧洲的教育家和心理学家才开始运用实验、测量与统计的方法研究儿童心理发展及教育问题。其中为科学心理学与教育的结合作出早期努力的是冯特的学生莫伊曼（E.Meumann）。莫伊曼十分推崇实验研究对教育工作的重要性，他与教育家拉伊（A.Lay）所倡导的实验教育学运动，以及重视对儿童身心发展与改进教育方法进行实验研究的思想，深深地打动欧美许多教育家和心理学家。而在科学心理学与教育结合的历程中迈出决定性一步，并对教育心理学

的创建作出突出贡献的则是美国心理学家桑代克。桑代克在19世纪末开始用实验、测量的方法研究学习及个别差异问题，于1903年出版了《教育心理学》。他所提出的"教育心理学的研究是以了解人性及改变人性而实现教育目的"的观点为本学科的性质及其与教育的关系做出了明确的定位。

桑代克《教育心理学》一书的问世，不仅使教育心理学走向科学化，而且也带动整个的教育学研究走向科学化。他认为教育将"依赖其领导者用科研结果而不是一般意见指导其方法选择的程度而得到改进"。许多过去视为不证自明的传统教育观点，现在都要经过实证研究的检验而决定存废。例如：

教育心理学对学习迁移现象（学习中的"举一反三"）有过各种理论解释。其中一种早期的迁移理论"形式训练说"认为：注意力、记忆力、思考力、意志力等，都是每个人的"心"所具有的官能（能力），它们可以通过训练而得到加强。学习的迁移则是在一个领域通过训练而得到加强的心理官能，在其它领域自动地发挥作用。特定的学科对于训练人的某种心理官能可能具有独特的作用。早已退出生活领域的拉丁语，据说就是因其有利于训练学生的观察、推理、记忆等官能，故仍值得学生花费大量的时间去学习。按照这一观点，学生所学知识、技能的内容并不重要，重要的是它能否使学生的思考力、记忆力、想象力得到训练。这种观点在欧美流行了200多年，对我国教育亦产生过一定的影响，可以说是长期误导了学校教育实践。正值教育心理学诞生之际，桑代克和伍德沃斯（R. S. Woodworth）于1901年以大学生为被试，在知觉领域进行了一系列实验，结果表明：形式训练说的许多假设都没有经得起科学检验的证据。该研究的结论影响深远。对于当今我国教育界一些人在倡导"素质教育"时忽视知识技能掌握的观点也有警示作用。它警示人们，轻视学科知识、技能、认知策略的学习和迁移，而一味地追求素质、能力的普遍提高，只能是南辕北辙。

2　教育心理学与教育关系发展的历史脉络

教育心理学诞生后，科学心理学与教育实践的关系，走过了百年曲折的发展道路。梅耶（R.E.Mayer）将这种关系的发展比喻为三条道路：单向道，死胡同，双向道。

2.1　单向道时期

20世纪初至30年代一段时期，心理学家对科学心理学改进学校教学的作用普遍持有乐观态度。在这一精神鼓舞下，美国教育学界推动了一场以桑代克为主导的教育科学运动，开展了4项大型的教育心理学研究，包括贾德主持的儿童阅读心理研究；桑代克主持的智力测量研究；推孟主持的天才儿童研究；全国教育研究会负责的先天遗传与后天教养问题的研究。这场以教育心理学为主导的教育科学运动有力地促进了当时教育心理学的发展。这一时期的教育心理学家对教育现实问题都表示极大的关注；注重用科学精神和严密的科学方法指导教育问题的解决。但对教育问题的症结缺乏深层次的把握。

2.2　死胡同时期

由于前一时期教育科学运动的成就没有达到预期的理想，20世纪30-60年代一段时间内，许多教育心理学家开始脱离学校教育实际，回到他们擅长的实验室工作，热心于根据动物学习实验中得到的资料去建立各种庞大的学习理论体系。行为主义学习理论在当时占有主导地位。由于这些理论都不是以学校情境中的学习活动为基础，所以很难在解决学校教育问题上发挥作用。其中行为主义者斯金纳（B.F.Skinner)在其强化原理基础上倡导的程序教学，由于符合学习的部分规律，曾在许多国家引起反响，并对教学技术现代化产生过积极影响。

2.3　双向道时期

20世纪60-90年代，心理学与教育的关系进入双向道时期。心理学在适应教育实践需要方面取得了进展，同时教育也成了推动心理学发展的动力。

由于 60 年代初美苏国防竞赛的压力，美国人普遍增强了视教育为国防的观念，通过对教育的反省，认识到加强中小学知识教学，提高国民知识水平的重要性，在教育上掀起了"恢复基础运动"。教育心理学研究也开始由行为主义范式转向认知范式，心理学家从人为控制实验室重新回到面向教育实际的研究，学科学习心理和认知研究受到关注。布鲁纳以其结构主义学习观为依据所倡导的课程改革运动，对许多国家教育改革产生了强大的推动作用。奥苏伯尔对有意义言语学习的过程、条件、心理机制的研究，加涅对人类学习的分类及内外条件的研究，维特洛克（M.C.Wittrock）通过对阅读教学和自然科学教学的考察而开展的对生成过程（意义建构过程）的研究，以及信息加工理论的许多代表人物对学习中知识表征和内部加工过程的研究，在认知领域的学习和教学规律探讨方面都取得了可观的进展。

3 教育心理学的内容框架、研究方法与应用价值

3.1 教育心理学的内容框架

教育心理学研究学校教育情境中学生"学"的心理规律，同时也研究旨在有效地指导这种学习的教师的"教"的心理规律，而以学生学习的基本心理学规律为其研究主线。其基本内容框架包括：

学生学习的性质、特点和分类。研究涉及学习的实质，内部结构，学习中人的行为改变的心理机制，学习、记忆与脑的关系，各类学习（机械学习与有意义学习，陈述性知识学习与程序性知识学习，外显学习与内隐学习等）独特的过程和特点。

学生学习的过程。揭示学习的一般过程，也揭示各类学习的特殊过程。例如认知心理学家将阅读理解分为解码过程、字面理解过程、推理性理解过程和理解监控过程，使我们得以了解阅读过程中所发生的心理事实，从而为分析学生阅读能力差异提供了框架，也为对阅读障碍儿童实施有效干预指引了思路。

影响学生学习的因素。包括对影响学生学习的个体因素和外部因素及其复杂的交互作用的考察。在内部因素方面最为受到关注的是学生的认知结构特征（背景知识）和动机情感因素。外部因素则包括课堂里的社会心理因素、学校人际关系、家庭变量、社会文化背景、计算机与网络的技术环境因素等。

基于科学心理学的教学设计研究。本领域研究旨在将对学生学习的心理规律的了解转化为合理的教学原则、教学组织和教学设计。

3.2 教育心理学研究方法的特殊性

心理学研究使用的一般方法都被大量地用于教育心理学研究。计算机模拟、反应时实验、出声思考、作业展示以及眼动技术和某些认知神经科学的方法在对学生阅读、解题过程的研究中也都有广泛应用。例如：

四则运算是学生必须掌握的自动化技能，布朗和范莱恩（Brown & Vanlehn）提出了一种减法能力模型，据此编制了减法运算的计算机程序，并假定学生减法错误是由于使用了错误规则（称为"程序障碍"）。对所编制的程序作出各种改变，就可以模拟学生所犯的各类错误。利用这种模拟的程序可以对学生的错误进行分析，获得诊断信息，选择适合学生需要的补救教学措施。

由于教育心理学研究对象和研究目的的特殊性，在研究方法选用时，也有一些独特的问题需要考虑：

（1）强调真实教学环境中的研究。教育心理学研究特别注重实验室研究与现实课堂研究的结合、量的研究与质的研究的结合。在开展面向教育实际的问题研究时，重视自然实验法、各种准实验设计以及改进的观察法、深度访谈、学生作品分析等质性研究方法的运用。这也符合 20 世纪 80 年代以后出现的儿童与教育心理学研究中的"生态化运动"的基本趋势。

（2）注重包含多变量的综合研究。由于影响学生学习的因素众多，各种生理的、认知的、情感的、社会的因素交互作用，故在研究中不宜总是简化变量及变量间的关系，而是需要更多地采用多因素设计、更有弹性的理论模型和处理数据的多元统计方法，

提高研究的内部和外部效度。

3.3　教育心理学的应用价值

教育心理学主要服务于学校教育实际，宏观层面上，它能为课程改革、教学内容、方法的改革提供理论支持。20世纪中期世界上影响较大的教育改革运动，如美国布鲁纳（J.S.Bruner）倡导的课程改革运动，苏联赞可夫（Л.В.Занков）主持的小学教育体制改革，都是受到教育心理学理论的推动而兴起的；在微观层面上它能为解决学校教育与课堂教学中的实际问题、为改善有特殊需要儿童的学习提供建议。

除此以外，教育心理学的基本原理和研究成果，特别是作为其核心部分的学习心理学，因其本身就带有基础性质，故在其他有关领域亦具有应用价值。例如，行为疗法的理论基础，就是学习联结理论中关于两种条件反射的形成、消退、强化、惩罚、接近学习、交互抑制以及观察学习、生物反馈等一系列原理；教育心理学的知识体系也部分适用于成人教育、员工培训、罪犯改造；美国心理学会成立学校心理学家分会时曾将学校心理学定义为应用临床与教育心理学（1945），反映了教育心理学在学校心理服务中的应用价值；班杜拉（A.Bandura）的观察学习论主要用来解释人的社会行为的习得过程，对于社会公众的态度改变和社会文明建设具有参考价值。

4　教育心理学的研究进展

NolenAL对美国心理学与教育学杂志中2007年影响因子居高端的6种权威期刊（《教育心理学杂志》、《教育心理学家》、《学习科学杂志》、《学习和个体差异》、《教育心理学评论》和《当代教育心理学》）进行分析。这些期刊2003年~2007年间总共发表了758篇论文。采用SPSS TAS组织和分析数据，从这些论文的主题词中形成25个类属。各类属在758篇论文中出现的比率反映了其内容的相对重要程度。结果显示，排在前5位的是：36%的论文（n=279）聚焦于课堂成就，33.2%的论文（n=251）

关注学习与记忆。情绪/动机/信念、认知/推理，教学的主题分别占总论文的31%，21%和21%（一篇论文可聚焦于多个类属）。这说明近期教育心理学研究的内容重心集中在课堂成就、认知/推理、学习/记忆、情绪/动机/信念，以及教学几个方面。

Mitchell和McConnell Ⅲ对1995-2010年在《当代教育心理学》杂志上发表的440篇专业论文进行了内容分析，发现教育心理学研究的几大主题是：学科、认知过程、个体差异、方法论及专业思考、教与学。以下仅就若干研究主题，对教育心理学近一、二十年来的研究进展做些说明。

4.1　认知与学习

涉及的主题有学习过程中的认知负荷、注意力、理解、记忆、推理、元认知和迁移；学生的各类知识（观念性理解、自动化技能和认知策略）在新知识获得和问题解决中的作用；阅读、数学学习、记忆、问题解决过程中的学习与大脑的关系等。不少研究探查了"基于计算机的协作和合作学习"这种新的学习形式的性质和条件，以及多媒体、超媒体和网络条件下的学习者的自我效能感、自我调节等问题。

4.2　学习动机与学业情绪

学习动机是研究得最多的主题之一，学习者个体认知因素，个体信念，如自我效能感、成就目标定向、学业成败的归因、自我价值等对动机的作用在近期研究中受到重视。研究内容有成就目标定向的发展、升学对目标定向的影响以及课堂目标结构对个体目标的影响、自我效能感与考试焦虑。

学业情绪是指发生于学习过程中与学生学业相关的情绪体验，如自豪，满足，焦虑、内疚，羞愧，无助和厌倦。德国心理学家Pekrun等按唤醒度和愉悦度的高低将学业情绪分为四类。并基于社会认知的视角，提出了社会认知控制—价值学业情绪理论（Social Cognitive Control-Value Theory of Achievement Emotions），引发了关于学业情绪的一系列实证研究。该理论认为，学业情绪受学习者个人的控制感（自我概念、自我效能感和归因）和成就价值的影响，也受父母期望、教学质量、同伴关系等外部环境的

影响。

4.3 学科学习与教学

认知心理学为当代的教学研究提供了深厚的基础。研究的问题有：课堂环境不同方面，如班级目标结构、性别比例、教师特征、师生关系等对学生学习有何影响；计算机和网络条件下的学习对于不同知识经验、认识能力、学习动机的学生有何积极作用和消极作用；什么样的多媒体和网络学习的内容和形式能产生最佳效果。一些研究探查了基于计算机的合作学习以及网络学习对学生知识结构、问题解决的影响。此外，具体科目（阅读、写作、数学、科学等）的认知过程和个别差异的研究也占有很高的数量比例。

5 教育心理学的研究的发展趋势

5.1 研究内容的变化与拓展

近20多年来教育心理学研究内容显现出一些新特点，实验室中人为控制下对认知过程的研究转变为课堂里具体学科学习中的认知研究；教学心理研究受到重视，力图将"描述性"的学习理论与"处方式"的教学理论结合起来；教师的作用和训练的研究受到关注，通过新手和专家教师的比较研究，试图揭示专家型教师成长的途径；从20世纪90年代以后开始运用认知神经科学技术来探讨阅读、学习效率与脑部活动的关系，预计将会对深化教育心理学若干主题的研究（如能力的实质、复杂的学习过程、教学活动）产生积极影响。

5.2 研究视角的多元化

当代教育心理学研究主要采用认知、社会认知、元认知、信息处理、建构主义和行为主义6种视角。其中认知和社会认知视角采用得最为广泛。认知视角下的研究取得了令人瞩目的成就，对信息加工过程的精细描述，促进了学习和思维策略的研究以及认知策略教学的发展，多项研究显示了学习和思维策略教学成功地改善了学生在各科学习领域的表现。采用社会认知视角的研究者倾向于将个体特征与环境因素整合到一起考察。随着建构主义学习理论的兴起，教育心理学家开始从对个体的关注转到对社会文化情境中个体的关注，越来越倾向于探索基于真实生活和具体情境中的个体学习与合作学习，关注社会的、人际的和文化的环境对学习者的信念、态度和认知的影响。

5.3 环境变量和技术因素的影响

Berliner（2006）认为教与学通常是在"教师 × 学生 × 任务 × 情境"的交汇处发生。而近20年来情境的变化最为迅速。随着计算机辅助教学、多媒体教学、网上课程、视频会议实现的合作学习的日益增多，探讨环境如何激发学习者的认知和动机投入，网络课程设计如何符合学生的认知特点，避免学生认知负荷过重和信息迷航，使其成为有效地自主调节学习者，就成为必须研究的问题。专家认为，计算机、网络等技术作为一种环境因素，在学习中的作用应成为"真正重要的问题之一"而列入21世纪教育心理学的研究日程。

6 发展我国教育心理学的建议

6.1 加强教育心理学研究与教育实践的联系

科学心理学独立以后，教育心理学承担着沟通心理学与教育理论、教育实践的桥梁作用。如何在坚持心理学科的方法规范和学术专长的同时关注教育现实问题的解决，则是教育心理学永远无法绕开的话题。如果在解决我国教育实际问题时，教育心理学研究严重缺位，教育者单凭权威人物意志、对西方某种流行理论的迷信、经验常识、理论思辨等去制定教育政策和教育措施，必定会使有关的教学理论和教育改革实践失却科学基础和实证依据。1996年台湾教育当局为改进小学数学教育，在缺乏充分准备的情况下全面推行建构式教学。实施6年造成200万小学生数学能力普遍降低，最后在社会抨击声中于2002年全面喊停，即是一个深刻的教训。反之，如果像20世纪20~50年代的大多数教育心理学家那样坚持方法中心主义，一味地热心于在实验室情境中

建立精密的学习理论体系，而置开发学生智力和创造力、完善学生人格、解决教育实际问题的基本目标于不顾，教育心理学必定会走入发展的"死胡同"，重蹈历史覆辙，丢失本学科的应用价值。

6.2 扩充教育心理学研究的视野

"活到老，学到老"是中国的古训，也是符合"终身学习"观点的先进理念。建设学习型组织，学习型社会已是建设中国和谐社会的重要内容。教育心理学主要研究中小学教育情境中学与教的心理学规律。在坚守这一重要研究领域的同时，可以适当将研究的视野扩展到成人，探讨各类组织中的员工学习、领导和管理人员学习以及教师培训与专业成长的心理学规律，以及高层次专业人才、创新人才成长的心理规律。在所要考察的学习的形式上，除了学校课堂学习外，社区的学习、基于博物馆、科技场馆的学习的特有规律也应该纳入教育心理学的研究范围。

6.3 提高人才培养质量

参照 Berliner 的有关观点，我们对我国教育心理学专业研究生培养提出如下建议：（1）使学生注重方法论思考和实证研究方法训练，能将量化研究和质性研究相结合，能运用多因素研究设计和多变量统计技术；（2）对我国教育现实问题有较深入的了解；（3）须从事为期一年的教育实践；（4）教师为研究生提供复杂环境下进行学术研究的实践机会；（5）确保学生对教育政策有相当程度的了解。

参考文献

胡锦涛 . (2007). *高举中国特色社会主义伟大旗帜，为夺取全面建设小康社会新胜利而奋斗——在中国共产党第十七次全国代表大会上的报告，见《十七大报告》辅导读本*，人民出版社，36–37.

胡谊 . (2007). 改良教育心理学：来自认知神经科学的影响，*心理学探新，27*（1），15–18.

李季湄 . (2001). 教育心理学的发展历程综述 – 梅耶的四隐喻说，*心理科学，24*（4），454–457.

林崇德 . (2005). 试论发展心理学与教育心理学研究中的十大关系，*心理发展与教育*，（1）：1–5.

皮连生 . (2004). *教育心理学*，第三版，上海教育出版社，10–13.

吴庆麟 . (2000). *认知教学心理学*，上海科技出版社，305–309.

张春兴 . (2005). 从思想演变看教育心理学发展宜采的取向，*北京大学教育评论，3*（1），77–93.

Azevdo. R Using hypermedia as a metacognitive tool for enhancing student learning? The role of self-regulated learning. *Educational Psychologist, 40*, 199–209.

Berliner, D. C. (2006). Educational psychology: searching for essence throughout a century of influence. In Alexander，P A, Winne, P H (Eds.), Handbook of educational psychology (2nd ed). *Mahwah: Erlbaum,* 1–27.

Berliner, D. P. (2003). *Toward a future as rich as our past*. Carnegie Initiative on the Doctorate, Carnegie Foundation for the Advancement of Teaching, Stanford CA.

Calfee, R. (2006). Educational psychology in the 21st century. In Alexander，P A, Winne, P H (Eds.), Handbook of educational psychology (2nd ed.). *Mahwah: Erlbaum,* 29–42.

Elliott，A. J. (2006). *A conceptual history of the achievement goal structure*. In Elliott，A J, Dweck, C S (Eds.), Handbook of Competence and Motivation. New York: Guilford Press.

Elliott, A. J., Murayama, K., Pekrun, R. A. (2011). 3 × 2 Achievement Goal Model. *Journal of Educational Psychology,* 103(3), 632–648.

Mandl, H., Ertl, B., Kopp, B. (2006). Computer support for collaborative learning environments. In Verschaffel, L, Dochy, F (Eds), Instructional psychology: Past, present and future trends. *Netherlands: Elsevier,* 223–237.

Miller, G.，Reynolds, W. (2003). *Future Perspectives in Educational Psychology*. IN Reynold W M，Miller G E (Eds), Handbook of Psychology V7: Educational

Psychology. New Jersey: John Wiley & Sons, 609–628.

Mitchell, A. W., McConnell III, J. R. (2012). A historical review of Contemporary Educational Psychology from 1995 to 2010. *Contemporary Educational Psychology*. doi: 10.1016/j.cedpsych.2011.11.001.

Moos, D. C., Azevedo, R. (2009). Learning with computer–based learning environments: A literature review of computer self–efficacy. *Review of Educational Research, 79*, 2, 576–600.

Nolen, A. L. (2009). The Content of Educational Psychology: an Analysis of Top Ranked Journals from 2003 Through 2007. *Educ Psychol Rev, 21*, 279–289.

Pekrun, R., Goetz, T., Titz, W., et al. (2002). Academic emotions in students' self–regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37,* 91–106.

Pekrun, R. (2006). The control–value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*, 315–341.

Simons, R., deLaat, M. (2006). E–pedagogies of networked learning. In Verschaffel, L, Dochy, F (Eds), Instructional psychology: Past, present and future trends. *Netherlands: Elsevier*, 239–255.

Educational psychology: A bridge between psychology and education

LIU Huashan; GONG Shaoying; XIONG Junmei

（School of Psychology, Central China Normal University 430079 Wuhan）

Abstract　Educational Psychology is an important applied discipline of psychology science investigating the psychology of man and laws for behavior modification. The primary laws for learning and instruction in school context are the foci of Educational Psychology. Educational Psychology has been the product of psychology and education. When it was founded as an independent discipline at the beginning of the last century, it enormously promoted the scientific study of education. In the last twenty years, there has been great advances for Educational Psychology in learning and cognition, motivation, emotions and belief, instruction and subject learning, and individual differences in learning, etc. The trends of research development are: The study of general cognitive process has shifted to the cognitive study of concrete subjects in classroom settings; The psychology of teaching has been more emphasized; Multiple perspectives for research are preferred; Educational psychologists have begun to pay attention to persons in social cultural context instead of only paying attention to individual characteristics; The impact of computer and internet technology on learning has been transformed into a new research field. Lastly, the authors put forward some suggestions as on how to further nurture the development of Educational Psychology, with its function as the communicative platform for psychology and education.

Keywords　educational psychology; developmental thread; research advancement; trends of development

国外心理健康服务体系目标概览及启示

刘华山

摘 要 在查阅有关若干国家心理健康服务目标的文献资料的基础上，本文重点介绍了美、英、澳制定的国家或地区心理健康服务目标的内容梗概，并从心理健康服务目标的内容构成，所遵循的基本原则和理念、服务运行与管理机制三方面，引申出一些重要启示，以便为我国心理健康服务目标体系的构建提供借鉴。

关键词 心理健康服务；目标

完善的心理健康服务体系是现代社会的重要特征，是我国和谐社会构建的重要内容。而心理健康服务体系目标是心理健康服务体系建设的预期结果，它要指明经过一定时段的建设后，我国心理健康服务体系所要达到的状态、所要具备的特征或所能完成的功能。心理健康服务体系目标的建构可为我国心理健康服务体系的建立和完善提供指引，也可为心理健康服务体系的管理和评估提供标准。

要构建这一目标体系，最基本的是要考察我国心理健康服务工作现有基础、我国公众对心理健康服务的需求，以及我国经济和社会发展状况为心理健康服务事业所能提供的人力物力条件；同时借鉴国外的有关经验也是非常必要的。本课题组采用文献检索法，查阅了澳大利亚、美国、英国、加拿大、加沙地区、丹麦、白俄罗斯、泰国、德国、新加坡等国有关心理健康服务目标状况的资料 76 份（其中，外文资料46 份，中文资料30 份），旨在了解国外心理健康服务的现状、心理健康服务目标的内涵、目标制定所遵循的理念和原则，从中寻找必要的借鉴和启示。

1 国外心理健康服务目标撮要

1.1 美国纽约州心理健康服务目标

纽约州心理健康主管部门在 2005 年 11 月制定

的"纽约人心理健康保护系统战略计划框架"从结果、心理健康服务和系统运作三方面阐述了州心理健康服务的目标。该目标体系共有 8 个支撑性目标（supporting goals），42 项具体目标。

目标 1：促进公共心理健康。通过一种有效的针对公众和服务提供者的教育，改善公共心理健康水平，幸福感水平以及所有纽约人的"韧性 / 挫折容忍力"。包括 5 项具体目标，内容涉及：使公众更多地了解"自杀"现象，提高公众对各种危险以及防范措施的意识；面对各种自然和人为灾害时，能够保证迅速有效地为大众提供心理健康支持；使公众能更好地了解儿童的各种情绪障碍和成人的各种心理疾病的成因、危害和治疗措施；完善对饮食障碍的心理学测查、早期干预和治疗；协调从事基础护理的医生、其他健康护理服务者和社区服务者的工作，以完善各种甄别、早期干预和预防。

目标 2：促进儿童、家庭和成人的积极发展。通过行之有效的治疗措施，改善有严重情绪障碍的儿童和成人的发展状况。包含 3 个具体目标，内容涉及：对已证明行之有效的心理健康服务做更普遍的推广；提高消费者和家庭对治疗过程的投入和参与程度；对有潜在危险的群体提供更多服务，以帮助个体在生活、学习、工作和社会环境中达成满意和成功。

目标 3：实践研究。加强从事基础研究、临床研

究和服务等科学团体和消费者团体间的联系以减少各种疾病负担。包括 6 个具体目标，内容涉及：增进对心理疾病成因和治疗措施的了解；促进新的治疗措施的研发；增强行之有效的服务措施对于不同文化和语言环境的适用性；缩短研究成果运用于主要社会受益者的预备时间；促使研究者为服务者和决策者提供更好更多的技术支持。

目标 4：持续地提高服务质量。提高针对严重情绪障碍的儿童和严重心理障碍的成人的心理健康服务的质量，包含 6 个具体目标，内容包括：确保以人为中心的护理原则的贯彻，综合例行护理中的各种被证实为行之有效的常规方法，来提高服务质量；尽可能减少因各种伤害、忽视或不当的护理、治疗而造成的不良后果；确保州政府和各城镇具备评定、监督护理质量的各种条件、资源和行政能力。

目标 5：服务途径。使公众能够更便捷地得到适当有效的服务，特别是针对弱势群体和服务缺失的群体的服务，包含 9 个具体目标，内容包括：确保有严重情绪障碍的儿童、面临发展危机的儿童、涉及司法刑罚的个体、刚成年的个体、居住于养老院且有心理障碍的个体、能够获得适当有效的心理健康服务的途径。

目标 6：服务系统的能力。提高州政府和地方政府达成行政目标的能力，包含 6 个具体目标，内容涉及：完善州和地方服务系统提供适当有效服务的能力；完善系统对不同文化和不同语言的群体提供服务的能力；完善人员招募和聘用；完善系统对工作人员的技能和能力的培训；发展和优化服务系统的评定和监督服务的成本和效益的能力。

目标 7：政府的职责。各级政府部门要承担起对服务的获得、服务的质量和有效性以及服务的费用等方面的责任，包含 4 个具体目标，内容涉及：完善州和地方心理健康服务方案和监督制度，以提高其责任感；完善对成人和儿童的医疗措施的监管；完善服务提供者的认证过程和执照发放过程；消除不同文化、年龄和性别群体在服务获得和服务质量上存在的差距。

目标 8：合作护理。加强针对个体需求和意愿的医疗、自助、社会支持等服务环节的统筹部署。

包含 3 个具体目标，内容涉及：与其它州级儿童服务机构合作，确保能为有严重情绪障碍的儿童提供全面有效的服务；协调为那些已获得各种层次护理的个体所提供的服务；完善对有多次住院记录但少有适当院外服务的个体提供的服务。

1.2 澳大利亚国家心理健康服务目标

澳大利亚国家心理健康政策目标共 38 项，要点如下：

1.2.1 消费者权利。

目标 1~2 涉及消费者权利，内容包括：规定消费者享有"澳大利亚卫生部权利和责任书"、联合国有关"保护有心理疾病个体决议"以及心理健康立法"提高心理健康保健法"中规定的所有权利；确保自治政府和所有其他州政府有保护以上权利的机制。

1.2.2 心理健康服务和整个健康部门间的关系。

目标 3~6 涉及心理健康服务部门与整个健康部门的联系，内容包括：扩大综合医院看护急性精神科住院病人的范围；在各直辖地区维持或建立一个整合的心理健康方案；将有关整体健康服务的主流组织，如地区的管理系统等整合到心理健康服务中来；推介那些可以提高保健的持续性和为有心理健康问题的人们提供综合服务的系统。

1.2.3 心理健康服务和其他部门联系。

目标 7~9 涉及心理健康服务系统与系统外社会各部门的联系，内容包括：在健康部门内外的方案和服务中都要消除对有严重心理健康问题和心理障碍个体的任何外显或内隐的歧视；整合当地政府和各地区间的联系和服务安排，确保有严重心理问题的人获得的服务能够反映他们的需求。

1.2.4 服务的协调综合。

目标 10~15 涉及心理健康服务系统内部的协调综合，内容涉及：各州就各种服务的综合制定专门计划；降低现有精神病院的规模，同时另外提供足够的急诊医院和基于社区的服务；升级现存的精神医疗设备；分散精神病院提供的服务，确保所有地

区都能获得一般的住院病人服务和社区服务，包括危机服务、测评和治疗；分析州政府对心理健康资金投入的空缺及可能造成的损失，并予以解决。

1.2.5 促进和预防。

目标16~18涉及心理障碍预防方面的目标，内容包括：制定有关心理障碍知识的公众教育计划；鼓励对心理障碍起因的深入研究，鼓励对初级预防干预的发展和评估等。

1.2.6 初级保健服务。

目标18~20内容为：确保针对初级保健工作的教育计划对心理健康问题有适当的涵盖；确保初级保健工作者，尤其是农村和偏远地区的初级保健工作者能获得专业的心理健康资源。

1.2.7 保健工作者和非政府组织。

目标21~22内容为：支持非政府组织的发展壮大，使其通过提供信息、参与心理健康服务决策和资金投入，为保健工作者提供协助。

1.2.8 心理健康人力资源。

目标23~26涉及心理健康服务人员的供给和配置，内容包括：确保毕业生的数量，确保他们所具备的相关护理技能能够符合专业心理健康服务的要求；在各州、各地区，在公共的、私人开业、半专业的机构中对精神病理学工作者进行更合理配置；鼓励对所有心理健康专业人员进行继续教育。

1.2.9 立法。

目标27~28涉及立法方面的目标要求，内容包括：确保澳大利亚境内的心理健康立法一致，并确保这一立法关注目标1提到的消费者应享有的所有权利；确保其他部门的立法和国家心理健康政策中的原则相一致。

1.2.10 研究和评估。

目标29~31内容为：扩大心理健康基础和应用研究，促进研究成果在预防和干预方案中的应用；对于针对有严重心理障碍的人的服务结果进行定期审查，将其作为实施心理健康服务的一个核心成分。

1.2.11 标准。

目标32~36是为心理健康服务达到合格标准设

立的，包括：鼓励设立心理健康服务结果的国家标准体系；确保所有的心理健康服务都有质量保证；每个心理健康机构都被独立的、受到认可的

评审机构认定完全合格。

1.2.12 监督和问责。

目标37~38的内容为依照本政策目标和相关法规目标，开展国家认可的行为评估，及时根据评估结果编制公开的年度报告。

1.3 英国威斯敏斯特心理健康服务目标

英国威斯敏斯特心理健康服务2004年部分目标要点有：

1. 提高咨询水平以确保服务接受者和其他相关人员的观点能够整合到服务的发展和监督中来。

2. 按照国家目标，减少自杀事故。

3. 为遭受严重心理疾病困扰的个体开展有效服务。

4. 改善针对因精神（心理）障碍而犯罪者的服务。

5. 实施针对老年人的心理健康国家服务框架。

6. 为无家可归者提供一定的住房和心理健康服务。

7. 针对有心理疾病个体开展就业培训。

8. 对保健人员给予认定和支持。

1.4 基于学校的心理健康服务目标：预防、早期干预和治疗

美国基于学校的心理健康服务的目标要点如下：

目标1：提升公众对儿童心理健康问题的知晓度，减少心理疾病引起的耻辱感。内容包括：提高儿童在社会、情绪、行为三个方面的幸福感；宣传包括心理健康内容在内的促进儿童发展的指导纲要；识别心理健康问题的早期征候；提升教育系统提供足够预防性心理健康服务的能力。

目标2：增强对儿童心理健康需求的评估和识别。内容包括：鼓励对现行教育、青少年司法、物质滥用治疗体系中心理健康需求的早期鉴别；和家长、看护者、儿童讨论心理健康问题，为进一步的评估和干预提供恰当的参考；培训教育人员使其能识别有特殊健康保健需要的儿童、破碎家庭儿童以及父

母有着心理健康障碍或物质滥用障碍的儿童的心理问题的早期征兆；在学校层面上，发展积极有效的行为支持系统；认识到未加治疗的心理健康问题有造成儿童和青少年犯罪的潜在危险。

目标3：缩小心理健康服务获得方面的种族、民族和社会经济差异。内容包括：增加那些有文化针对性的、经过科学证实且适合青少年及其家庭能力和需求的服务；积极发展和整合那些要求家庭参与预防和干预措施的活动、方法；支持针对有犯罪倾向青少年心理健康问题的心理健康计划；加强针对不同种族、民族、性别、性取向和社会经济地位的团体的心理健康问题的诊断、预防、治疗和服务。

目标4：增加获得心理健康服务的可能，加强服务的协调。内容包括：采用一套通用语言来描述儿童的心理健康状况，这对于促进跨系统服务非常重要；在治疗中对儿童进行追踪，并评估个案的治疗结果；修正教育系统采用的定义和评价步骤，以识别有心理健康需求的儿童和青少年并为他们服务；注意保密，以尊重家庭的权利和隐私，但鼓励不同系统服务者间的相互协调与合作；将青少年容纳进治疗方案中，以适当的方式向他们提供有关服务的直接信息，允许他们对推荐的干预措施进行决策和选择。

2 有关心理健康服务目标的一些重要理念的启示

国外心理健康服务目标体系中包含的若干理念，体现了这一领域的重要价值和哲理，对于我国制定心理健康服务目标体系颇多有益启示。

2.1 广义、积极的心理健康概念。

通过考察了美国哈佛大学、麻省理工学院、斯坦福大学和耶鲁大学四所美国一流大学的心理健康教育状况，人们发现这些学校的心理健康工作不仅涉及心理健康，也包括身体健康、生活态度的健康和环境的健康。例如哈佛大学心理健康服务中心认为，好的健康状态是保证生命质量的重要因素，而一个好的健康生活方式应包括健康的身体、健康的精神和健康的环境。

注重积极目标。一些发达国家的心理健康服务目标中除了强调为一些重点人群、为遭受严重心理疾病困扰的个体开展有效服务、减少自杀事故一类的"消极目标"外，还开始关注"提高健康生活的质量和年限"这样的积极目标。如纽约州心理健康服的第1项目标就是"提高公共心理健康水平，幸福感水平以及所有纽约人的韧性/挫折的复原力"（resiliency）。有的国家目标中还将对精神病的治疗与支持个体为实现健康生活发展各种能力和职业技能结合起来。

2.2 心理健康服务是一种公共事业。

心理健康服务不能仅仅依靠家庭、朋友和同伴的私人间帮助、社会团体帮助来维持，也不仅仅是一种志愿者的行动，它是由在政府掌控下的公共事业机构中工作的从业人员完成的、旨在满足社会公众需要的一种有组织的行动。私人、宗教团体、慈善机构及其他社会团体、都可以赞助公共事业。而各级政府在制定计划和法规；完善心理健康服务专业人员的招募、聘用、培训；完善资格认定和执照授予过程；设立心理健康服务国家标准，对服务系统进行评估、监督和问责；加强对各部门、各地区、各心理健康服务机构力量的协调合作；解决对心理健康服务的资金投入的空缺部分；在面临重大自然灾害和事故时，组织迅速有效的心理健康支持；减少各类人群在获得心理健康服务的途径和质量上社会差距等方面，都起着不可替代的重要作用。

2.3 尊重服务对象的个人权利、需要和意愿。

在心理健康服务中要保护消费者权益，注意消除对有心理障碍者的公开的或隐性的歧视。例如加拿大不列颠哥伦比亚省在心理健康服务执行指导中提出的多项指标，主要体现了对服务对象的需要和意愿的充分尊重：（1）可接受性。提高消费者对所获得心理健康服务的满意度；注意服务的文化敏感性。（2）可获得性。使患者能够在恰当的地点和时间获得所需的服务和照料。（3）适当性。所提供的服务和看护与患者的需求有关，符合既定标准。（4）能力。确保提供服务的人员有足够的相关专业知识和能力。

（5） 延续性。使消费者能得到不间断的跨方案、跨从业者、跨组织、跨服务层次和跨时间的协同服务等。

2.4 确保心理健康服务的社会公平。

美国、加拿大、澳大利亚等国的心理健康服务目标中都列入了消除不同文化、年龄、性别、种族、民族和社会经地位的群体在服务获得和服务质量上存在的差距方面的内容。特别强调要为那些弱势群体、没有参与保险的儿童、涉及司法刑罚的个体、农村和边远地区的社区提供便捷的、有效的心理健康服务。美政府曾先后制定了"健康人 1990"、"健康人 2000"、"健康人 2010"系列方案。其中，"健康人 2000"设定了三大健康目标，分别是增加全美人民的健康生活年限；缩小全美人民的健康差距；争取所有人民都能获得预防性服务。"健康人 2010"有两大总体目标：提高健康生活的质量和年限，即帮助各年龄段的个体提高生活期待、改善生活质量；消除不同人群间的健康差距。

2.5 强调心理问题的预防和早期干预。

美国、澳大利亚、新加坡等国的心理健康服务都非常重视提升公众对儿童心理健康的知晓度，增加公众获得预防性健康服务的途径，明确心理健康问题的早期指标，强调心理问题早期发现、早期干预。与此有关的是，重视由公共卫生机构、社区诊所或个人开业者提供的初级卫生保健（primary health care）的作用，为初级保健工作者提供专业支持，确保他们能获得专业的心理健康服务资源。

2.6 充分发挥社区服务的作用。

加拿大、美国、澳大利亚在其国家心理健康政策目标中都涉及到：降低现有精神病院的规模，分散精神病院提供的服务，扩大综合医院看护急性精神科住院病人的范围，另外提供足够的急诊医院和基于社区门诊服务。社区服务有许多独特的功能，如对基层心理健康服务有广泛的覆盖；帮助严重心理疾病者在最少限制的环境中尽可能的达到最佳功能水平；相对于住院治疗是一种兼顾成本和有效性的选择；产生较高的来访者及其家属的满意度。

2.7 鼓励服务对象及其家庭的积极参与。

将青少年吸纳进治疗方案中，提高消费者和家庭对治疗过程的投入和参与程度，允许他们对推荐的干预措施进行决策和选择。积极发展和整合那些要求家庭参与预防和干预措施的活动、方法，如男孩女孩俱乐部咨询、患者自助服务、家庭自助服务。

2.8 提倡基于证据的实践（Evidence–based Practice）。

要求特定的心理服务实践、步骤、计划通过特定的研究标准和证据标准的系统检验，并证实能产生积极的结果。符合标准的服务实践分为两类：完善的治疗，可能有效的干预。达到"完善的治疗"标准要求有 2 个或 2 个以上实验研究来证明他们比药物、安慰剂或其他替代性的治疗高出一筹；"可能有效的干预"标准则要求有一个符合"完善治疗"的实验研究或 3 个个案研究的支撑。提倡基于证据的实践，实质是要求通过服务实践、效果评价、科学研究之间的密切结合，来提高心理健康服务的专业性和科学性。

2.9 整合多种资源，提升合作协调能力。

美国德克萨斯州、加拿大近年来都把推进协调服务，建立一个和谐有效的服务和支持系统作为心理健康改革的中心目标。这包括各服务环节（医疗、自助、住宿、职业技能训练等）的配合；服务者、患者、家庭、学校观点的整合；针对同一些对象的不间断的跨方案、跨从业者、跨组织、跨服务层次和跨时间的协同服务；心理健康服务与其他健康部门之间的合作；心理健康服务部门与政府及部门外机构的配合；心理健康服务目标与一个地区的其他社会发展目标的协调一致。

2.10 心理健康服务目标的一致性和特殊性。

许多国家如美国、澳大利亚、新加坡都有全国整合的心理健康方案、心理健康服务的目标，确保境内心理健康立法的一致，且要确保国家其他部门的立法和国家心理健康政策中的原则相一致。但在承认这种一致性的前提下，也要求各州、各地区确立适合自身特点的心理健康服务目标。同时还可针

对特定人群制定心理健康服务目标，如美国卫生局 2001 年正式出台《儿童心理健康国家行动议程》中设置的目标、美国基于学校的心理健康服务目标、英国针对老年人的心理健康国家服务框架、白俄罗斯军队心理卫生工作的目标等。

3　心理健康服务运行与管理机制的启示

欧美许多国家在心理健康服务途径、运行机制和服务管理方面也有许多创造性的观点和做法，值得在制定我国心理健康服务目标时效法和借鉴。

3.1　大众健康框架下的学校心理健康服务

大众健康关注群体的健康，并且考虑影响人的健康的社会因素，如生活方式、社会经济地位和预防教育等。大众健康模式将预防分为三级：普遍性预防、选择性预防和指示性预防。

大众健康模式用于学校心理健康服务时，教育与心理健康从概念上被视为不可分割的整体，心理健康服务被看做学生学习的先决条件和保证。在这一模式下，研究与服务工作都要作出相应的调整。服务人员就理论和知识基础来说，很有必要深入地理解组织心理学和系统理论；从方法上讲，需要加强系统水平上的咨询技巧和能力；研究方向应集中于问题解决而不仅仅是理论问题的探讨，例如在学校暴力预防研究中，需要查明那些预防行动或方法在经验上是有效的，以及这些方法如何运用到学校情境中去。在对心理服务效果的评价上，应重点考查群体的变化，关注整个学校中的积极或消极结果的比率增高或降低情况。

3.2　美国和德国的自助服务

传统的与新产生的心理服务途径多种多样。如社区的临床服务（或者说是门诊），半住院服务，住院治疗，居家治疗，基于家庭的服务，治疗寄养服务，案例管理服务，健康图书馆，自助团体等。

其中自助团体在许多发达国家近一、二十年有了快速发展。在美国，这种团体已经从 1963 年的 300个发展到 1992 年的 500000 个，拥有超过一亿五千多万名成员，成为正式心理咨询与治疗机构的有效附属品。自助团体基本上是由有相似问题的人们所组成的互相帮助的群体，作为自助团体解决的焦点问题包括酗酒、滥用毒品、强迫性赌博、丧亲、过食、恐怖症、被强奸受害、失业以及生理疾病。自助团体有两大优势，一是更加便宜，二是由一些有类似问题的人提供支持，它比专业的领导团体更加容易亲近和接受。

老年孤独是德国人晚年幸福生活的最大障碍之一，克服这一障碍的有效方法是成立老年人自助团体，推进老年人的互助和助人服务。活动形式多样："上门服务"内容包括静心聆听其他老年人倾诉苦闷，帮助视力衰退者阅读书籍或读写信件、陪伴他人散步、办手续以及出行参加各种活动。"绿衣女士"（或"绿衣男士"）是活跃于德国医院中的身着绿色大褂的男女老人志愿者。他们义务帮助患者寻找病房或诊室，或用书籍或游戏解除候医与住院病人的烦恼，或是伴随病人克服生理和心理上的困难。老年人互助与助人活动不但为需要者解决了现实生活困难，而且使许多助人者为自己确立了新的生活目标、结识了新的朋友、开拓了新的兴趣天地，使老年人通过对社会作贡献而赢得社会尊重，强化了老人间的情感联系和互助精神。

3.3　协调服务机构者和消费者的管理治疗计划

管理治疗是由保险公司提供的，与医院、患者以及私人服务提供者挂钩的卫生保健系统的保护伞。据估计，管理治疗组织目前在美国覆盖了 70%~80% 的团体健康保险。它是涉及整个健康保健系统的管理模式，对于我国心理健康服务管理也有参考价值。

在提供主要的服务或在交付治疗费用之前，管理治疗保险公司必须利用外部评价机制对治疗进程作同行评估，以便在所推荐的治疗中确定一种能符合患者需求的最经济、最有效的治疗方式。评估需要回答的问题有：当事人的问题是什么？建议的治疗方式是什么？为什么推荐这样的治疗？有什么证据可以支持这种治疗？这种推荐的治疗在费用上划

算吗？以下几种管理治疗计划比较有用。

健康维护组织（HMO） 是最有组织性并能提供最全面的卫生服务的管理治疗计划，采用固定的提前付费方式。HMO 组织有他们自己的诊所，或有专的专家组或医生系统。治疗在主要治疗医生与登记员的协调下进行。其主要的优点是费用较低，中小企业将其视作一个让员工享受医疗保障的最便宜的途径。而缺点在于患者只能局限于 HMO 的诊所和医师。

选择服务者组织（PPO）：是医生和医院都同意以固定进度提供卫生服务的组织。他们给预期的患者一份服务者名单，患者可以直接找他们中的任何一个。

指向服务的计划（POS）：也有自己的服务者系统，但是患者可以同时在系统之外寻求治疗，只要他们愿意担负一些额外的费用。

管理治疗计划的优点是考虑到消费者的意愿，提供了更好的服务机会和新的转诊资源；使服务对象有选择医院、医生、服务方式的自由；在一定程度上可以控制医疗费用的上涨。但它也受到一些人的抨击，他们指责保险公司提供了装备线式的服务，破坏了医患关系；确立了不同的付费机制却没有建好的服务机制。

4 心理健康服务目标内容构成的启示

心理健康服务目标到底应包含哪些内容？各国有关心理健康服务的政策、计划、目标要点等都为我们考虑这一问题提供了参照和启示。前述"纽约人心理健康保护系统战略计划框架"是从结果、心理健康服务过程和系统运作三个方面来阐述了州心理健康服务的目标的。第一方面涉及促进公共健康和促进个体积极发展两个内容；第二方面涉及如何保证心理健康服务质量的途径和措施；第三方面涉及政府的职责和各种力量的协调。澳大利亚国家心理健康目标包含消费者权利，心理健康服务的主要功能（促进、预防、初级保健等）和实现保障（人力资源，科研，评估），政府的职责（制定标准、立法、监督问责），

心理健康服务的协调整合等。

吸收若干国家心理健康服务目标中的合理部分，结合我国及某些地方现有有关心理健康服务的政策法规精神，现提出我国心理健康服务目标内容的基本构想。设想的我国心理健康服务目标体系可归入服务结果的目标、服务资源建设目标、服务运行体系目标和服务保障目标 4 个目标群，具体包含以下 10 个方面：

4.1 心理健康服务的结果性目标。

这是心理健康服务所要达到的终极目标，即心理健康服务在预防心理疾病、促进精神疾患康复、提高国民心理素质方面预期带来的积极变化。如提高国民幸福感，改善有严重情绪障碍的儿童和成人的发展状况，降低自杀率等。

4.2 心理健康服务机构的建设目标。

这是心理健康服务机构建设方面应达到的目标，指各类心理健康服务机构的性质、数量、布局、覆盖区域、场地面积、设施、资料、经费保障及其应达到的要求或标准。

4.3 心理健康服务人员资格认证、配置、培养目标。

这是心理健康服务队伍建设上应达到的目标，服务人员的类别、层次、职能、专业资格、编制、人口配比、培养途径进修提高方式、职务晋升渠道等方面的要求。

4.4 心理健康服务科学研究方面的目标。

科学研究是提高心理健康服务实效性的专业保障。本部分目标涉及有关本领域基础研究与应用研究的人员、经费、与服务实践结合的研究取向、开展国际国内学术交流等方面的要求。

4.5 心理健康服务运行体系方面的目标。

这是心理健康服务的过程性目标。包括心理健康服务对象（服务对象的普遍性，各类重点人群的健康服务，服务对象对服务的可获得性、多样性、延续性、经济性、社会公平性的评价），心理健康服务运行模式，心理健康服务主要功能（教育、评估、预防、咨询、心理治疗、危机干预、护理、康复、

重大自然灾害与人为事故下的心理救助），心理健康服务途径，心理健康服务各部门、各机构的协调合作方面应达到的要求和标准。

4.6 心理健康服务的监督、评估和问责方面的目标。

本目标要求政府或主管部门完成其对心理健康服务过程和结果的监督、评估、问责方面的职能，并建立起相应的制度、方案和队伍。

4.7 心理健康服务整合与协调方面的目标。

心理健康服务的协调涉及心理健康服务内部的协调、心理健康与与整个卫生部门的协调、心理健服务的内外协调。本目标要求明确协调的各方面主体在心理健康服务中的领导隶属关系、职责分割，建立相对稳定的协调机制。

4.8 制度、法律、政策、法规建设目标。

应由政府或主管部门制定有关心理健康服务全国性或地域性基本制度、法律、政策；重要社会生活领域（医护、学校、军队、社区等）心理健康服务的法律、法规；针对特殊人群（残障人、精神病、物质成瘾者、涉及司法刑的个体等）的心理健康服务的法律、法规；有关心理健康服务人员的资格、培训标准与职业服务规范；新闻媒体对自杀事件报道的职业规范；与心理健康服务有关的社会控制（如毒品、农药管理）方面的制度和规定。

4.9 政府职能实现方面的目标。

这是政府及其职能部门在对所属地区的心理健康服务的领导、规划、组织、制定标准、监督评估、提供保障等方面预期达到的目标。

4.10 心理健康服务信息管理的目标。

这是有关国民心理健康及其服务状况信息的收集、分析、报告、更新、管理、运用方面的目标。信息的收集和分析可以为心理健康服务的科学研究和领导决策提供可靠依据。

上述 10 项目标中，第 1 项为结果性目标；第 2、3、4 项为条件性目标（资源建设目标）；第 5、6、7 为服务运行体系方面的目标；第 8、9、10 为保障性目标。我们试图以这一目标内容框架为出发点，通过调查

分析，修改补充，确立具体指标和标准，构建符合中国国情的心理健康服务目标体系。

参考文献

李明忠 . (2006). 美国一流大学心理健康教育工作的特色分析 [J]. *比较教育研究*，（1）.

赵俊茹 . (2003). 外国学校心理健康研究新进展 [J]. *天津教科院学报*，（5）.

George, E. Pataki., Sharon, E., Carpinello, Barbar. a L., Cohn, Keith. Simons. *Strategic Plan Framework*，2009– 1– 24 取自 http://www.omh.state.ny.us/omhweb/statewideplan/2006/html/appendix1.html

Summary of National Mental Health Policy objectives. 2009– 01– 24 取自 www.healthyactive.gov.au/internet/main/publishing.nsf/Content/B95EAF9352E4FBD4CA25722200141705/$File/repap15.pdf

Mental Health Services in Westminster (WestminsterPrimary Care Trust). 2009–01– 24 取 自 pct.nhs.uk/pdfs/ boardpec/wpct_board28sept04MENTAL_HEALTH_SERVICES_IN_WESTMINSTER.pdf

School–Based Mental Health Services [Prevention, Early Intervention and Treatment for Alcohol and Other Drug Abuse(AODA)]. 2009– 01– 24 取自 http://www.aasd.k12.wi.us/ACA/mental%20health.htm

Kimberley, McEwan & Elliot M. Goldner. Accountability and Performance Indicators for Mental Health Services and Supports.2009– 01– 24 取 自 www.phac– aspc.gc.ca/mh– sm/pdf/apimhss.pdf

William, R., Burger. Human Services in Contemporary America（7th ed），Thomson Brooks/Cole,2007.1

作者单位：华中师范大学心理学院

道德推理与道德行为关系的元分析 *

吴 鹏 [1]，刘华山 [2,3]

([1] 湖北大学教育学院心理学系，武汉，430062)

([2] 华中师范大学心理学院)

([3] 青少年网络心理与行为教育部重点实验室，武汉，430079)

摘 要 目前道德心理学中存在对道德推理作用的质疑，这一质疑源自哲学领域的著名争论。从经典道德心理学理论来说，道德推理应该是道德行为的重要决定因素，但新近的观点则否定这一重要作用。本研究采用元分析技术探讨道德推理与道德行为的关系。通过文献搜索与检查，获得了 50 项研究和 83 个独立效应量，共包含 16738 名被试。检验表明发表偏差不会影响元分析的结果，选择随机效应模型是准确的。通过随机效应模型的元分析表明，道德推理与道德的行为有显著的正相关，与不道德的行为有显著的负相关。调节效应分析表明，道德推理测量工具的类型会影响道德推理与道德行为的关系，被试年龄阶段会影响道德推理与不道德行为的关系。这些结果肯定了道德推理的作用，也强调了研究过程中要关注测量工具的类型，指出了需要开发更全面的道德推理测量工具。

关键词 道德推理；道德行为；元分析；调节效应

分类号 B849:C91

1 问题提出

在哲学领域，存在一个千百年来都未得到解决的问题：到底是情绪还是认知决定着人的道德？古今中外，有很多学者对这个问题发表过观点。在众多观点中，以休谟 (David Hume) 和康德 (Immanuel Kant) 两派的思想最具代表性。休谟认为情感驱动我们的道德判断，理性可以对道德判断发挥作用，但它必须依靠情感。他认为在人类的道德中，非理性因素是最为重要的。与之相反，康德的观点认为理性才是影响道德判断的首要因素，推理决定着道德判断，非理性不能影响道德判断。与休谟和康德的截然相反的观点相对应，现代心理学的道德理论存在着两种相对立的视角。

1.1 道德心理学中的两个视角

现代心理学对于道德的早期研究持有一个普遍的观点——人类有意识的道德推理决定着后续的道德判断和道德行为。这一思想是源自皮亚杰 (Jean Piaget) 和科尔伯格 (Lawrence Kohlberg) 从心理学角度对道德发展的开创性研究，他们共同的理论假设就

是个体的道德发展阶段是以其不同的道德推理水平来划分。显然，这一理论的哲学根源是康德的理性主义。康德的理性主义强调有意识的推理在道德判断过程中的重要作用，当需要我们判断行为或决定的对错时，个体是通过将外在的推理原则应用到一个具体情境中 (Murphy, Wilde, Ogden, Barnard, & Caldera, 2009)。据此哲学思想，皮亚杰从心理学的角度提出了道德发展理论，这也就成为心理学中道德研究的开创性、基石性理论。此后，另一位道德心理学家科尔伯格，在皮亚杰的基础上提出来自己的道德发展阶段理论。皮亚杰和科尔伯格都认为道德判断来自道德推理，他们只强调和承认认知因素对道德判断的作用，不承认非认知因素的作用。自此之后，广大道德心理研究者均秉承他们的道德判断的认知观而开展道德心理学的研究。因此，到目前为止的大量道德心理学的研究探讨了道德推理的重要作用。

另外一个道德心理学的理论视角，可以追溯到19世纪末。当时的心理学开始重视实证研究，道德心理学者认为推理并不重要。比如弗洛伊德 (Sigmund Freud) 就认为人们的判断来自于无意识的动机和感觉，然后才会用公众接受的理由进行辩解。而行为主义学家将道德推理看作道德行为的附带结果，认为道德行为只是社会奖励或惩罚的结果。伴随着上世纪60年代的认知革命，心理学家们开始一边倒的关注道德推理。但在上世纪末，受情感革命的影响，一些道德心理学家们开始认识到过去几十年的道德理论和研究过分夸大了有意识的理性因素（特别是道德推理）的作用而忽视无意识因素的影响，于是休谟的哲学思想又开始占据道德心理学理论。持这一观点的学者们认为至少有些道德判断是无意识心理过程的结果，这些无意识的心理过程也可以被认为是直觉性的。于是，道德心理学研究者又重新将研究焦点集中到非认知因素，其中受到更多关注的是道德直觉和道德情绪 (Hauser, Cushman, Young, Jin, & Mikhail, 2007)。

在目前的道德心理学研究中，研究的出发点几乎都是上述两种视角中的一个。而最近的研究者更多是以道德直觉和道德情绪来探讨道德心理，他们认为在道德心理和行为中推理并不会起重要作用。这其中以 Haidt 的社会直觉模型 (Social Intuitionist Model, SIM) 最为有名，Haidt 认为道德推理几乎不会直接影响道德判断，除非需要才会有慢速的、事后的道德推理来影响道德判断 (Haidt,2001; 徐平，迟毓凯，2007;)。从 Haidt 的观点来看，个体的道德推理与其道德行为没有直接关系。但大量的研究结果均表明道德推理与道德行为是有关系的，这里就存在一个疑问：道德推理与道德行为有关系吗？

1.2 道德推理、道德判断

道德推理、道德判断是道德心理学家最早关注的因素之一，也是目前为止道德心理学理论中最为重要的构成因素。道德推理是指个人运用已有的道德概念和道德认识，对道德现象进行分析、评价、推断和选择的心理过程 (余宏波，刘桂珍，2006)。道德判断则是指个体（基于内心的道德原则）对哪些是道德的进行决策和判断的能力，以及能够根据这些判断付诸行为的能力 (杨韶刚，吴慧红，2006)。简言之，可以将道德判断定义为对一个人的行为或特性的评价（好或坏）(徐平，迟毓凯，2007)。从定义上可以清晰发现这两者之间的区别，但是大量的研究将这两者混为一体，很多的研究者将道德推理冠以道德判断 (Maeda, Thoma, & Bebeau, 2009; Narvaez & Gleason, 2007)。虽然这两者都是道德心理中的认知因素，但它们是不同的。道德推理侧重于推理与分析，它关注道德现象背后的理由与解释。而道德判断则关注评价与判断，不一定涉及判断的理由。因此，道德推理可以看作是道德判断的深入，是对判断结果的探讨。这两者是完全不同的，个体可能有相同的道德判断，但是其道德推理可以完全不一样。比如，小学生可能都会认为帮助同学是好的，但认为这样做可以得到老师的认可与可以得到他人的回报是两种完全不同的道德推理，也反映个体不同的道德水平。

1.3 道德行为

道德行为是一个涵盖面很广的道德因素，很多学者给出了自己的定义。比如道德行为是指在道德

意志支配下表现出来的符合社会道德规范的行为 (刘华山 , 2008)。彭蕾则认为道德行为是人们在道德方面有意识的行动，是指个体在一定道德意识的支配下所表现出来的有利或有害于他人与社会的实际行动，是个体道德品质的外在表现 (彭蕾 , 2004)。心理学是从知、情、意、行来划分道德结构，这里面的行就是指道德行为。道德行为的另一面应该是非道德领域的行为（如学习、工作），即非道德行为。从上述定义来看，道德行为应该包含两部分——道德的行为与不道德的行为，而在目前的研究中学者们则会使用很多与这两者相近似的名词，比如助人行为、亲社会行为、利他行为、攻击行为、反社会行为等 (Eisenberg, Zhou, & Koller, 2001; Janssens & Dekovi 1997; Manning & Bear, 2011; Shumaker, 2006;Wyatt & Carlo, 2002)。但总的来说，道德行为是道德研究的最终点，所有的研究都是为了促进道德的行为、抑制不道德的行为。我们对于道德行为的探讨应该包含上述两方面的内容，这样才能完整地展现道德行为。

1.4 道德推理与道德行为的关系

针对道德的行为的研究中，Eisenberg 的团队成果最为丰富，他们探讨了青少年阶段、成年早期的亲社会行为发展，发现道德推理与其亲社会行为相关 (Eisenberg , Carlo, & Murphy,1995; Eisenberg, Miller, Shell, McNally, & Shea, 1991; Eisenberg et al., 2002)。此外，研究者也发现儿童、成人、护士、会计师、商业人士的道德推理水平与道德的行为有显著的正相关 (Ketefian, 1981; Krebs & Rosenwald, 1977; Malti, Gasser, & Gutzwiller–Helfenfinger, 2010; Ryan, 2001)。同时，有一些研究结果发现道德推理并不会与道德的行为有关系 (Lai, Siu, Chan, & Shek, 2012; Schonert–Reichl, 1999; Simmons & Zumpf, 1986; 朱丹 , 李丹 , 2005)。

另一方面，大量的研究发现个体的道德推理与攻击行为、青少年犯罪、逃学等有显著的负相关 (Gasser & Malti, 2012; Guzman, 2006; Wyatt & Carlo, 2002)，但也有很多的研究发现道德推理与不道德行为并没有联系 (Lai et al., 2012; Richards, Bear, Stewart, & Norman, 1992)。

对于道德推理与道德行为的关系，以认知发展理论的角度来看，道德认知因素应该起着重要的作用 (Blasi, 1980)。而道德推理应该是最为重要的认知因素，因此道德推理理应与道德行为有密切关系，这一推论也符合康德的思想。但近 10 年来，道德心理学研究中开始重视非认知因素的作用、减弱认知因素的作用 (Eisenberg, 2000;Haidt, 2001; Tangney, Stuewig, & Mashek, 2007)。从这些学者的理论来看，非认知因素才是道德行为的重要影响因素，道德推理等认知因素可能与道德行为没有关系。针对这两种观点，本研究想通过元分析来探讨道德推理与道德行为（道德的行为与不道德的行为）的关系。依据皮亚杰与科尔伯格的开创性研究与理论，我们提出假设：道德推理与道德的行为之间有正相关，道德推理与不道德的行为之间有负相关。

1.5 道德推理与道德行为关系的调节变量

元分析技术不仅仅是得到一个合成效应量，它还可以就这个效应量的影响因素展开分析。由于本元分析是探讨道德推理与道德行为之间的关系，所以可以将效应量的影响因素看作是一个调节变量，而本研究想探讨测量工具和被试年龄阶段这两个调节变量。

1.5.1 道德推理测量工具的类型

在几十年的道德研究中，出现了很多的道德推理测量工具。我们认为可以分为两大类：非结构化测量和标准的结构化测量。第一类工具是以访谈的形式提供一些小故事，然后让被试判断行为的合理性或允许性，并指出其理由。在数据分析时，研究者则根据一定的标准进行计分，或将被试的道德推理分成不同的类型。最为常用的故事当属科尔伯格编制的列车困境和天桥困境 (Cushman, Young, & Hauser, 2006; Greene et al., 2009)。此外，研究者也采用自编的故事，如哭泣的婴儿、背叛的妻子、严刑拷问等 (Banerjee, Huebner, & Hauser, 2010; Tarrant, Branscombe, Warner, & Weston, 2012; Ugazio, Lamm, & Singer, 2012; Vandello, Michniewicz, & Goldschmid, 2011)。

第二类工具是按照标准的心理测量学程序

编制而成，同时在施测时也有严格的操作要求。这类工具中使用较多的有 MJT(the Moral Judgment Test)、DIT(the Defining Issues Test) 和 PROM(Prosocial Reasoning Objective Measure)。它们都依据科尔伯格的理论，向被试呈现一定数量的小故事，被试需要做出行为选择，最后还要对行为选择的理由进行评定。对于被试的作答，MJT、DIT 和 PROM 以自己独特的计分方式给每个被试一个道德推理分数。

大量的道德研究采用上述两类测量工具 (Banerjee et al., 2010; Hardy, 2006, Eisenberg et al., 2001; Maeda et al., 2009; Mouratidou, Barkoukis, & Rizos, 2012; Ugazio et al., 2012)，但我们认为第一类工具存在以下问题：①道德两难情景不统一。很多研究者会采用自编的故事来进行测量，但这些测量工具并没有进行严格的心理测量学检验。就算是采用列车困境或天桥困境的研究，对于故事的描述也存在不一致。比如有的研究中会要求被试回答是否应该救多数铁道工，而另一些研究则询问是否不该杀害那个铁道工。已有的研究早已表明，这一不同表述会影响个体的道德推理 (Broeders, van der Bos, Muller, & Ham, 2011; Christensen & Gomila, 2012)，因此这种测量方法有很大的研究特异性。②道德困境多涉及生死。这一主题的特殊性可能使其完全不同于其他道德话题（如违反规则等），对这一问题的推理是否可以完全反映被试的道德推理，值得商榷 (Graham et al., 2011)。③工具计分的主观性强。没有一个统一的严格计分方式，导致不同的研究结果不能进行比较。而主试的语言表达能力会极大影响被试的作答，被试的语言表达能力也会影响主试的记录 (Carlo, Eisenberg, & Knight, 1992; Rest, Cooper, Coder, Masanz, & Anderson, 1974)。因此，我们认为道德推理测量工具会影响元分析的结果，假设采用不同测量工具的研究中道德推理与道德行为的关系有显著差异。

1.5.2 被试年龄阶段

皮亚杰与科尔伯格的道德理论都认为个体的道德推理能力是发展的，不同发展阶段的个体有不同的道德推理水平或类型。已有研究指出不同道德推理水

平导致的道德行为是不一样的 (Blasi, 1980; Comunian & Gielen, 1995; Eisenberg, Cumberland, Guthrie, Murphy, & Shepard, 2005)，而不同年龄阶段（儿童、青少年和成人）的个体在道德稳定性与认知能力上是有差异的。因此假设对不同年龄阶段的被试，其道德推理与道德行为之间的关系存在显著差异。

2 研究方法

2.1 文献搜索

本研究全面搜索了相关文献，包含了中文和英文文献搜索。中文文献的搜索过程如下：首先，在 CNKI 数据库、中国科技期刊数据库、万方数据库、中国优秀硕士学位论文全文数据库以及中国博士学位论文全文数据库中，以（*道德推理或道德判断*）或（*道德行为、亲社会行为、助人行为、利他行为、不道德行为、攻击行为、反社会行为或青少年犯罪*）为关键词进行搜索。此外，也在互联网 google 学术中以相应关键词进行搜索。英文文献的搜索过程：在 PsycARTICLES，PsycINFO，JSTOR，SAGE，Springer，Elsevier，ProQuest 博硕士论文全文数据库中，以 (moral reasoning 或 moral judgment) 或 (moral behavior、moral conduct、moral action、prosocial behavior、helping behavior、altruistic behavior、altruism、immoral behavior、immoral conduct、immoral action、aggression、aggressive behavior、antisocial behavior、delinquency、truancy) 为关键词进行搜索。同时，也在互联网 google 学术中进行搜索。

对于搜索到的、但没有结果内容的文献记录，我们尽量通过可以寻找到的联络方式给作者发送电子邮件以获取全文或结果。

2.2 文献纳入的标准

对于搜索到的相关研究，我们按照以下标准来决定是否将其纳入后面的元分析：①必须报告了数字结果的实证研究，而纯理论的、综述性的研究被排除。②如果仅仅只进行了道德判断测试，即只呈现了被试的判断结果，没有进一步测试判断理由的研究将

表 1 元分析中纳入的原始研究

研究	样本量	性别群体	年龄	结果变量类型	工具类型	相关系数
Aleixo&Norris, 2000	100	M	Ad	IB	U	−0.230
Barriga et al., 2001	193	B	A	IB	U	−0.200
Bear &Richards, 1981	91	B	C	IB	U	−0.290
Bear & Rys, 1994	133	B	C	IB	U	−0.228
Bear & Rys, 1994	60	F	C	IB	U	−0.040
Bear & Rys, 1994	73	M	C	IB	U	−0.380
Bear, 1999	77	B	C	IB	U	−0.290
Blair, Monson, & Frederickson, 2001	102	B	C	IB	U	−0.034
Blair, Monson, & Frederickson, 2001	102	B	C	IB	U	−0.005
Blair, Monson, & Frederickson, 2001	102	B	C	IB	U	−0.070
Blair, Monson, & Frederickson, 2001	102	B	C	IB	U	−0.138
Bredemeier, 1994	106	B	C	IB	U	−0.386
Bredemeier, 1994	106	B	C	IB	U	−0.375
Bredemeier, 1994	106	B	C	IB	U	−0.280
Bredemeier, 1994	106	B	C	IB	U	−0.464
Bruggeman & Hart, 1996	221	B	Ad	IB	S	−0.030
Carlo & Randall, 2002	249	B	A	MB	S	0.450
Carlo et al., 2003	80	B	A	MB	S	0.250
Carlo et al., 2003	58	B	A	MB	S	0.220
Carlo et al., 1996	130	B	A	MB	S	0.270
Carlo et al., 1996	55	M	A	MB	S	0.260
Carlo et al., 1996	75	F	A	MB	S	0.280
Cummings, 2001	145	B	Ad	IB	S	−0.202
Eisenberg−berg & Hand, 1979*	35	B	C	IB	U	0.000
Eisenberg et al., 1985	58	B	C	MB	U	0.310
Eisenberg et al., 1991	64	B	A	MB	U	0.300
Eisenberg et al., 1995	32	B	A	MB	S	0.510
Eisenberg et al., 2002	30	B	Ad	MB	S	0.420
Eisenberg, Zhou, & Koller, 2001	149	B	A	MB	S	0.300
Gasser & Malti, 2012	118	B	C	IB	U	−0.230
Gasser & Malti, 2012	118	B	C	IB	U	−0.180
Guzman, 2006	192	B	A	IB	U	−0.140
Guzman, 2006	195	B	A	IB	U	−0.200
Hardy, 2006	91	B	Ad	MB	S	0.040
Janssens & Dekovic, 1997	125	B	C	MB	S	0.170
Ketefian, 1981	79	B	Ad	MB	S	0.280
Krebs & Rosewald, 1977	31	B	Ad	MB	S	0.525
Kumru et al., 2012	330	B	A	MB	S	0.180
Kumru et al., 2012	1252	B	A	MB	S	0.170
Lai et al., 2012	566	B	A	IB	S	−0.050
Lai et al., 2012	566	B	A	MB	S	0.100
Malinowski & Smith, 1985	53	M	Ad	IB	S	−0.390
Malti, Gasser, & Gutzwiller, 2010	312	B	C	IB	U	−0.170
Malti, Gasser, & Gutzwiller, 2010	312	B	C	MB	U	0.060
Manning & Bear, 2011	216	B	C	IB	U	−0.270
Manning & Bear, 2011	132	B	A	IB	U	−0.210
Miller et al., 1996	74	B	C	MB	U	0.240
Palmer & Hollin, 1996	64	B	Ad	IB	U	−0.200
Palmer & Hollin, 2000	58	B	Ad	IB	U	−0.250
Palmer & Hollin, 2001	94	B	A	IB	U	−0.440
Raaijmakers, Engles, & Hoof, 2005	846	B	A	IB	S	−0.120
Raaijmakers, Engles, & Hoof, 2005	846	B	A	IB	S	−0.150
Raaijmakers, Engles, & Hoof, 2005	846	B	A	IB	S	−0.170

续表 1

研究	样本量	性别群体	年龄	结果变量类型	工具类型	相关系数
Richards et al., 1992	143	B	C	IB	U	−0.100
Rubin & Schneider, 1973	57	B	C	MB	U	0.310
Ryan, 2001	116	B	Ad	MB	S	0.210
Schonert–Reichl, 1999	54	F	A	MB	U	0.320
Schonert–Reichl, 1999	54	M	A	MB	U	0.160
Schonert–Reichl, 1999	54	F	A	IB	U	−0.080
Schonert–Reichl, 1999	54	M	A	IB	U	−0.220
Shumaker, 1993	64	B	A	MB	U	0.260
Simmons, 1996	428	B	Ad	IB	S	−0.010
Simmons, 1996	428	B	Ad	IB	S	−0.050
Simmons, 1996	429	B	Ad	IB	S	−0.020
Simmons, 1996	429	B	Ad	IB	S	−0.010
Tarry & Emler, 2007	475	M	A	IB	U	−0.080
Turner et al., 2007	74	B	As	MB	S	0.260
Wyatt & Carlo, 2002	80	B	A	MB	S	0.280
Wyatt & Carlo, 2002	76	B	A	MB	S	0.280
Wyatt & Carlo, 2002	58	B	A	MB	S	0.220
Wyatt & Carlo, 2002	80	B	A	IB	S	−0.420
Wyatt & Carlo, 2002	76	B	A	IB	S	−0.270
Wyatt & Carlo, 2002	58	B	A	IB	S	−0.100
Wyatt & Carlo, 2002	76	B	A	IB	S	−0.260
Wyatt & Carlo, 2002	58	B	A	IB	S	−0.430
洪丽, 2005	481	B	A	MB	U	0.049
李炜, 2012	203	B	Ad	MB	S	0.680
刘美辰, 2012	580	B	A	IB	U	−0.353
刘志军, 2001	286	B	A	MB	U	0.014
毛静思, 2001	667	B	A	MB	U	0.083
彭蕾, 2004	410	B	C	MB	U	0.202
朱丹 & 李丹, 2005	217	B	C	MB	U	0.026

注：B 表示研究样本含男性与女性，M 表示研究样本为男性，F 表示研究样本为女性，C 表示样本为儿童，A 表示样本为青少年，Ad 表示样本为成人，MB 表示结果变量为道德的行为，IB 表示结果变量为不道德的行为，S 表示道德推理为标准化工具，U 表示道德推理为非标准化工具。

被排除。③研究探讨的行为必须是道德领域的，非道德领域行为的研究将被排除。④没有报告完整效应量的研究将被排除。最终，我们得到符合元分析要求的文献有 50 篇。其中公开发表的文献有 43 篇，中文文献有 7 篇。

2.3 文献编码

对纳入元分析的文献进行如下编码：文献信息（作者名 + 文献时间），样本性别群体（男性、女性或两者均有），样本年龄（儿童、青少年或成人），结果变量类型（道德的行为、不道德的行为），道德推理测量工具类型（标准化、非标准化），见表 1。

针对每一个独立样本，得到一个效应量。同时，考虑有的研究针对多种行为变量（道德的行为与不道德的行为），有的研究则报告了不同样本群体的结果（男性与女性），有的研究则报告了多次测量的结果（追踪研究）。我们分别呈现每一个研究文献中包含的多个独立效应量，于是有的研究文献会包含多个独立效应量。最后，我们一共得到 83 个独立的效应量。

2.4 元分析过程

2.4.1 效应量

最终纳入的 50 篇研究文献主要报告了道德推理与道德行为的相关系数，因此我们的元分析以相关系数作为效应量。其中，在研究文献 (Bear & Rys, 1994) 中，研究者分别报告了男性与女性被试道德推理与道德行为的相关系数和样本量，我们采用相关系数

合成的方法 (*r–Fisher Z*) 得到被试总样本道德推理与道德行为的相关系数 (张厚灿 , 徐建平 , 2004)。另一文献中 (Krebs & Rosenwald, 1977)，研究者没有报告相关系数，但呈现了不同道德推理水平被试是否实施助人行为的具体人数，我们通过计算得到道德推理与道德行为的相关系数。

2.4.2 模型的选定

目前的元分析主要采用固定效应模型或随机效应模型，这两者最主要的区别在于权重成分的不一样。固定效应模型假设元分析中所有研究背后只存在一个真效应量，而每个研究效应量的不同是由抽样误差引起的。随机效应模型则认为每个研究的真效应量都是不同的，每个研究效应量的不同是由真效应量的不同和抽样误差共同引起的 (Borenstein, Hedges, Higgins, & Rothstein, 2009)。两个模型的不同假设会导致元分析中平均效应量的显著性检验、区间估计以及调节变量的显著性检验方法不同 (Hunter & Schmidt, 2000)。在进行元分析之前，研究者就应该从理论与实际层面选定好模型。而不能先假设一个模型开始分析，结果发现与假设不符又换另一个模型进行分析 (Borenstein et al., 2009)。在模型的选定上，Borenstein 等建议主要考虑元分析的研究是否拥有一个共同的效应量以及元分析的目的。具体来说，如果认为元分析中的研究在功能上是相同的，而我们的元分析得到的总效应量只是针对包含的研究所涉及的总体，不推广到其它总体的话，我们应该使用固定效应模型。相反，如果元分析中包含的研究中被试群体、测量工具不同，并且有理由相信这种不一样会影响结果时，就不能假设存在一个真效应量，此时使用随机效应模型更加合理 (Borenstein et al., 2009)。

在我们最终确定的 50 篇研究文献中，被试包含儿童、青少年、成年人等，被试职业涵盖学生、商务人士、护士等。要进行元分析的研究文献中被试各异，元分析得到的效应量不能只局限于某一个研究所涉及的样本群体，因此不适合采用固定效应模型。此外，我们的元分析本来就想探讨测量工具的调节作用，

因此我们有理由相信随机效应模型更适合本元分析。在后面的元分析中，将采用异质性检验来验证我们的模型选择。

2.4.3 发表偏差

当被发表的研究文献系统性地不能代表该领域已经完成的研究总体时，就认为发生了发表偏差 (Rothstein, Sutton, & Borenstein, 2006)。发表偏差的结果就是某一领域的研究文献不完整，这会严重影响元分析的结果。任何一个元分析研究都应该关注发表偏差的问题，因为它会导致最终得到的效应高于真实值 (Kuppens, Laurent, Heyvaert, & Onghena, 2013)。针对发表偏差的问题，我们首先在文献搜索阶段尽可能获取了没有发表的文献。在后面的元分析过程中，我们还会采用漏斗图 (funnel plot)、Rosenthal's *Fail–safe N* 与 Egger's 检验等方法来评估本元分析的发表偏差。

2.4.4 元分析过程及软件

我们的元分析首先想探讨道德推理与道德行为之间的关系，考虑到道德行为分为道德的行为与不道德的行为，而道德推理与这两种行为之间的相关方向是相反的，于是我们分别针对道德的行为与不道德的行为来计算总效应量。在调节变量的检验过程中，我们同样是分开分析。本研究采用 CMA 2.2(Comprehensive Meta Analysis 2.2) 进行元分析。

表 2　效应量异质性检验结果

结果变量	Q	df	p	I^2	τ^2
道德的行为	168.567	34	<0.001	79.830	0.022
不道德的行为	147.342	47	<0.001	68.101	0.011

3　研究结果

3.1 异质性检验

针对道德的行为与不道德的行为分别进行异质性检验，结果见表 2。

从表 2 的结果来看，两个结果变量的 Q 检验均显著，表明元分析中各研究的效应量是异质的。另外，依据 Borenstein 等人对 I^2 的解释 (Borenstein et al.,

2009)，针对道德的行为的元分析的 I^2 为 79.830，说明在道德推理与道德的行为的关系研究中有 79.83% 的观察变异是由这一关系中真正差异所造成的。针对不道德的行为，I^2 为 68.101，说明在不道德推理与道德的行为的关系研究中有 68.10% 的观察变异是由这一关系中真正差异所造成的。σ^2 表示真效应量的方差，两个 σ^2 表明真效应量都有一定的变异。异质性检验的结果表明，我们选定以随机效应模型来进行元分析是准确的。

3.2 发表偏差检验

首先，通过漏斗图 (funnel plot) 来检查本元分析的发表偏差，两类道德行为的漏斗图见图 1 和图 2。从漏斗图来看，涉及道德的行为的研究文献并未均匀分布于总效应量两侧，多数研究位于总效应量右侧。而涉及不道德的行为的研究文献基本均匀分布于总

效应量两侧。这一分布特点表明，针对道德的行为的研究可能存在发表偏差，而针对不道德的行为的研究不存在发表偏差。因为漏斗图仅仅只能从主观的角度初步检查发表偏差，为了更准确的检验发表偏差，我们紧接着进行了 Rosenthal's *Fail-safe N* 与 Egger's 检验，结果见表 3。

从 Egger's 检验的结果来看，涉及道德的行为的研究与涉及不道德的行为的研究均存在一定的发表偏差。从 Rosenthal's *N* 值来看，需要再纳入大量 (>2200) 涉及两个行为的研究文献才可能使两个总效应量不显著，这说明涉及两个行为的本研究并不存在严重的发表偏差。

上述 3 个发表偏差检验中，有两个结果 (漏斗图和 Rosenthal's *N*) 表明针对不道德的行为的元分析不存在发表偏差，一个结果 (Rosenthal's *N*) 表明针对道德的行为的元分析不存在发表偏差，都没有得到三
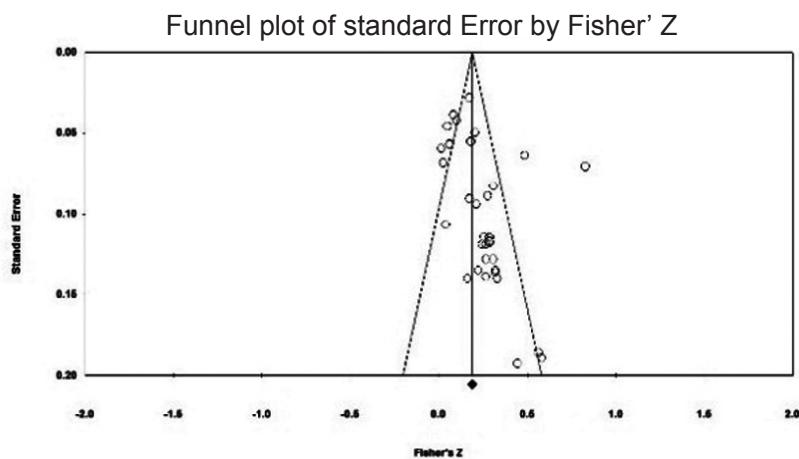
图 1　涉及道德的行为的研究的漏斗图
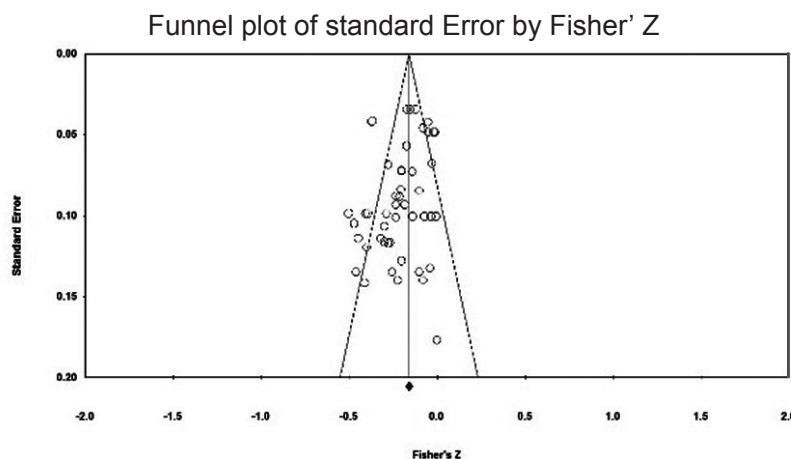
图 2　涉及不道德的行为的研究的漏斗图

表 3 发表偏差检验结果

结果变量	Rosenthal's N	Egger's Intercept	SE	LL	UL	p
道德的行为	2285	1.78240	0.72154	0.31443	3.25038	0.019
不道德的行为	3293	−1.52362	0.54727	−2.62521	−0.42203	0.008

注：LL、UL 表示 Egger's Intercept 的 95% 置信区间的下限与上限。

表 4 道德推理与道德行为关系的随机效应模型分析结果

结果变量	N	k	r	LL	UL	Z	p
道德的行为	6663	35	0.238	0.181	0.293	7.981	<0.001
不道德的行为	10065	48	−0.188	−0.225	−0.151	−9.734	<0.001

注：N 表示样本量，k 表示研究个数，LL、UL 表示 r 的 95% 置信区间的下限与上限。

个检验都认可的结果。但按照 Borenstein 等人的看法，发表偏差的检验目的应该是确定元分析结果属于以下哪种类型：①偏差的影响可以忽略不计；②偏差的影响不能忽略，但研究结果还是有效的；③研究结果可能存在问题 (Borenstein et al., 2009)。因此需要作进一步分析，我们采用 Duval 和 Tweedie 提出的剪粘法

(Trim and Fill) 来检验发表偏差对元分析结果造成的影响 (Duval & Tweedie, 2000)。结果发现，剪粘研究文献后，针对两种行为分别采用随机效应模型得到的总效应仍然都显著。此外，最终进行元分析的文献中，未发表的文献占 14%，这一比例已经很大。综合以上结果表明虽然本研究的两个元分析中可能存在轻
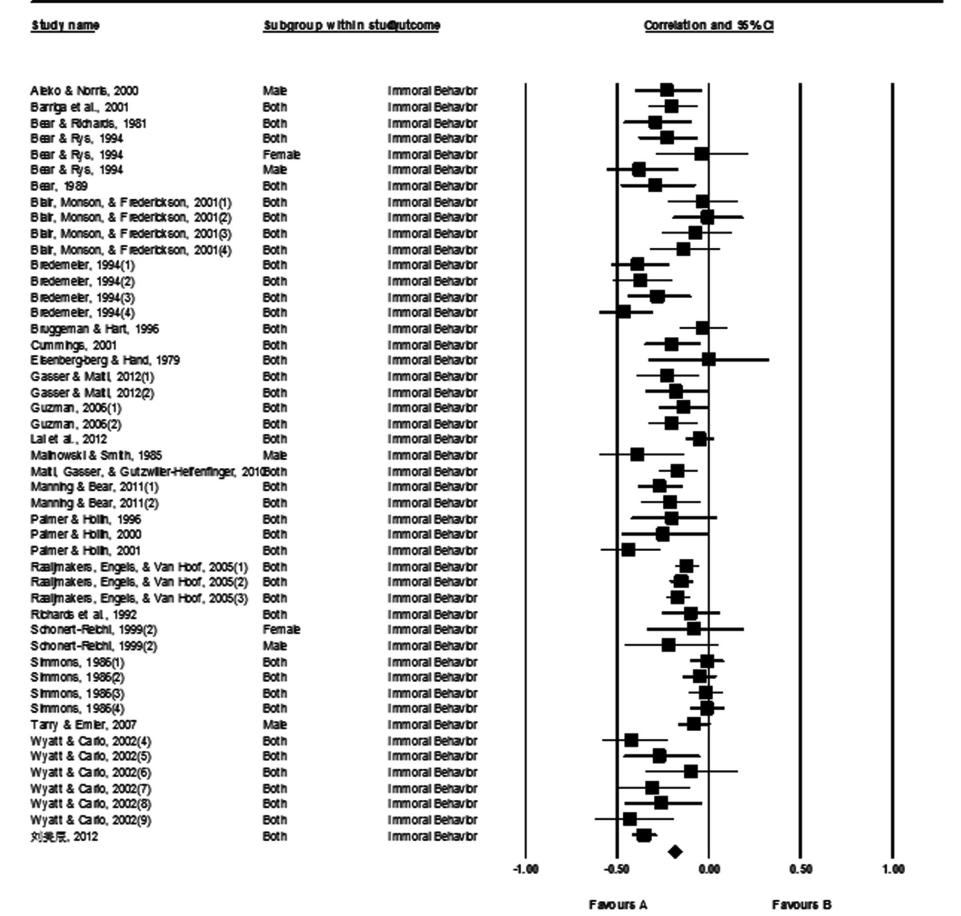
图 3 针对道德的行为的效应量的分布图

微的发表偏差，但是元分析的主要结论还是有效的。

3.3 主效应

从整体检验道德推理与道德行为的关系，结果见表 4。结果表明，共有 35 项独立的道德推理与道德的行为的效应量，被试总数为 6663，道德推理与道德的行为的整体相关系数为 0.238，见图 3。共有 48 项独立的道德推理与不道德的行为的效应量，被试总数为 10065，道德推理与不道德的行为的整体相关系数为 –0.188，见图 4。

3.4 调节效应检验

从图 3 与图 4 来看，各个研究的效应量分布于总效应量（图中菱形）左右两侧，而且各研究的效应量之间存在很大的变异。为了分析这一变异，我们分别检验道德推理测量工具类型（标准化与非标准化）、被试年龄（儿童、青少年与成人）对道德推理与道德行为关系的调节作用，结果见表 5。

从调节效应分析的结果来看，道德推理测量工具的类型可以影响道德推理与道德的行为间的关系 (Q_b=9.577，p= 0.002)，也可以影响道德推理与不道德的行为间的关系 (Q_b= 5.550，p= 0.018)。被试的年龄只能影响道德推理与不道德的行为间的关系 (Q_b= 10.183，p=0.006)。

4　讨论

本研究是道德心理学领域首次通过元分析技术
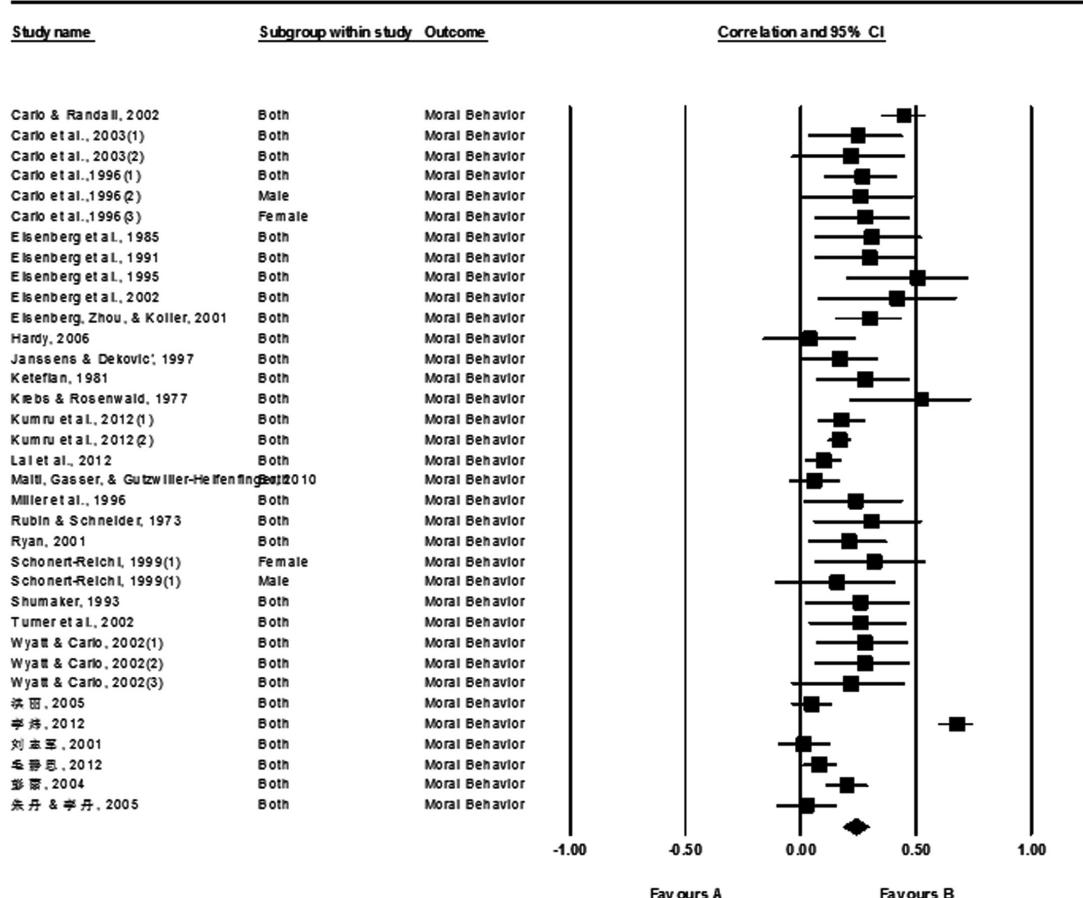
图 4　针对不道德的行为的效应量的分布图

表 5 道德推理与道德行为关系的调节效应分析结果

结果变量	调节变量	类别	k	r	LL	UL	Q_b	p
道德的行为	工具类型	标准化	22	0.288	0.209	0.363	9.577	0.002
		非标准化	13	0.133	0.075	0.191		
	被试年龄	儿童	7	0.157	0.076	0.236	3.129	0.209
		青少年	21	0.213	0.156	0.268		
		成人	7	0.361	0.123	0.559		
不道德的行为	工具类型	标准化	17	−0.136	−0.186	−0.084	5.550	0.018
		非标准化	31	−0.218	−0.262	−0.172		
	被试年龄	儿童	19	−0.216	−0.275	−0.157	10.183	0.006
		青少年	19	−0.206	−0.259	−0.153		
		成人	10	−0.090	−0.153	−0.027		

注：k 表示研究个数，LL、UL 表示 r 的 95% 置信区间的下限与上限，Q_b 表示类别间异质性的 Q 检验。

来整合以往相关的研究结果，以探讨道德"理性取向"的合理性。具体来说，我们整合了以往研究中道德推理与道德行为之间的关系，并探究这一关系的影响因素。我们的元分析结果支持了道德的"理性取向"（道德推理与道德行为有显著关系），而调节效应的分析结果表明道德推理测量工具的类型与被试的年龄都会影响道德推理与道德行为之间的关系。

4.1 道德推理与道德行为的关系

主效应分析结果表明道德推理与道德的行为的平均相关系数为 0.238，与不道德的行为的平均相关系数为 −0.188，这两个总效应量均说明道德推理与道德行为之间存在一定联系，这与以往的研究结果相似 (Carlo, Hausmann, Christiansen, & Randall, 2003; Gasser & Malti, 2012; Malti et al., 2010)。这一显著联系可能的原因有，首先从道德推理的类型来看，道德推理是道德中的认知因素，代表了个体对道德问题的理性思考与认知，不同道德推理类型的个体代表其有不同的道德水平。以 Carlo 等人对道德推理类型的划分为例，个体有 5 种基本的道德推理：享乐主义的、基于需求的、基于认可的、刻板印象的与内化的。通过对这 5 种基本推理的加权可计算出个体的道德推理得分，其中内化被赋予最大的权重 (Carlo et al., 1992; Eisenberg et al., 1995)。显然，道德推理得分越高，表明个体道德水平越高，也就更认可社会普遍接受的道德准则。在不同文化的共同道德准则中，道德的行为都受到推崇与赞许，不道德的行为都受到谴责与责备。于是，道德推理水平高的个体更可能实施道德的行为，也更

可能抑制不道德的行为。再从内容来看，高水平的道德推理涉及为了他人而牺牲自己，而道德的行为总是会导致实施者做出一定的"牺牲"，可能是体力、时间、金钱甚至生命。相反，不道德的行为会带给他人某种"牺牲"而使自己获得"利益" (Eisenberg, Boehnke, Schuhler, & Silbereisen, 1985)，高水平的道德推理与道德的行为具有同样的目标，而与不道德的行为则目标相反。因此，道德推理与道德行为之间存在正相关，与不道德的行为之间存在负相关。

需要指出的是，我们的研究结果也表明道德推理与道德行为的相关程度还是很小的，也就是说道德行为还与其他很多因素有关。相对于 Haidt 等学者对道德认知的批判，Greene 等学者则整合了认知因素与非认知因素在道德中的作用，他们提出的道德双加工模型 (dual−process model) 认为情绪与推理都应该在道德中有一定的作用 (Koven，2011)。而 2008 年 Leffel 针对道德行为提出的道德动机社会直觉模型则明确指出在激发的道德情景与道德行为的产生之间有以下 6 个影响因素: 道德直觉、道德情绪、道德美德、道德价值观、道德推理和道德意志 (Leffel，2008)。此外，道德行为还会受到道德同一性、道德推脱、道德敏感性等影响 (吴鹏，刘华山，鲁路捷，田梦潇，2013)。总体来说，道德推理与道德行为确有一定的联系，但道德行为还与很多其他因素相关。

4.2 道德推理与道德行为关系的影响因素

4.2.1 道德推理测量工具类型的作用

调节效应分析表明，道德推理测量工具的类型

既可以影响道德推理与道德的行为间的关系，也可以影响道德推理与不道德的行为间的关系。首先，就道德推理测量工具的类型对道德推理与道德的行为间的关系的影响，研究结果表明在使用标准化工具来测量道德推理的研究中，道德推理与道德的行为间的平均相关系数为 0.288；在使用非标准化工具来测量道德推理的研究中，道德推理与道德的行为间的平均相关系数为 0.133。测量工具类型的作用显著则表明，采用标准化工具能更准确地探讨道德推理与道德的行为间的关系。出现这一结果的原因可能是，本研究区分标准化的主要依据是测量工具的测试过程、计分方式与测试内容的标准化程度，非标准化工具有很强的随意性与主观性，测试过程中主试与被试的个体特征会影响整个测试，而计分方法也容易受主试的影响。因此，非标准化的测量工具并不能准确、客观地评价个体的道德推理水平 (Carlo et al., 1992; Rest et al., 1974)。这就导致在使用非标准化工具的研究中，道德推理与道德的行为间的相关较低。

其次，就道德推理测量工具的类型对道德推理与不道德的行为间的关系的影响，研究结果表明在使用标准化工具来测量道德推理的研究中，道德推理与不道德的行为间的平均相关系数为 –0.136；在使用非标准化工具来测量道德推理的研究中，道德推理与不道德的行为间的平均相关系数为 –0.218。测量工具类型的作用显著则表明，采用标准化工具来探讨道德推理与不道德的行为间的关系会比较弱。造成这一结果的原因可能是，从早期科尔伯格开始采用道德两难困境来测试个体的道德推理开始，大量道德推理测量工具一直都采用这一模式。这些道德推理测试都会呈现几个道德情景，让被试对某些行为选择进行推理。以本研究界定的标准化工具为例，它们主要针对于个体的道德的行为展开，如"海因茨偷药"、"工厂风波"、"救助他人"等故事情景 (Carlo et al., 1992; Rest, Narvaez, Thoma, & Bebeau, 1999; 吴慧红, 2005)。显然，这样的工具更多关注了对道德的行为的推理，并没有涉及对不道德的行为的推理。加上标准化的工具更加准确、有效、稳定地测试了个体的

道德推理 (Carlo et al., 1992)，于是就会导致使用标准化工具的研究中道德推理与不道德的行为相关较低。道德推理测量工具类型的调节效应的分析结果说明，在探讨道德的行为时，我们应该采用标准化工具来测查道德推理；而在研究不道德的行为时，目前的标准化工具也是不适当的，我们应该开发针对不道德行为的道德推理测量工具。

4.2.2 年龄阶段的作用

检验被试年龄阶段调节作用的结果表明被试的年龄阶段只影响道德推理与不道德的行为间的关系，儿童的道德推理与不道德的行为的相关系数为 –0.216，青少年的等于 –0.206，成人则等于 –0.090。被试的年龄阶段的调节作用显著说明，相比于儿童与青少年，成人的道德推理与不道德的行为的关系较弱。这一调节效应的可能原因是就个体道德发展来说，儿童与青少年正处于道德逐步稳定的阶段 (Bar-Tal & Nissim, 1984; Cushman et al., 2006)。他们会接受很多道德教育，学校与家庭会教给儿童与青少年正确的价值观、道德推理与道德判断，这些符合社会文化的道德因素会让儿童与青少年认知到"不道德行为是不被接受的、不能实施的"。按照科尔伯格的理论，儿童与青少年的道德属于"绝对化阶段"，他们对于道德行为是做出绝对对错的判断，不会寻求理由使不道德的行为合理化。同时，相比于道德的行为，学校与家庭可能更加关注不道德的行为，更凸显这一行为的不良后果，儿童与青少年会被要求去抑制自己的不道德行为。而儿童与青少年越是按社会文化要求进行道德推理，越能认识到不道德行为的危害，也越不会做出不道德行为。因此儿童与青少年的道德推理得分越高，其不道德行为应该越少，从而使儿童与青少年的道德推理与不道德的行为有一定的负相关 (Gasser & Malti, 2012; Guzman, 2006)。而就成人来说，其认知能力已经很稳定与成熟，因此都会进行社会普遍采用的道德推理 (Eisenberg et al., 2002; Reynolds & Ceranic, 2007)，这就导致成人间的道德推理水平差异不大。而成人间的不道德行为会有很大的差异，也就是说相比于儿童与青少年，成人之间的不道德行为数量会有显著的差异，有些成人会

经常做出不道德行为 (Eisenberg et al., 2002)。因为不道德行为往往会带来一定的"好处"，比如金钱、权势等等，成人会有更强烈的动机去获得这些"好处"。另一方面,成人更能够为自己的不道德行为寻找理由、做出辩解 (Detert, Trevino, & Sweitzer, 2008)，这也就可以减少不道德行为可能带来的负性影响。于是成人间的不道德行为会有很大的差异，但其道德推理已经趋于稳定与成熟，不会有很大的差异 (Eisenberg et al., 2002; Reynolds & Ceranic, 2007)，两者之间不对称的变异就导致了成人的道德推理与不道德行为的相关很弱。

本研究没有发现被试年龄阶段对道德推理与道德的行为间关系的调节作用，原因可能是：①道德的行为是整个社会大力提倡的行为，每个年龄阶段的群体都会要求自己做出道德的行为。儿童和青少年会经常被要求或奖励去做出道德的行为，学校、家庭与社会也会教育他们进行正确的道德推理，这两方面的作用就使儿童和青少年的道德推理与道德的行为之间存在一定的正相关 (Eisenberg et al., 2002; Malti et al., 2010)。而相比于儿童，成人对自己做出道德的行为会有更强烈的要求 (Carlo, Crockett, Randall, & Roesch, 2007; Maeda et al., 2009)。成人又具有更强的体力、更好的能力来完成道德的行为，因此成人会有更多的道德行为。同时，随着年龄的增长，个体的道德推理水平会越来越高 (Eisenberg et al., 2002; Narvaez & Gleason, 2007)。于是，道德推理与道德的行为随着年龄共同增长，成人的道德推理与道德的行为会都强于儿童与青少年。但相关关系只是关注这两者之间的联系，随时间共同的增长可能不会显著影响两者之间的相关系数，于是年龄阶段不会影响道德推理与道德的行为之间的关系。②正如上述提到的，目前的道德推理测量工具主要针对于道德的行为，因此道德推理与道德的行为关系可能更为密切。而一些道德推理的测量工具针对不同年龄群体有不同的版本，比如 PROM 有儿童版、青少年版与成人版。测量工具针对各个年龄阶段进行修订，确保各个年龄阶段被试都能准确作答，也就提高了道德推理的测量准确性，这也就保证了在各个年龄阶段的

道德推理与道德的行为之间都能展现出紧密的联系。③本元分析涉及的原始研究中探讨青少年的道德推理与道德的行为之间关系的研究数量 (21) 远多于针对儿童或成人的研究 (7)，3 种年龄群体的研究数量的不均衡可能也会影响元分析中调节效应的分析结果 (Borenstein et al., 2009)。

4.3 不足与展望

本研究的不足：①没有考虑性别对道德推理与道德行为间关系的可能影响。道德心理研究领域中，性别通常都是一个重要的考虑因素。由于最终纳入元分析的原始研究基本都没有报告不同性别的效应量，也就无法分析性别可能的作用。②调节效应分析中的样本较少且分配不均衡，这都会影响分析结果。③本元分析所纳入的原始研究中，未发表的国外文献数量较少。作为重要的心理学研究主题，探讨道德行为的影响因素具有很大的理论与现实意义，未来的研究可以：①在探讨道德心理与行为中理性因素的作用时，同时考虑非理性因素的作用，从"道德双加工"角度全面探讨道德心理。②针对不道德行为编制道德推理测量工具，以准确测量不同道德行为的推理。③随着年龄的增长，个体的认知能力逐步成熟。在不同的年龄阶段，道德推理与道德行为之间的关系可能有一定的差异。甚至不同年龄阶段，理性因素与非理性因素的作用强度可能会有不同，今后的研究应该关注这一方面。

5 结论

本元分析发现道德推理与道德行为之间存在联系，道德推理测量工具的类型可以影响道德推理与道德行为（道德的行为或不道德的行为）的关系，被试的年龄阶段只能影响道德推理与不道德的行为间的关系。

参考文献

Aleixo, P. A., & Norris, C. E. (2000). Personality and

moral reasoning in young offenders. *Personality and Individual Differences, 28*(3), 609–623.

Banerjee, K., Huebner, B., & Hauser, M. (2010). Intuitive moral judgments are robust across variation in gender, education, politics and religion: A large-scale web-based study. *Journal of Cognition and Culture, 10*(3), 253–281.

Barriga, A. Q., Morrison, E. M., Liau, A. K., & Gibbs, J. C. (2001). Moral cognition: Explaining the gender difference in antisocial behavior. *Merrill-Palmer Quarterly, 47*(4), 532–562.

Bar-Tal, D., & Nissim, R. (1984). Helping behaviour and moral judgment among adolescents. B*ritish Journal of Developmental Psychology, 2*(4), 329–336.

Bear, G. G. (1989). Sociomoral reasoning and antisocial behaviors among normal sixth graders. *Merrill-Palmer Quarterly, 35*(2), 181–196.

Bear, G. G., & Richards, H. C. (1981). Moral reasoning and conduct problems in the classroom. *Journal of Educational Psychology, 73*(5), 664–670.

Bear, G. G., & Rys, G. S. (1994). Moral reasoning, classroom behavior, and sociometric status among elementary school children. *Developmental Psychology, 30*(5), 633–638.

Blair, R. J. R., Monson, J., & Frederickson, N. (2001). Moral reasoning and conduct problems in children with emotional and behavioural difficulties. *Personality and Individual Differences, 31*(5), 799–811.

Blasi, A. (1980). Bridging moral cognition and moral action: A critical review of the literature. *Psychological Bulletin, 88*(1), 1–45.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2009). *Introduction to meta-analysis*. West Sussex, UK: Wiley & Sons.

Bredemeier, B. J. L. (1994). Children's moral reasoning and their assertive, aggressive, and submissive tendencies in sport and daily life. *Journal of Sport & Exercise Psychology, 16*(1), 1–14.

Broeders, R., van der Bos, K., Muller, P. A., & Ham, J. (2011). Should I save or should I not kill? How people solve moral dilemmas depends on which rule is most accessible. *Journal of Experimental Social Psychology, 47*(5), 923–934.

Bruggeman, E. L., & Hart, K. J. (1996). Cheating, lying, and moral reasoning by religious and secular high school students. *The Journal of Educational Research, 89*(6), 340–344.

Carlo, G., Crockett, L. J., Randall, B. A., & Roesch, S. C. (2007). A latent growth curve analysis of prosocial behavior among rural Adolescents. *Journal of Research on Adolescence, 17*(2), 301–324.

Carlo, G., & Randall, B. A. (2002). The development of a measure of prosocial behaviors for late adolescents. *Journal of Youth and Adolescence, 31*(1), 31–44.

Carlo, G., Hausmann, A., Christiansen, S., & Randall, B. A. (2003). Sociocognitive and behavioral correlates of a measure of prosocial tendencies for adolescents. *The Journal of Early Adolescence, 23*(1), 107–134.

Carlo, G., Koller, S. H., Eisenberg, N., Da Silva, M. S., & Frohlich, C. B. (1996). A cross-national study on the relations among prosocial moral reasoning, gender role orientations, and prosocial behaviors. *Developmental Psychology, 32*(2), 231–240.

Carlo, G., Eisenberg, N., & Knight, G. P. (1992). An objective measure of adolescents' prosocial moral reasoning. *Journal of Research on Adolescence, 2*(4), 331–349.

Christensen, J. F., & Gomila, A. (2012). Moral dilemmas in cognitive neuroscience of moral decision-making: A principled review. *Neuroscience and Biobehavioral Reviews, 36*(4), 1249–1264.

Cummings, R., Dyas, L., Maddux, C. D., & Kochman, A. (2001). Principled moral reasoning and behavior

of preservice teacher education students. *American Educational Research Journal, 38*(1), 143 – 158.

Comunian, A. L., & Gielen, U. P. (1995). Moral reasoning and prosocial action in Italian culture. *The Journal of Social Psychology, 135*(6), 699 – 706.

Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*(12), 1082 – 1089.

Detert, J. R., Trevino, L. K., & Sweitzer, V. L. (2008). Moral disengagement in ethical decision making: A study of antecedents and outcomes. *Journal of Applied Psychology, 93*(2), 374 – 391.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel - plot - based method of testing and adjusting for publication bias in meta - analysis. *Biometrics, 56*(2), 455 – 463.

Eisenberg, N. (2000). Emotion, regulation, and moral development. *Annual Review of Psychology, 51(1),* 665 – 697.

Eisenberg, N., Boehnke, K., Schuhler, P., & Silbereisen, R. K. (1985). The development of prosocial behavior and cognitions in German children. *Journal of Cross– Cultural Psychology, 16*(1), 69 – 82.

Eisenberg, N., Carlo, G., & Murphy, B. (1995). Prosocial development in late adolescence: *A longitudinal study. Child Development, 66*(4), 1179 – 1197.

Eisenberg, N., Cumberland, A., Guthrie, I. K., Murphy, B. C., & Shepard, S. A. (2005). Age changes in prosocial responding and moral reasoning in adolescence and early adulthood. *Journal of Research on Adolescence, 15*(3), 235 – 260.

Eisenberg, N., Guthrie, I. K., Cumberland, A., Murphy, B. C., Shepard, S. A., Zhou, Q., et al. (2002). Prosocial development in early adulthood: A longitudinal study. *Journal of Personality and Social Psychology, 82*(6), 993 – 1006.

Eisenberg, N., Miller, P. A., Shell, R., McNalley, S., & Shea, C. (1991). Prosocial development in adolescence: A longitudinal study. *Developmental Psychology, 27*(5), 849 – 857.

Eisenberg, N., Zhou, Q., & Koller, S. (2001). Brazilian adolescents' prosocial moral judgment and behavior: Relations to sympathy, perspective taking, gender - role orientation, and demographic characteristics. *Child Development, 72*(2), 518 – 534.

Eisenberg–Berg, N., & Hand, M. (1979). The relationship of preschoolers' reasoning about prosocial moral conflicts to prosocial behavior. *Child Development, 50*(2), 356 – 363.

Gasser, L., & Malti, T. (2012). Children's and their friends' moral reasoning Relations with aggressive behavior. *International Journal of Behavioral Development, 36*(5), 358 – 366.

Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology, 101*(2), 366 – 385.

Greene, J. D., Cushman, F. A., Stewart, L. E., Lowenberg, K., Nystrom, L. E., & Cohen, J. D. (2009). Pushing moral buttons: The interaction between personal force and intention in moral judgment. *Cognition, 111*(3), 364 – 371.

Guzman, C. M. (2006). *Moral reasoning and truancy in early adolescence*(Unpublished Doctorial Dissertation), Fordham University, New York.

Haidt, J. (2001). The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychological Review, 108*(4), 814 – 834.

Hardy, S. A. (2006). Identity, reasoning, and emotion: An empirical comparison of three sources of moral motivation. *Motivation and Emotion, 30*(3), 205 – 213.

Hauser, M., Cushmen, F., Young, L, Jin K–X., & Mikhail, J. (2007). A dissociation between moral judgments and

justifications. *Mind & Language, 22*(1), 1 – 21.

Hong, L. (2005). *A study on the relationship among altruistic behavior, empathy and moral judgment of senior middle school students.*（Unpublished Master thesis）, Fujian Normal University.

[洪丽 . (2005). *高中生利他行为与移情，道德判断关系研究*（硕士学位论文）. 福建师范大学 .]

Hunter, J. E., & Schmidt, F. L. (2000). Fixed effects vs. random effects meta - analysis models: Implications for cumulative research knowledge. *International Journal of Selection and Assessment, 8*(4), 275 – 292.

Janssens, J. M., & Dekovi, M. (1997). Child rearing, prosocial moral reasoning, and prosocial behaviour. *International Journal of Behavioral Development, 20*(3), 509 – 527.

Ketefian, S. (1981). Moral reasoning and moral behavior among selected groups of practicing nurses. *Nursing Research, 30*(3), 171 – 176.

Koven, N. S. (2011). Specificity of meta–emotion effects on moral decision–making. *Emotion, 11*(5), 1255 – 1261.

Krebs, D., & Rosenwald, A. (1977). Moral reasoning and moral behavior in conventional adults. *Merrill–Palmer Quarterly of Behavior and Development, 23*(2), 77 – 87.

Kumru, A., Carlo, G., Mestre, M. V., & Samper, P. (2012). Prosocial moral reasoning and prosocial behavior among Turkish and Spanish adolescents. *Social Behavior and Personality, 40*(2), 205 – 214.

Kuppens, S., Laurent, L., Heyvaert, M., & Onghena, P. (2013). Associations between parental psychological control and relational aggression in children and adolescents: A multilevel and sequential meta–analysis. *Developmental Psychology, 49*(9), 1697 – 1712.

Lai, F. H. Y., Siu, A. M. H., Chan, C. C. H., & Shek, D. T. L. (*2012*). Measurement of prosocial reasoning among Chinese adolescents. *The Scientific World Journal, 2012*, Article ID 174845. doi:10.1100/2012/174845

Leffel, G. M. (2008). Who cares? Generativity and the moral emotions, part 2: A "Social Intuitionist Model" of moral motivation. *Journal of Psychology and Theology, 36*(3), 182 – 201.

Li, W. (2012). *Research on the applicability of the internet altruistic behavior scale and the relationship between moral judgment and internet altruistic behavior of post graduates* (Unpublished Master thesis). East China Normal University, Shanghai.

[李炜 . (2012). *硕士研究生网络利他行为的量表适用性及其与道德判断间关系* .（硕士学位论文）. 华东师范大学 , 上海].

Liu, H. S. (2008). Formation of morality. In L. Mo, (Eds.), *Educational Psychology* (pp. 228). Beijing: Educational Science Publishing House.

[刘华山 . (2008). 品德的形成 . 见 莫雷 (编), *教育心理学* (pp. 228). 北京：教育科学出版社].

Liu, M. C. (2012). *Research on the relationship between moral judgment and aggression of middle school students: The effect of moderator of moral disengagement* (Uxnpublished Master thesis), Sichuan Normal University.

[刘美辰 . (2012). *中学生道德判断与攻击行为的关系：道德推脱的调节效应* (硕士学位论文). 四川师范大学 .]

Liu, Z. J. (2001). Research on the relationship between moral judgment ability, peer relationship and prosocial behavior of middle school students. *Psychological Science, 24*(5), 629 – 630.

[刘志军 . (2001). 中学生的道德判断推理水平，同伴关系和亲社会行为关系的研究 . *心理科学，24*(5), 629 – 630].

Maeda, Y., Thoma, S. J., & Bebeau, M. J. (2009). Understanding the relationship between moral judgment development and individual characteristics: The role of educational contexts. *Journal of Educational Psychology, 101*(1), 233 – 247.

Malinowski, C. I., & Smith, C. P. (1985). Moral reasoning and moral conduct: An investigation prompted by Kohlberg's theory. *Journal of Personality and Social Psychology, 49*(4), 1016–1027.

Malti, T., Gasser, L., & Gutzwiller–Helfenfinger, E. (2010). Children's interpretive understanding, moral judgments, and emotion attributions: Relations to social behaviour. *British Journal of Developmental Psychology, 28*(2), 275–292.

Manning, M. A., & Bear, G. G. (2011). Moral reasoning and aggressive behavior: Concurrent and longitudinal relations. *Journal of School Violence, 10*(3), 258–280.

Mao, J. S. (2012). *Research on the relationship among moral judgment ability, guilt and pro–social behaviors of middle school students* (Unpublished Master thesis) , Sichuan Normal University.

[毛静思 . (2012). *中学生道德判断能力，内疚与亲社会行为的关系研究* (硕士学位论文). 四川师范大学].

Miller, P. A., Eisenberg, N., Fabes, R. A., & Shell, R. (1996). Relations of moral reasoning and vicarious emotion to young children's prosocial behavior toward peers and adults. *Developmental Psychology*, 32(2), 210–219.

Mouratidou, K., Barkoukis, V., & Rizos, S. (2012). Achievement goals and moral competence in sport: Examining the moderating role of demographic characteristics. *European Psychologist, 17(1)*, 34–43.

Murphy, F. C., Wilde, G. Ogden, N. Barnard, P. J., & Caldera, A. J. (2009). Assessing the automaticity of moral processing: Efficient coding of moral information during narrative comprehension. *The Quarterly Journal of Experimental Psychology, 62(1)*, 41–49.

Narvaez, D., & Gleason, T. (2007). The relation of moral judgment development and educational experience to recall of moral narratives and expository texts. *The Journal of Genetic Psychology, 168*(3), 251–276.

Palmer, E. J., & Hollin, C. R. (1996). Sociomoral reasoning, perceptions of own parenting and self–reported delinquency. *Personality and Individual Differences, 21*(2), 175–182.

Palmer, E. J., & Hollin, C. R. (2000). The interrelations of socio - moral reasoning, perceptions of own parenting and attributions of intent with self - reported delinquency. *Legal and Criminological Psychology, 5*(2), 201–218.

Palmer, E. J., & Hollin, C. R. (2001). Sociomoral reasoning, perceptions of parenting and self - reported delinquency in adolescents. *Applied Cognitive Psychology, 15*(1), 85–100.

Peng, L. (2004). *The development situation and the relativity research between moral judgment and moral behavior of primary and middle school students* （Unpublished Master thesis）, Yunnan Normal University.

[彭蕾 . (2004). *中小学生道德判断与道德行为的发展现状及二者的相关研究* . 硕士学位论文 . 云南师范大学].

Raaijmakers, Q. A., Engels, R. C., & Van Hoof, A. (2005). Delinquency and moral reasoning in adolescence and young adulthood. *International Journal of Behavioral Development, 29*(3), 247–258.

Rest, J., Cooper, D., Coder, R., Masanz, J., & Anderson, D. (1974). Judging the important issues in moral dilemmas: An objective measure of development. *Developmental Psychology, 10*(4), 491–501.

Rest, J. R., Narvaez, D., Thoma, S. J., & Bebeau, M. J. (1999). DIT2: Devising and testing a revised instrument of moral judgment. *Journal of Educational Psychology, 91*(4), 644–659.

Reynolds, S. J., & Ceranic, T. L. (2007). The effects of moral judgment and moral identity on moral behavior: An empirical examination of the moral

individual. *Journal of Applied Psychology, 92*(6), 1610–1624.

Richards, H. C., Bear, G. G., Stewart, A. L., & Norman, A. D. (1992). Moral reasoning and classroom conduct: Evidence of a curvilinear relationship. *Merrill−Palmer Quarterly, 38*(2), 176–190.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2006). *Publication bias in meta−analysis: Prevention, assessment and adjustments.* Hoboken, NJ: Wiley.

Rubin, K. H., & Schneider, F. W. (1973). The relationship between moral judgment, egocentrism, and altruistic behavior. *Child Development, 44*(3), 661–665.

Ryan, J. J. (2001). Moral reasoning as a determinant of organizational citizenship behaviors: A study in the public accounting profession. *Journal of Business Ethics, 33*(3), 233–244.

Schonert−Reichl, K. A. (1999). Relations of peer acceptance, friendship adjustment, and social behavior to moral reasoning during early adolescence. *The Journal of Early Adolescence, 19*(2), 249–279.

Shumaker, D. M. (1993). *Altruism, empathy, and moral reasoning in high−risk children* (Unpublished Doctorial Dissertation). University of South Carolina.

Simmons, C. H., & Zumpf, C. (1986). The gifted child: Perceived competence, prosocial moral reasoning, and charitable donations. *The Journal of Genetic Psychology, 147*(1), 97–105.

Simmons, R. C. (1986). *Relationship between moral reasoning and participation in and acceptance of library theft−behaviors among undergraduates in a large academic library* (Unpublished Doctorial Dissertation). University of Illinois at Urbana−Champaign.

Tangney, J. P., Stuewig, J., & Mashek, D. J. (2007). Moral emotions and moral behavior. *Annual Review of Psychology, 58,* 345–372.

Tarrant, M., Branscombe, N. R., Warner, R. H., & Weaton, D. (2012). Social identity and perceptions of torture: It's moral when we do it. *Journal of Experimental Social Psychology, 48*(2), 513–518.

Tarry, H., & Emler, N. (2007). Attitudes, values and moral reasoning as predictors of delinquency. *British Journal of Developmental Psychology, 25*(2), 169–183.

Turner, N., Barling, J., Epitropaki, O., Butcher, V., & Milner, C. (2002). Transformational leadership and moral reasoning. *Journal of Applied Psychology, 87*(2), 304–311.

Ugazio, C., Lamm, C., & Singer, T. (2012). The role of emotions for moral judgments depends on the type of emotion and moral scenario. *Emotion, 12*(3), 579–590.

Vandello, J. A., Michniewicz, K. S., & Goldschmied, N. (2011). Moral judgments of the powerless and powerful in violent intergroup conflicts. *Journal of Experimental Social Psychology, 47*(6), 1173–1178.

Wu, H. H. (2005). *A new visual angle of moral research: The theory and empirical research of moral judgment test (Unpublished Master thesis)*, Nanjing Normal University.

[吴慧红 . (2005). *道德研究新视角：道德判断测验的理论和实证研究* (硕士学位论文). 南京师范大学].

Wu, P., Liu, H. S., Lu, L. J., & Tian, M. X. (2013). The relationship between adolescent immoral behaviors on internet and parenting style. *Psychological Science, 36*(2), 372–377.

[吴鹏 , 刘华山 , 鲁路捷 , 田梦潇 . (2013). 青少年网络不道德行为与父母教养方式的关系——道德脱离、责任心、道德同一性的中介作用 . *心理科学* , *36*(2), 372–377].

Wyatt, J. M., & Carlo, G. (2002). What will my parents think? Relations among adolescents' expected parental reactions, prosocial moral reasoning, and prosocial and antisocial behaviors. *Journal of*

Adolescent Research, 17(6), 646–666.

Xu, P., & Chi, Y. K. (2007). A review on the social intuitionist model of moral judgment. *Psychological Science, 30*(2), 403–405.

[徐平 , 迟毓凯 . (2007). 道德判断的社会直觉模型述评 . *心理科学* , *30*(2), 403–405].

Yang, S. G., & Wu, H. H. (2006). Research on the moral judgment abilities of the adolescents. *Psychological Exploration, 26*(2), 55–60.

[杨韶刚 , 吴慧红 . (2006). 青少年道德判断能力的研究 . *心理学探新* , *26*(2), 55–60].

Yu, H. B., & Liu, G. Z. (2006). Advance of research on the relationship between empathy, moral reasoning, perspective taking and prosocial behavior. *Psychological Development and Education, 22*(1),

113–116.

[余宏波 , 刘桂珍 . (2006). 移情 , 道德推理 , 观点采择与亲社会行为关系的研究进展 . *心理发展与教育* , *22*(1), 113–116].

Zhang, H. C., & Xu, J. P. (2004). *Modern Psychology and Education Statistics.* Beijing Normal University Press.

[张厚粲 , 徐建平 . (2004). *现代心理与教育统计学* . 北京 : 北京师范大学出版社]

Zhu, D., & Li, D. (2005). Moral reasoning, empathetic response, and prosocial behavior: Their interrelations in middle school students. *Psychological Science, 28*(5), 1231–1234.

[朱丹 , 李丹 . (2005). 初中学生道德推理 , 移情反应 , 亲社会行为及其相互关系的比较研究 . *心理科学* , *28*(5), 1231–1234].

Association between moral reasoning and moral behavior: A systematic review and meta–analysis

WU Peng[1] ; LIU Huashan[2,3]

([1]Faculty of Education, Hubei University, Wuhan 430062, China)

([2]School of Psychology, Central China Normal University, Wuhan 430079, China)

([3]Key Laboratory of Adolescent Cyberpsychology and Behavior, Ministry of Education, Wuhan 430079, China)

Abstract　According to the theory of Piaget and Kohlberg, whose focuses were the role of moral reasoning in morality, some researchers are inclined to identify moral reasoning as a key factor in predicting moral behavior, moral judgment or moral decision. Over the past decade, some moral psychologists put forward that cognitive factors played a trivial role in morality, while non–cognitive and unconscious factors having great impacts on morality. The controversy of whether moral reasoning can affect moral behavior remained in both theoretical researches and empirical studies. A systematic review was conducted to synthesize empirical results about relationship between moral reasoning and ethical behavior or immoral behavior. Through literature retrieval and selection, in terms of the criteria for inclusion in the meta–analysis, 83 independent effect sizes (50 studies, 16738 participants) were pick out as meta–analysis unit. After coding of data, independent effect sizes were analyzed by CMA 2.2 program. There are four analyses in this research, including heterogeneity test, publication bias test, main effect analysis and moderation effect analysis. In addition, in terms of tentative review analysis and research hypotheses, random effects model was used as meta–analysis model. The test for heterogeneity illustrated that there was significant heterogeneity in 83 independent effect sizes, and also random effects model was a appropriate model for subsequent meta–analysis. The publication bias test indicated that the impact of publication bias was modest but the major finding remained still valid. The research revealed that a positive association between moral reasoning and ethical behavior was found out ($r = 0.238$) and contrarily a negative relationship between moral reasoning and immoral behavior ($r = -0.188$) was disclosed. The moderator analysis revealed that the standardization of moral reasoning measurements affected the relationship between moral reasoning and moral behavior, and additionally participant's age could affect the relationship between moral reasoning and immoral behavior. Specifically, there was a stronger link between moral reasoning and ethical behavior in the process of using standardized instrument, while there being a weaker link between moral reasoning and immoral behavior. Meanwhile, the association between adult moral reasoning and immoral behavior was weaker than adolescent' or children' association. The results suggested that moral reasoning could play an important role in moral behavior (ethical behavior and immoral behavior). Researchers, what's more, are expected to pay much attention to measurement instruments of moral reasoning and standardized instruments in the research of moralities, and especially, the area of ethical behavior. Overall, the findings

provided an evidence to prove moral reasoning being the key factor of morality and suggested that developing moral reasoning measurement in connection with immoral behavior is in urgent need.

"热"执行对注意缺陷多动障碍和阅读障碍儿童言语工作记忆的影响 *

张　微[1]，刘翔平[2]，宋红艳[3]

([1] 华中师范大学心理学院暨湖北省人的发展与心理健康重点实验室 , 武汉 , 430079)

([2] 北京师范大学心理学院 , 北京 , 100875)

([3] 华中师范大学汉口分校心理咨询中心 , 武汉 , 430070)

摘　要　ADHD 儿童在与背外侧前额叶 (DLPFC) 相关的"冷"执行功能上的缺陷已大量证实，但在与眼眶和中前额叶皮层 (OMPFC) 相关的"热"执行功能上，ADHD 儿童是否存在缺陷则未可知。与儿童赌博任务实验范式（该任务中"热"启动对"冷"执行起到抑制作用）不同，本研究考察趣味言语 N–back 任务是否对 ADHD 和阅读障碍儿童的成绩有促进作用。枯燥 N–back 任务考察的是言语工作记忆的纯认知特征，而趣味任务则卷入了"热"执行对"冷"执行的影响。结果表明，在枯燥任务上，ADHD 和阅读障碍儿童的成绩均明显低于正常儿童，二者之间差异不显著，在趣味任务成绩上，ADHD 儿童与正常儿童的成绩没有显著差异，而阅读障碍儿童成绩落后于正常控制组和 ADHD 组，在成绩变化的趋势上，ADHD 儿童在趣味任务上成绩明显提高，而阅读障碍儿童则无明显改善。结果说明 ADHD 和阅读障碍儿童的言语工作记忆均存在明显的缺陷，但是机制不同，"热"执行对提高 ADHD 的言语工作记忆有明显的促进作用，ADHD 的"冷"执行缺陷能够通过"热"执行的调节得到改善。

关键词　注意缺陷多动障碍；阅读障碍；"热"执行；言语工作记忆

分类号　B844

1 前言

注意缺陷多动障碍（Attention Deficit Hyperactivity Disorder, ADHD）是一种临床常见的儿童行为问题，核心症状是注意缺陷、多动和冲动 (APA,1994)。通常发病于学龄前期，在小学阶段，症状表现尤为明显，随着年龄的增长，症状逐渐得到改善或者消失，国内调查显示，ADHD 的检出率在 3.1%–6.3% 之间 (张微，刘翔平，廖冉，顾群，2007)。

以 Wilding（2005）为代表的研究者认为，执行功能（executive function, EF）缺陷是 ADHD 的核心缺陷，能够解释 ADHD 症状的全部特征，Wilding 对不少研究进行分析总结后，将持续性注意任务上的准确性差异归结于 ADHD 儿童的执行功能的基本缺损，执行功能的缺陷会影响 ADHD 儿童反应准备，反应

速度，刺激—反应组织，焦点的保持，目标特征的组合，目标的定义甚至更多的过程。

考虑到在前额叶皮层功能的差异，Zelazo 和 Müller（2002）区分了两类执行功能，与背外侧前额叶 (DLPFC) 相关的"冷"执行功能，更多涉及的是 EF 的纯认知方面，眼眶和中前额叶皮层 (orbital and medial prefrontal cortex，OMPFC）则与"热"执行功能有关，该区分得到了 Habib, Boulanger, Soubias, Delarbre 和 Joly-Pottuz（2003）的进一步确认。冷认知相对抽象，去情境化，目前多数用来测试 ADHD 执行功能的任务，如 Stroop, flanker, Go/No-Go, Stop, 持续性操作任务和工作记忆任务，均属此类。而在"热"执行的研究情境中，需要进行可能的冒险决策，或者需要对刺激的情感意义进行灵活性的估价，卷入情感和动机。Zelazo 和 Muller（2002）认为 ADHD 是一种"冷"执行功能障碍，而 Castellanos, Sonuga-Barke, 和 Tannock（2006）则认为，注意缺陷症状与"冷"执行功能有关，多动冲动症状与"热"执行有关，一些 ADHD 个体主要是"热"执行功能缺损，另一些 ADHD 个体的缺陷则主要表现在"冷"执行功能上，还有一些则表现出双重缺陷。

暂不论 ADHD 儿童的执行缺陷是以"热"执行还是"冷"执行缺陷为主，但"热"执行对"冷"执行加工能起到调节作用，且是以特定的神经回路为基础的。根据 Haber（2000）模型，皮质—纹状体—丘脑—皮质螺旋回路（spiraling cortico-striato-thalamo-cortical circuits）担当此任。此复杂的，非交互的通路包括纹状体—黑质—纹状体（SNS）和丘脑—皮质—丘脑网络。这是通过与 OMPFC 通路相关的情绪及动机影响"冷"执行的 DLPFC 通路的解剖学基础，而 DLPFC 接着影响了运动区。具体是，OMPFC 投射到壳区（shell），DLPFC 主要投射到中部纹状体，前运动皮层和运动皮层主要投射到背外侧纹状体。壳区投射到中脑两个区域：腹侧被盖区（ventral tegmental area，VTA）和黑质致密部（substantia nigra, pars compacta，SNc）。VTA 投射到壳区，形成 SNS 回路。这种通过这些螺旋回路的非交互成分形成的单向信

息流通道的本质特征表明了动机和情感对认知加工的影响的层次性。认知加工能够调控动作输出并且与 ADHD 等相关的障碍明显有关系。

另有研究者 Sonuga-Barke（2002）虽然未明确提出 ADHD 在"冷""热"执行机制上的表现差异，但其双通道理论也强调了 ADHD 在奖励相关的动机加工上的缺陷，认为 ADHD 抑制缺陷（EDF）与厌恶延迟（DEL）遵循不同的神经通路，ADHD 是 EDF 和 DEL 双通道作用的结果。一方面，前额叶问题导致抑制缺陷，另一方面，与奖励环路（reward circuit）中央 - 边缘多巴胺分支（meso-limbic dopamine branch）相关的延迟满足能力的改变促成了 ADHD 的特殊的动机风格。ADHD 既表现出执行功能障碍，也表现出厌恶延迟满足的动机风格。枯燥的任务测查的是"冷"认知功能，而能够立刻得到满足的任务则与"热"认知功能相关，ADHD"冷"认知任务上存在缺陷，是因为其在前额叶多巴胺系统的神经通路上出现障碍，在"热"认知任务上，由于奖励回路被激活，ADHD 儿童能够通过情绪和动机的激活，提高任务卷入程度，改善了其兴奋、抑制等认知功能，从而弥补了其在 EDF 神经通路上的缺陷。也就是说，ADHD 儿童在"冷"认知任务上缺陷严重，但在"热"认知任务上的则有可能表现正常。

基于背外侧前额叶、扣带回以及纹状体等为基础的以抑制缺陷为代表的"冷"执行缺陷在 ADHD 身上已经被毫无争议地证实。但是，对 ADHD 儿童在"热"执行任务上的关注并不多，也没有形成明确的结论。来自强化依随（reinforcement contingencies）范式的总体研究结果表明，奖励和反应代价在 ADHD 和控制组的成绩和动机水平上有积极的效果，但 ADHD 儿童在奖励和反应代价条件下成绩比无强化依随条件下成绩提高的程度要比控制组大；另外，与控制组相比，ADHD 更多选择即刻满足（Luman, Oosterlaan, & Sergean, 2005）。但是来自生理记录方面的研究却发现，ADHD 在心理生理上（皮肤电和心律指标）对强化依随的敏感性低于控制组 (Crone, Jennings, & Van der Molen, 2003)。来自儿童赌博任

务的研究发现 (Habib, Boulanger, Soubias, Delarbre, & Joly-Pottuz, 2008)，ADHD 儿童的在此任务上的成绩落后于正常控制组，他们并不偏爱少的有利的选择，且 ADHD 儿童在此任务上的失败与 Stroop 测验上的成绩没有明显相关，Habib 等人认为 ADHD 在奖励机制上存在缺陷。国内朱昭红 (2006) 的研究发现，在赌博任务中儿童更倾向于不利选择，两种亚型儿童在赌博任务上的选择模式相似，均倾向于不利选择，表明儿童在情感性决策中的缺损源于对奖励的高敏感性，并且可能与 OFC 损伤者有同样的机制。

据上所述，尽管 Zelazo 和 Muller 认为 ADHD 可以看成是"冷"执行功能障碍，但是仍然有研究发现了 ADHD 在"热"执行上的缺陷，这有任务范式的原因，或可以假定在不同性质的实验任务中，"热"执行所发挥的协调作用并不相同，甚至截然相反。凡是情境化的，能够很好地卷入情感动机的任务，均涉及到"热"执行和"冷"执行的参与，但是情绪的高度卷入和动机的加强是否一定对有效地执行任务起到促进作用呢？在强化依随的任务中，每当个体作出有效的反应，都有可能得到积极的奖励，追求奖励的动机与能否有效地作出反应并不存在冲突，是一致的，为了获取高奖励，个体需要坚持有效地进行反应，动机越高，坚持性越强，成绩也越好，例如根据个体对计算机屏幕中出现信号的注视时间越久，就能得到越多的奖励。研究发现，奖励对 ADHD 儿童自我评定的动机水平有积极的效果，强化依随对 ADHD 儿童的动机促进作用较正常儿童更为明显 (Carlson, Mann, & Alexander, 2000；McInerny, & Kerns, 2003； Scheres, Oosterlaan, & Sergeant)。在儿童赌博任务中，研究者鼓励儿童尽量多赢得奖励，与反向择物任务相似，表面上的高奖励并不一定意味着高奖励，可能带来更高的损失，但是被试在每次试验中，首先感受到的是所作选择的有利面（例如笑脸，代表奖励），然后才有可能出现不利面（打开被遮盖的纸牌的下半区，可能露出数量不等的哭脸，意味着损失），此时涉及到一个对优势心理表征进行有效抑制的心理过程，动机和情绪的高度卷入

构成了不利因素，它会使不利的优势表征过度加强，从而提高了有效抑制的难度，对最终的任务结果产生负面影响。此时，动机和情感的过度卷入阻碍了任务的执行，由此带来的过"热"的冲动反应需要被有效抑制，这显然是与前面强化伴随的实验任务不同。

然而，遗憾的是有一类"热"执行任务没有引起重视。这类任务中，任务本身就很有趣，完成任务本身就是奖励，不刻意需要额外的人为的强化施加，个体的动机和情绪就得到加强，被试对任务本身的喜爱要甚于对完成任务的结果的奖励趋向，任务本身的积极特性激发了个体的动机和情感卷入，同时高度的动机和情感卷入并不阻碍任务本身的有效执行。此类任务可以认为是纯趣味任务。与强化依随任务相比，趣味任务所激发的动机指向于完成任务的过程，而强化依随任务本身是无趣的，动机指向于结果，并且要求个体人为地将反应和强化建立联系，当条件发生变化时，需要中断旧的反应-强化联系，建立新的反应-强化联系。根据刺激强化值的灵活表征假设 (Rolls，2000)，眶额叶（OFC）卷入了根据情感或动机对刺激的再估价，以及对刺激与强化刺激联系的分离，OFC 的损伤则导致了刺激-奖励转换能力的损伤。但如果 OFC 存在缺陷，可能不会不会影响儿童在趣味任务上的表现，这是因为在此类趣味任务中，刺激奖励之间的联系并不重要，也不涉及到对刺激强化值的灵活表征和刺激-奖励转换。研究发现，ADHD 儿童在那些比较有意思的游戏任务中的工作记忆并不存在明显的缺陷，但是，在那些枯燥的记忆任务中，缺陷比较明显，研究中，任务是需要对连续的空间位置进行记忆，结果 ADHD 青少年却不存在明显缺损 (Barkley, Edwards, Laneri, , Fletcher, & Metevi, 2001)。

阅读障碍与 ADHD 关系极为密切，伴随发生率很高，语音加工缺陷被广泛认为是阅读障碍的核心问题，而阅读障碍的执行功能缺陷则并没有得到广泛认同。两种障碍在特定的神经心理加工缺陷上既有相似性又有差异，它们在"冷"言语工作记忆任务上的缺陷已得到大量研究的证实（Bowers, Steffey, & Tate,

1988；Xavier, Castellanos, & Tannock,2002）。但在"热"言语工作记忆任务上是否存在缺损则不甚明了。"热"执行对于儿童阅读障碍的言语工作记忆的成绩是否有影响呢？两种障碍是否表现出同样的模式呢？探索这些问题将有利于解释 ADHD 和阅读障碍言语工作记忆缺陷的核心问题，以及能够说明"热"执行对于"冷"执行的影响机制。

因此，本研究以阅读障碍被试和正常被试作为参照，采用 N-back 范式的两类性质言语工作记忆任务（趣味任务和枯燥任务）来要考察 ADHD 儿童的表现，以探索"热"执行对于 ADHD 和阅读障碍儿童言语工作记忆的影响。

2 方法

2.1 被试

共有 86 名来自某学习障碍儿童咨询矫治机构的障碍被试和 25 名来自学校样本的正常控制组被试参加了研究。ADHD 儿童最近半年内未服用过利他林等精神类药物治疗，无明显的器质性损伤，不伴随明显的品行障碍、情绪障碍、语言障碍和智力缺陷，通过 DSM–IV，Conners 家长和教师量表，自编 ADHD 问卷和 IVA–CPT 视听整合持续性操作测验来评估选择 ADHD 儿童，以瑞文标准推理测验的成绩作为智力衡量标准，排除得分在 25% 以下和 90% 以上儿童。

从北京市一所城区小学选取正常控制组被试 20 名，年龄、性别、智力以及教育水平和教育环境都与实验组匹配。发展性阅读障碍的选择主要工具为《小学生识字量测验》，以智力与识字水平的不匹配作为选择依据。

ADHD 儿童入选标准及甄选程序：在甄别 ADHD 时，我们对家长和教师（班主任）进行了访谈，访谈的工具有 DSM–IV，Conners 问卷，自编 ADHD 问卷（包括 ADHD 症状以及一些排除 ADHD 诊断的可能情况），另外，我们对儿童施测了 IVA — CPT 测验。诊断标准主要以 DSM–IV 家长评估为依据。在此量表上，至少有一个维度得分在 6 分或 6 分以上，

在 IVA–CPT 上被诊断为 ADHD（少数被试在 DSM–IV 家长评定上接近可诊断水平，CPT 得分较低，也可被视为 ADHD），同时，考虑到教师评定尺度较宽松，教师评定的 DSM–IV 得分不一定要求在 6 分或 6 分以上，但是需与家长评定一致，且相应维度得分不能在 3 分以下。另外，以 Conners 家长和教师评定量表以及自编的 ADHD 问卷作为辅助工具，凡符合或者边缘符合 DSM–IV 的 ADHD 诊断的儿童在 Conners 量表上得分较高（10 分或 10 分以上），且符合自编 ADHD 问卷的多数症状标准，可进一步确认，少数在不同诊断标准上有明显矛盾，且不能为访谈所证实为 ADHD 的儿童，将被排除在研究之外。正常儿童的取样来自于一所普通小学，其家长评定的 DSM–IV 两个维度得分均在 3 分以下，否则被排除。

确定阅读障碍的标准：我们采用了两个工具，分别是《小学生识字量测验》（上海教育出版社，信效度均为 0.98）和《标准化阅读理解测验》（本测验分低中高三个年级版本，分别适用于一二年级、三四年级和五六年级。项目构成如下：低年级阅读理解测验由 4 篇 600 左右的小短文及其与短文内容相关的问题构成，其中 3 篇记叙文，一篇说明文，共 19 个问题，均为四选一的选择题；中年级测验由 4 篇 800 字左右的短文和相应的题目构成，其中议论文 1 篇，说明文 1 篇，记叙文 1 篇，散文 1 篇，每篇各附 5 个选择题；高年级阅读理解测验由 5 篇约 1000 字左右的短文组成，其中说明文 2 篇，议论文 2 篇，记叙文一篇，各附 4 道选择题。低年级、中年级和高年级测验的分半信度分别为 0.83、0.82 和 0.86，效标效度分别为 0.71、0.87 和 0.77（通过计算阅读理解测验成绩与教师采用五点量表对学生阅读理解能力的主观评估的相关获得），信效度均符合要求。），纳入标准：识字量落后 0.5 年级，且阅读理解成绩在 50% 以下。非阅读障碍被试识字量必须在 –0.2 年级以上，且阅读理解成绩必须在 25% 以上。（识字量为主要的筛选标准，阅读理解测验仅为参考）

最后确定有效被试共 77 人，其中正常被试 17 人，ADHD 不伴随阅读障碍被试 21 名，ADHD 伴随阅读

障碍被试 24 名，阅读障碍（不伴随 ADHD）被试 15 名，年龄在 7 到 12 岁之间的二到五年级儿童，男 56 人，女 21 人，平均年龄 8.5 岁。障碍被试均来自临床样本，而正常被试均来自某小学。四组被试年龄、智力相匹配，识字量存在明显差异，见表 1：

从上表可以看出，四组被试的年龄和智力水平没有差异，但在识字量上存在明显差异。事后检验结果表明，识字水平上，控制组 =ADHD，ADHD>ADHD+RD，ADHD+RD=RD（前后两两比较，

$P=0.315$，$P<0.001$，$P=0.695$）。这说明被试分组是合理的。

2.2 实验材料与设计

程序采用 AB 6.0 中文版编写。实验包括两个独立的任务：枯燥任务和趣味任务。两个任务的材料、难度、任务要求完全一致，只是在趣味性上有很大差异。为防止练习效应，枯燥任务和趣味任务并不在一个时间段完成（相隔 1 周）。且在完成任务的顺序上，进行了平衡。实验材料均为大写字母。

表 1　四组被试瑞文成绩、识字量和年龄的比较

统计项	控制组（17）	ADHD（21）	ADHD+RD（24）	RD（15）	F	P
年龄（岁）	8.6（1.50）	8.42（1.01）	8.38（1.01）	8.87（1.30）	0.88	0.457
瑞文	91.81（8.22）	73.92（28.02）	79.89（16.21）	70.83（24.47）	2.29	0.087
识字量（相对年级）	0.61（.68）	0.37（.62）	−0.81（.48）	−0.73（.46）	25.53	0.000

注：识字量为相对年级水平，正表示在正常水平之上，负表示落后于正常水平。ADHD：儿童注意缺陷多动障碍不伴随阅读障碍，ADHD+RD：儿童注意缺陷多动障碍伴随阅读障碍，RD：阅读障碍不伴随儿童注意缺陷多动障碍。下同。

考虑到被试的特殊性，我们对经典 N-back 任务进行了改进，降低了难度，与 Kiss, Pisio, Francois 和 Schoflocher (1998) 改进后的 N-back 范式相似，有良好的效度。枯燥任务中所有 trial 形式相同，所有字母字体、颜色相同，没有反馈，没有积分，完成任务没有奖励，背景和前景色单调。趣味任务的趣味性明显增加。登陆时有非常漂亮的登录界面，并且伴随非常舒缓动听的背景音乐。登录后进入指导语界面，在非常漂亮的背景上呈现字体美观的指导语。指导语暗示被试任务将是一个有趣的小游戏。为了让被试明白指导语要求和奖励情况，主试需耐心给被试讲解。点击"进入游戏"按钮后，进入练习任务窗口，背景音乐消失。呈现大写字母序列，不同的 trial 的大写字母序列字体、颜色、大小各异，同一个 trial 内的字母大写、颜色、字体相同。练习共 6 题，至少答对 5 道题方可进入正式实验。正式实验中，屏幕背景色为红色，字母的颜色与屏幕背景色有区分。完成正式实验后，会有一个最后总分提示的窗口。实验结束后，根据被试的得分挑选奖品。具体任务操作见实验程序部分。

趣味游戏实验材料为大写字母，同一个 trial 中，目标刺激不能有相同的字母。共计 60 个 trial，其中

长度为 4 个字母、5 个字母和 6 个字母的 trial 各 20 次，正负反应各半，随机分布。共有五种字体，分别是隶书，幼圆，Aribal，方正姚体，方正舒体，字体大小也有五种。根据情况挑选了 15 中字体类型、大小和颜色组合，即 trial 中字母的外观上有 15 种可能，均出现 4 次。

趣味任务中，被试对每个 trial 作出反应之后都有反馈，每次反馈的内容都不相同。用于反馈的材料为动画、声音、文字。动画为 72 项大小为 150mm × 150mm 的动画，均取材于数码宝贝、三国等儿童熟悉的游戏动画人物，随机出现；声音 72 项，长度在 3 秒左右，均是非常动听的音乐；文字有 180 项，分为答对了的反馈、答对了且很快的反馈、答错了的反馈，例如："真是又快又准！加 15 分！"，这些反馈内容随机出现在各个 trial 的对应的反应条件下。反馈窗口持续时间为 5s。练习实验同正式实验形式相同。各种类型的 trial 随机出现在实验中。只有在正确回答的条件下，才会有各种各样的动画和音乐的反馈，在答错的条件下，每次反馈的内容都相同，即一张苦脸，一个"哦哟"的声音和文字鼓励"不要灰心！"，这样做的目的是要让被试努力表现好才能看到各种各样的动画和听到美妙的音乐。

实验为一对一进行，主试坐在被试身边约 30 厘米处，斜对着屏幕，随时监控实验的进行。被试正坐在计算机屏幕前，眼睛与显示器中央成 15 度水平视角，事先熟悉键盘和鼠标。被试戴上立体声端坐于屏幕前，显示器分辨率为 1024×768。实验步骤如下：每个实验任务都是先出现注视点（屏幕中心的一个蓝色的"+"字）300ms，消失，然后呈现空屏 300ms，呈现字母序列的第一个字母 500ms，消失后出现空屏 500ms，随后呈现第二个字母 500ms，消失后再次呈现空屏 500ms，呈现第三个字母 500ms，以此类推，直到呈现最后字母序列的最后一个字母 500ms 和空屏 500ms，要求被试判断最后一个字母与倒数第三个字母是否相同，并按键反应，正反应按"P"键，负反应按"Q"键，探测字母消失，在枯燥任务中，trial 结束，之后是 800ms 的空屏刺激准备时间，然后进入下一个 trial；在趣味任务中，被试反应后，呈现反馈窗口，持续 5000ms，然后自动跳入下一个 trial。系统自动记录反应时和反应正确与否。如果被试在探测刺激出现后的 10s 内未能做出有效反应，将进入下一个 trial。枯燥任务大约持续 15 分钟，趣味任务大约持续 25 分钟。

3 结果

考虑到两类任务操作上的差异，比较反应时的差异意义不大，本研究中只是比较四组被试在各个水平上的正确率的差异。剔除反应时低于 200ms 和反应时明显过长（大于 3 个标准差）的数据，四组被试在不同性质任务下的正确率以及四组被试在枯燥和趣味记忆任务上成绩的变化的差异（趣味任务成绩减去枯燥任务成绩）见图 1 和图 2。

进行重复测量方差分析，枯燥任务，趣味任务，趣味任务和枯燥任务成绩的差异的组间效应均达到了显著性水平（$F_{(3,73)}$=15.67，$p<0.01$；$F_{(3,73)}$=3.76，$p<0.05$；$F_{(3,73)}$=3.42，$p<0.05$）。进一步进行多重比较后发现，在枯燥记忆任务上，控制组的成绩（82.6%）明显高于 ADHD（70.7%）($p<0.01$)、
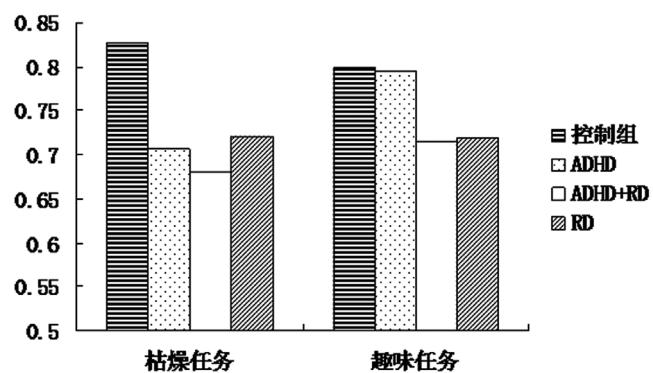
图 1 四组被试在枯燥和趣味言语工作记忆任务上的成绩（正确率）
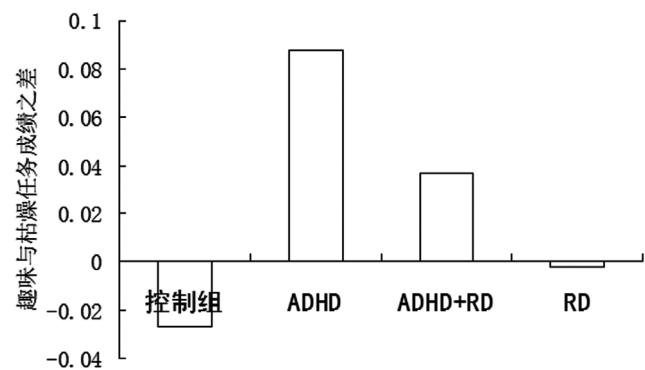
图 2 四组被试趣味与枯燥任务成绩（正确率）之差（趣味 − 枯燥）

ADHD 伴随阅读障碍（67.9%）($p<0.01$) 和阅读障碍组（72.1%）($p<0.01$)，虽然 ADHD 伴随阅读障碍成绩最低，但是差异检验发现，ADHD、ADHD 伴随阅读障碍和阅读障碍三组的成绩差异并不显著 (ADHD，ADHD+RD：$p=0.565$；ADHD+RD，RD：$p=0.952$；ADHD,RD: $p=0.560$)。在趣味记忆任务上，四组被试的成绩依次为 79.9%, 79.5%, 71.5%, 71.9%, 结果是：正常控制组 =ADHD,($p=0.912$);ADHD>ADHD+RD,($p<0.05$); ADHD+RD =RD($p=0.917$)。在趣味任务与枯燥任务的差异（趣味记忆任务的成绩减去枯燥记忆任务的成绩）上，正常控制组儿童由枯燥任务到趣味任务，成绩下降了 2.7%，ADHD 儿童提高了 8.8%，ADHD 伴随阅读障碍儿童提高了 3.7%，阅读障碍儿童下降了 0.2%，ADHD 儿童的成绩显著得到了提高 ($p<0.01$)，增幅最大（ADHD，ADHD+RD：$p=0.1$），其次是 ADHD 伴随阅读障碍儿童，成绩相

比正常组也有提高 (ADHD+RD,RD: $p<0.1$)，阅读障碍儿童在趣味记忆刷新任务上的成绩没有显著变化 ($p=0.753$)，反观正常儿童，成绩不但没有提高，反而下降 ($p<0.1$)。关于四组被试在枯燥趣味任务上的成绩变化趋势上，我们控制反应时后进行了方差分析得到了相同的结果 $F（54，21）=4.079$，$p<0.01$。

从以上结果可以看出，相比"冷"工作记忆任务，"热"执行明显提高了 ADHD 儿童的成绩，但是对于阅读障碍和正常儿童的成绩影响不大。

4　讨论

本研究对比了 ADHD 和阅读障碍儿童在同等难度的趣味和枯燥工作记忆言语 N-back 任务上的表现，该任务范式被认为既很好的涉及到语音回路的功能，同时又卷入了中央执行（处理前摄抑制和记忆更新操作）的参与，结果发现 ADHD 儿童在趣味工作记忆任务上的成绩有明显提高，相反，阅读障碍儿童和正常儿童的成绩则几乎没有变化，甚至有小幅下降。这与来自儿童赌博任务的研究得到的 ADHD 儿童在"热"执行上的缺陷结果不同，本研究结果说明：1）ADHD 和阅读障碍儿童都表现出言语工作记忆能力的缺陷，但是缺损模式不同；2）ADHD 伴随阅读障碍儿童在"热"执行上成绩一定程度的补偿说明了 ADHD 伴随阅读障碍的缺陷既有 ADHD 的特征又有阅读障碍的特征；3）在趣味任务上，ADHD 的"热"执行功能很好的发挥了作用，ADHD 可能并不存在特异性的"热"执行缺陷。这与 Barkely（2001）的研究一致，Carlson, Booth, Misung 和 Canu (2002) 的研究也证实了 ADHD 儿童更多通过外部的反馈来评量他们行为的水平，表现出更低的掌握目标定向水平，内部动机缺失，成绩不稳定，容易受到趣味性奖励的影响。

可以认为"热"执行是否能够很好的发挥作用关键在于依据特定的任务情境。趣味任务中"热"执行和"冷"执行共同参与，"热"执行对"冷"执行起到一定的协调作用。与儿童赌博任务不同，

赌博任务中，过度的情感反应容易导致冲动的不利判断，虽然提高了动机，但是更需要加以抑制，进行冷却再加工，而相反，趣味任务卷入的"热"执行不仅提高了动机水平，而且对任务相关的"冷"执行功能起到促进作用，例如抑制能力和认知资源的利用效率的提高，这些都有利于任务执行，因此，对于"热"的情感反应，无须刻意对抗。然而，这种"热"执行的激发作用对 ADHD 的效果明显，而对阅读障碍和正常儿童没有明显的促进作用，这可能与不同的障碍的缺损模式的差异以及"热"执行的作用机制有关。对于正常儿童，在枯燥任务时，正常儿童成绩已经很高，其成绩并没有像 ADHD 一样，在枯燥任务时受动机缺失的影响而受损，故面临趣味任务时，虽然提高了动机水平，但是并不明显提高成绩；另外，与障碍儿童相比，正常儿童的成绩已经处于一个相对较高的水平，再提高可能有瓶颈效应；再次，趣味任务过多的声音变化、trial 变化和动画变化对正常儿童而言除了提高动机，也形成了干扰刺激，从而影响了成绩，尽管同样也对 ADHD 儿童造成了干扰，但后者受到动机和情绪的激发效果是最主要的。对于阅读障碍儿童，成绩在趣味任务条件下没有明显提高可能受到了任务本身的干扰外，另一个可能的原因是动机的加强对于阅读障碍儿童的言语工作记忆功能并没有实质性的促进作用，依据 Haber 模型，通过皮质—纹状体—丘脑—皮质螺旋回路，与"热"执行相关的 OMPFC 区调节了与"冷"执行相关的 DLPFC 区，并进一步调节了动作输出，即动机和情绪的积极卷入提高了执行功能，从而改善了任务操作。在记忆刷新操作过程中，主要的执行操作是与更新操作相关的处理前摄抑制作用，之前的研究发现 (张微，2008)，ADHD 儿童的处理前摄抑制能力较差，并导致了记忆刷新成绩的下降，而阅读障碍儿童并不存在这个问题，后者主要的缺陷表现在更为基础的语音回路的编码和复述功能上，而热工作记忆任务并不能改善阅读障碍儿童的语音编码和复述能力，故没有明显提高其成绩。

从另一个角度，Sonuga-Barke（2002）等人强调

了ADHD特殊的动机风格以及对强化的过度敏感性，本研究中两类任务的差异在于趣味任务能够获得即时奖励和满足，枯燥任务则不能。Sonuga-Barke认为厌恶延迟，强烈的追求即时满足的倾向是ADHD儿童的动机风格，并提出了厌恶延迟假说，认为ADHD是潜在的动机风格的功能特征而非自我调节系统失能。ADHD儿童被推动去逃避或者避免延迟。他们的注意缺陷、多动、冲动可以被认为是厌恶延迟的功能表现。当面临即时满足和延迟满足，ADHD会好不犹豫选择立刻满足。当没有选择时，他们会为了在延迟中减少时间知觉，与环境发生作用，创造或者参与非当时的环境事件。认知缺损被假定为由于厌恶延迟所导致的二级缺损。特殊的动机风格的基础是奖励通路。相反，在那些能够得到即时满足的情景中，奖励回路的激活，伴随情感动机的卷入，ADHD儿童的认知功能的缺陷得到了补偿。类似的，Douglas和Parry（1994）曾认为ADHD核心问题之一是他们是否对奖励具有异常的敏感性。Seageant（2000）的认知能量模型认为，ADHD儿童可能在不同的认知水平上表现出不同程度的认知过程缺陷，认知的加工机制中编码、中央加工和运动反应三个阶段与能量机制唤醒、努力和激活三种"能量库"联系紧密，同时这些认知加工过程和状态因素受到一个更高级的执行功能控制系统(也称管理评价机制)的监督和调控，ADHD儿童主要缺陷是与认知能量模型的高级管理机制相联系的三种能量库，但是动机的提高可以提高认知能量水平，从而影响了执行功能。同样的，Philip，Vaughan，Terry，Sally，Diane，和Belinda（1999）也认为ADHD儿童的成绩较容易受到动机和唤醒的影响，在有奖励的情况下，他们在那些需要意志努力的控制加工任务的表现会明显得到提高。

虽然研究者从不同的角度对ADHD在卷入热认知功能的趣味任务中的成绩的提高做出了解释，但有一点是共同的，即ADHD儿童在趣味任务中，都毫无疑问的表现出的强烈的情绪和动机卷入，"热"执行对"冷"执行进行了积极的调节。

参考文献

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (4th ed.)*.Washington, DC: American Psychiatric Press.

Barkley, R.W., Edwards, G., Laneri, M., Fletcher, K &Metevi, L.(2001). Executive functioning, temporal discounting, and sense of time in adolescents with Attention Deficit Hyperactivity Disorder (ADHD) and Oppositional Defiant Disorder (ODD). *Journal of Abnormal Child Psychology, 29*, 541–556.

Bowers, P., Steffey, R., & Tate, E.(1988). Comparison of the effects of IQ control methods on memory and naming speed predictors of reading disability. *Reading Research Quarterly, 23*,304-309

Carlson ,C., Booth, J. E., Misung, S., & Canu. W.H. (2002). Parent-teacher-,and self-related motivational styles in ADHD subtypes. *Journal of Learning Disabilities,35*,2 104-114.

Carlson, C., Mann, M., & Alexander, D. K.(2000) Effects of reward and response cost on the performance and motivation of children with AD/HD. *Cognitive Therapy and Research, 24*, 87–98.

Castellanos, E., Sonuga-Barke, E.J., & Tannock,M.R. (2006).Characterizing cognition in ADHD: beyond executive dysfunction. *TRENDS in Cognitive Sciences,10*（3）：117-124

Crone, E., Jennings, J. R., & Van der Molen, M. W.(2003). Sensitivity to interference and response contingencies in attention-deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry, 44*, 224–226.

Douglas, V., & Parry, P.(1994). Effects of reward and nonreward on frustration and attention in attention deficit disorder. *Journal of Abnormal Child Psychology, 22*, 281–302.

Haber, S.(2000). Striatonigrostriatal pathways in primates

form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience, 20,* 2369‒2382

Haber, S.(2003). The primate basal ganglia: parallel and integrative networks. *Journal of Chemical Neuroanatomy, 26,*317‒330

Habib, M., Boulanger, C., Soubias, M., Delarbre, C., & Joly‒Pottuz, B. (2008). The neuropsychology of the human reward system: Impaired gambling performance in ADHD children and adults with psychopathic tendencies . *Brain and Cognition, 67.* 11‒47

Kiss, I., Pisio, C., Francois, A.,& Schopflocher, D.(1998). Central executive function in working memory: event‒related brain potential studies. C*ognitive Brain Research*，6，235‒247

Luman, M., Oosterlaan, J.,& Sergean, J. (2005). The impact of reinforcement contingencies on AD/HD: A review and theoretical appraisal. *Clinical Psychology Review, 25,*183‒213

McInerny, R.,& Kerns, K.(2003). Time reproduction in children with ADHD: Motivation matters. *Child Neuropsychology, 9,* 91‒108.

Philip, H. L., Vaughan, C. J., Terry, L. J., Sally, D. A.., Diane, H. M., & Belinda, B. M. (1999). Effortful and automatic information processing in boys with ADHD and specific learning disorders. *Journal of Child Psychology and Psychiaty,40,*275‒286

Rolls E.(2000). The orbitofrontal cortex and reward. *Cerebral Cortex., 10* :284‒294.

Scheres, A., Oosterlaan, J.,& Sergeant, J.(2001). Response inhibition in children with DSM‒IV subtypes of AD/HD and related disruptive disorders: The role of reward. *Child Neuropsychology, 7,* 172‒189.

Sergeant, J.(2000). The cognitive‒energetic model: An empirical approach to attention‒deficit hyperactivity.

*Neuroscience Biobehavioral Review, 24:*7‒12.

Sonuga‒Barke, E.J. (2002). Psychological heterogeneity in AD/HD:a dual pathway model of behaviour and cognition. *Behavioural Brain Research, 130,* 29‒36

Wilding, J. (2005).Is attention impaired in ADHD? British *Journal of Developmental Psychology, 23*(4):487‒505

Xavier, F., Castellanos, E., & Tannock. R. (2002). Neuroscience of attention deficit/Hyperactivity disorder: The search for endophenotypes. *Nature reviews. Neuroscience ,3，*617‒628

Zelazo, P., & M ü ller U. （2002）.Executive function in typical and atypical development. Hand book of childhood cognitive development. *Oxford; Blackwell.*

Zhang, W. (2008). *The unique deficit of working memory of ADHD: the evidence come from the comparison of ADHD and RD in memory updating (in Chinse).* Unpublished Doctoral dissertation. Beijing Normal University.

[张微 .(2008). *注意缺损多动障碍（ADHD）儿童的特异性工作记忆缺陷：来自 ADHD 和阅读障碍的记忆刷新比较研究证据* . 博士学位论文 , 北京师范大学].

Zhang, W. ,Liu, X., Liao, R. ,& Gu, Q.（2007）. An epidemiological investigation of ADHD in six cities (in Chinse). *Chinese journal of clinical psychology. 1，* 23‒25.

[张 微， 刘翔平， 廖 冉， 顾 群 .（2007）. 六 城 市 ADHD 流行病学调查，*中国临床心理学杂志，1，* 23‒25].

Zhu, Z. (2006). *The effect of reward and punishment on affective decision‒making on two subtypes of children with attention deficit hyperactivity disorder (ADHD)* . Unpublished Master thesis Shanxi Normal University.

[朱昭红 (2006). 奖惩对两种亚型儿童情感性决策的影响 . 硕士学位论文，陕西师范大学].

The influence of "Hot" executive function on the verbal working memory of attention deficit hyperactivity disorder and reading disability children

ZHANG Wei[1]; LIU Xiangping[2]; SONG Hongyan[3]

([1] School of Psychology, Huazhong Normal University , Hubei Province Key Lab for Human Development and Mental Health, Wuhan,

430079,China)

([2] School of Psychology, Beijing Normal University, Beijing, 100875, China)

([3] Center of Psychological counseling, Hankou Branch of Huazhong Normal University, Wuhan, 430000,China)

Abstract The deficit of "cool" executive function (EF) associated with dorsolateral prefrontal cortex (DLPFC) in attention deficit hyperactivity disorder (ADHD) has been substantially confirmed. But whether ADHD children show the deficit of "hot" EF associated with orbital and medial prefrontal cortex (OMPFC) remains unknown, and till now no research made an explicit exploration for interaction models between the two EFs. Commonly different from some studies related to the children's gambling task, in which the "hot" EF impeded "cool" EF, this study aims to explore the facilitation of the "hot" EF to "cold" EF in the entertaining verbal N−back task. Pure cognitive procession were involved in boring N−back task while both "hot" EF and "cold" EF were involved in the entertaining N−back task. Participants were 77 children age between seven and twelve, of whom 60 were classified as having ADHD and /or reading disability (RD). All the disorder participants were recruited at a clinic and normal children were recruited from a elementary school. A four−group mixed design consisting of reading disabilities only (RD, n =15), reading disabilities and ADHD (RD+ADHD, n =24),ADHD only (ADHD, n =21) and a comparison group (n =17) was utilized. In the experiment , two adapted N−back working memory paradigms were used to explore verbal working memory ability, one was a traditional N−back task, another was entertaining N−back task. There are the same difficulty and materials between the two tasks.These results indicate that ADHD and RD groups behaved worse than comparison group and no significant differences had been detected between ADHD and RD groups in the boring task. A significant increase in ADHD had been found when comparing entertaining task with boring task. No significant differences had been detected between ADHD and comparison groups. Also, no significant changes related to the task types had been found in RD children. All these findings suggest that ADHD and RD children both show verbal working memory problems, however they have different mechanisms. The "hot" EF facilitates the performance of ADHD in verbal working memory task while not to RD. These results support the Haber model indirectly. According to this model, "hot" EF modulates "cool" EF by a special pathway.

Keywords ADHD; reading disability; "hot" executive; verbal working memory

幼儿对威胁性刺激蛇的注意觉察：来自眼动证据 *

王福兴 [1]，李文静 [1]，颜志强 [2]，段朝辉 [1]，李　卉 [3]

（[1] 华中师范大学心理学院，武汉 430079）

（[2] 北京大学心理学系，北京 100871）

（[3] 天津师范大学心理与行为研究院，天津 300074）

摘　要　前人研究发现相对于中性刺激花，没有经验的婴幼儿会对威胁性刺激蛇产生更快的觉察反应。研究选取 4–6 岁幼儿和成人被试，改进了刺激材料呈现范式和线画的刺激材料，采用 3 × 3 刺激矩阵呈现的方式，利用眼动仪记录被试的视觉搜索过程，探索威胁性刺激蛇是否被更快注意定向、作为干扰刺激的蛇是否同样能被更快觉察，以及蛇的特殊外形是否在快速觉察中具有重要作用。实验 1 发现，相对于目标物花，成人和幼儿对蛇的首次注视到达时间更短，注视到目标前的注视点个数更少，首次注视的时间更短。实验 2 采用线画的方式去除了刺激材料的色彩和纹理，只保留了蛇蜿蜒的外形，结果发现儿童和成人仍然以更短的注视到达时间、更少注视次数注意到蛇，对蛇的首次注视时间更短。此外，对干扰物分析发现，蛇作为干扰物（花为目标物）仍然表现出更快注意定向。两个实验对比发现，被试对彩色、真实蛇的注视快于线画的蛇。结论认为，即使是对蛇具有较少经验的幼儿，也表现出了对蛇的快速注意偏向；蛇的色彩和纹理会促进蛇的快速觉察；蛇的低水平知觉特征（蜿蜒外形）确实对蛇的注意觉察具有重要作用。

关键词　蛇；威胁性刺激；幼儿；觉察；眼动

分类号　B842

1 引言

恐惧是人类的基本情绪之一，我们对外界事物所产生的恐惧随着年龄的发展而不断变换恐惧对象。婴儿通常会对陌生人恐惧 (Feiring, Lewis, & Starr, 1984)；学龄前的儿童会害怕鬼怪；之后儿童便开始发展出对动物的恐惧，比如：蛇 (LoBue, Rakison, & DeLoache, 2010)。人们对某些事物的恐惧（如：陌生人和鬼怪）可能只在某个发展阶段存在，但是有些恐惧可能会在人的一生中都存在 (Berger, 2010)，比如：蜘蛛 (Purkis & Lipp, 2009)、蛇 (Isbell, 2006; Soares, Esteves, Lundqvist, & öhman, 2009) 和狮子 (Penkunas & Coss, 2013a)。

蛇在很多文化中都被当作恶魔的象征，出现在各种文学、影视作品中 (Isbell, 2006)。研究发现，当要求人们在一些由动物图片组成的刺激矩阵（3×3 或 2×2）中搜索蛇、蜘蛛等威胁性动物时，对蛇和蜘蛛这类威胁性刺激搜索的反应时间更短 (öhman, Flykt, & Esteves, 2001)。对于人类为什么惧怕蛇，心理学研究却有不同的解释。如果作为一个有生活经验的成人，我们从书籍、电视、电影或别人的经验中习得了蛇是恐怖的、有毒的（尽管不是所有蛇都分泌毒素）、被咬到会致命。所以，一些研究者认为，我们对蛇这类威胁性动物的快速觉察可能来自后天的学习或生活经验。比如：öhman 和 Mineka(2001, 2003) 用猴子作为被试研究发现，在实验室养大的猴子对蛇并不会产生恐惧，但是通过观察自然中长大的猴子对蛇的恐惧反应，实验室饲养的猴子可以很快习得对蛇的恐惧。对儿童和成人的对比发现，由于成人对蛇具有更多与恐惧有关的经验，成人对蛇的觉察速度要快于儿童 (LoBue & DeLoache, 2008; LoBue et al., 2010)。此外，在对蜘蛛恐惧研究中也发现后天的经验和学习在恐惧中扮演着重要作用 (Gerdes, Alpers, & Pauli, 2008; Waters, Lipp, & Randhawa, 2011)。

也有一些研究者认为人类天生恐惧蛇，对蛇这类刺激的注意偏向是由于人类具有将蛇和恐惧联结在一起的先天倾向 (innate predisposition)，所以对这些刺激的反应时间更短，觉察速度更快 (öhman & Mineka, 2001, 2003; Quinlan, 2013)。人类对蛇这种威胁性刺激的视觉搜索优势可能是这种行为对人类有其特殊的生物学或进化意义 (Blanchette, 2006; Isbell, 2006; öhman & Mineka, 2003)。从进化的观点来看，人类如果能更有效、更快速地搜索令人感到恐惧的威胁性刺激（蛇、蜘蛛等），将更有可能成功地回避这种危险而生存下来 (相关评述见：Isbell, 2006; LoBue et al., 2010; öhman, 2009)。实验室饲养的猴子研究为这种假设提供了很好的实验支持。因为这些猴子完全是在实验室养大，没有任何与蛇有关的经验，它们仍然会对威胁性刺激产生更快的觉察（参见：öhman & Mineka, 2001; Shibasaki & Kawai, 2009)。 比 如：Shibasaki 和

Kawai(2009) 发现实验室饲养的日本恒河猴对威胁性刺激（蛇）图片的搜索反应时要显著快于非威胁性刺激（花）的图片。此外，近期有关生物对威胁性刺激恐惧的神经生物学研究也为这种假设提供了支持 (Keil et al., 2013)。虽然"先天倾向"假设看似合理，但是它并没有为揭示我们对蛇的快速觉察和反应机制是什么。

由于婴、幼儿没有与蛇等威胁性刺激有关的经验，如果婴、幼儿也能表现出类似于成人的快速觉察反应，这就为人类可能天生恐惧蛇（对蛇的觉察反应更快）提供了强有力的支持。LoBue 和 DeLoach(2008) 在实验中要求 3 ~ 5 岁的儿童在中性刺激（花、青蛙和毛毛虫）中搜索威胁性刺激（蛇）或在威胁性刺激中搜索中性刺激，发现儿童对威胁性刺激蛇的反应显著快于花，表现出了与成人一致的反应模式。Waters，Lipp 和 Spence (2008) 在 9 ~ 12 岁的儿童中也发现了相似的结果。随后，DeLoache 和 LoBue(2009) 采用视觉偏好范式对 7~9 和 14~16 个月婴儿的研究发现，当播放蛇和其它非蛇的视频片段时，婴儿能够将恐惧的声音和蛇的视频片段进行匹配。此外，她们对 8~14 个月婴儿也进行了类似的研究，发现婴儿对蛇的注视反应要快于中性刺激花 (LoBue & DeLoache, 2010)。类似地，Rakison 和 Derringer(2008) 采用视觉偏好和习惯化范式研究了 5 个月大婴儿对蜘蛛和花的注视模式，结果发现婴儿对蜘蛛图像的注视时间要长于花。这些结论似乎都支持人类对蛇具有天生的敏感和快速觉察。

到底是什么原因导致我们对蛇这种特殊的爬行动物的快速觉察？ LoBue 等人 (2013) 认为，蛇的低水平知觉特征（low-level perceptual features）可能起到了重要作用。LoBue 和 Deloache(2011) 实验结果发现，儿童与成人在卷曲的电线和卷曲的蛇之间的反应时没有差异；儿童与成人在非卷曲的蛇与花之间的反应时没有差异，这些都显示可能是蛇的形状在起重要作用。研究者提出了"知觉模板"（perceptual template）假设来解释蛇的外形在幼儿对蛇的注意偏向中的作用 (LoBue, 2013; LoBue et al., 2010; Rakison &

Derringer, 2008)。该假说认为，一些恐惧刺激（蛇、蜘蛛）在人脑中存在知觉模板，它可能包含恐惧刺激的一些特有特征，人们一旦看到这些特征信息就会自动激活这个模板，进而导致更快的注意觉察。如果人对于这类进化相关的恐惧刺激具有天生的快速觉察，那么无论成人还是儿童应该都表现出相同的知觉偏向。"知觉模板"假设在"先天倾向"基础上给出了更加合理的解释，而且也让研究对此类问题的验证更加可操作化，因而受到了研究者的关注。

从行为反应结果（按键）来看，成人研究确实发现威胁性刺激觉察要快于非威胁性刺激。但是，也有研究者认为，这种按键反应时差异可能反应的是对威胁性刺激的快速行动而不是快速觉察 (Flykt & Caldara, 2006; LoBue, Matthews, Harvey, & Stark, 2014)。所以，采用眼动来了解威胁性蛇的视觉搜索对于揭示是否被试真的在视觉搜索阶段就表现出更快的注意定向具有重要意义 (LoBue et al., 2014)。此外，以往研究中对于婴儿注意的测量，更多是采用婴儿对两个事物的注视偏好，并且依据录像编码来确定其加工时间 (参见：DeLoache & LoBue, 2009; LoBue & DeLoache, 2010; Rakison & Derringer, 2008)。尽管在年龄较大的 3~5 岁儿童实验中，研究者使用了改进的触屏范式，要求儿童把手放在一个手模 (handprints) 上，以保证收集到可靠的反应时数据 (具体见：LoBue & DeLoache, 2008)。但也仍然存在从看到刺激到做出动作反应的误差，尤其是当儿童的精细动作反应没有完全发展前。行为方法中的反应时测量的是从感觉器官接收到刺激到动作器官做出反应的全部过程时间。但是，本研究更加关心在做出反应前的视觉搜索阶段（从目标刺激呈现到注视到目标刺激），被试是否也能够更快的定位或觉察到刺激，以及对目标刺激的首次加工。婴、幼儿群体的语言尚未发展成熟，不能进行自我报告，基于眼动技术提供的注视数据就为了解婴、幼儿的认知加工提供了有效的手段 (评述见：Bornstein, Mash, & Arterberry, 2011; Feng, 2011; Gredeback, Johnson, & von Hofsten, 2010; Oakes, 2012)。眼动仪可以记录过程性信息或观看过程，可以为直接测量儿童注视或搜索威胁性刺激提供更加客观和准确的数据。

基于以上论述，本研究实验 1 为验证性实验，目的在于验证先前实验中的行为反应时结果是否会表现出相同的视觉搜索差异，即探索儿童、成人对威胁性刺激蛇是否会表现出更短的首次到达时间、更少的注视次数以及更短的首次注视持续时间。实验 2 在实验 1 的基础上采用线画图 (line drawing)，保留了蛇蜿蜒的外形特征，探讨蛇的低水平知觉特征在快速觉察中作用，并且进一步验证"知觉模板"假设。在先前研究中，LoBue 和 DeLoache(2011) 的实验中所用的图片是真实的蛇和真实的电线图片。实验 2 关心的是如果知觉模板假设成立，即蛇特殊的外形导致了其快速加工。实验使用线画图片把蛇的颜色、纹理结构、头部特征等最大化去除后，只保留蛇的外形特征。这就有助于进一步证实这样一个结论：对蛇的快速觉察不是由于其恐惧性颜色、花纹，而是其特殊的蜿蜒的外形 (LoBue & Deloache, 2011; LoBue et al., 2010)。由于婴、幼儿对蛇具有更少的社会经验，研究者推测婴、幼儿所表现出来的对蛇等威胁性刺激的快速注意偏向可能是由于知觉模板在起作用。而蛇特有的蜿蜒的外形就是知觉模板的关键特征之一。如果人类真的是对蛇弯曲的体形具有更快的反应，我们预期儿童和成人在实验 2 中所有对目标物注视上表现出一致的快速注视。由于实验 1 和实验 2 采用了相同的蛇～花配对，所以，研究也对比实验 1 和实验 2 的数据差异，从而了解蛇的色彩、纹理以及特殊的外形在其快速注意偏向中作用。

其次，LoBue 和 DeLoache(2011) 的研究发现作为干扰刺激的干扰程度对目标刺激的加工不产生影响。实验使用了相同的干扰刺激（马、鹿、兔子）来搜索目标物蛇和目标物青蛙。结果仍然发现蛇快于青蛙。我们对此持有不同的看法，因为原来的研究都是把目标物（蛇、蜘蛛）和干扰物（青蛙、花）配对呈现的。但是，LoBue 和 DeLoache(2011) 的研究中把所有干扰物替换为非威胁性刺激后，得出结论具有局限性和误导性。其最近的研究也发现，蛇作为目标物和干

扰物的首次注视时间没有区别 (LoBue et al., 2014)。如果蛇这类威胁性动物作为目标物会有更快的反应时或觉察，那么蛇作为干扰物仍然会有更快的觉察。对此，有研究者曾提出了"注意脱离困难"(disengaging difficulty) 假设，即对花的反应变慢可能是由于对蛇的注视导致注意脱离困难 (Fox, Russo, Bowles, & Dutton, 2001)。但是这个假设没有在现有同类刺激研究中得到验证。所以，本研究预期当蛇作为干扰物时，对蛇的反应同样会快于作为干扰物的花。对与目标物的配对干扰物进行分析，有助于验证实验的预期。为了验证当威胁性刺激作为干扰物时仍然会具有更快注视，我们对作为干扰物的蛇和花进行了分析。

最后，两个实验在实验范式和被试选取上做了一些改进。第一，有关威胁性刺激搜索研究发现，3×3 矩阵搜索范式的中间位置的刺激加工速度更快 (Blanchette, 2006)。其他研究发现，3×3 矩阵搜索中中心位置和四周位置存在反应时间差异 (Brosch & Sharma, 2005)。而以往采用搜索范式的研究中，都没有对中心位置效应进行平衡或处理。另外，考虑到眼动仪校准的注视点会出现在屏幕中央，如果目标刺激位置在矩阵中心，对目标物快速注视或搜索可能不是威胁性本身导致，而是其位置有利于加工导致的。所以，研究修改了原有的视觉搜索范式，在 3×3 矩阵搜索中，去除了中心位置刺激。第二，以往研究中大多数都是方便取样，选取被试的父母作为成人对照组（如：DeLoache & LoBue, 2009; LoBue & DeLoache, 2008, 2010)。相关研究发现，父母对蛇的恐惧经验、家族恐惧历史对个体恐惧都具有影响作用 (Fredrikson, Annas, & Wik, 1997; Murray & Foote, 1979)。所以本研究拟选取没有任何血缘关系的大学生被试群体作为对照的成人组，这样可以避免因为儿童与父母之间的生活经验相似性或遗传的影响。

2　实验 1

以往研究发现，儿童对威胁性刺激蛇的觉察反应要快于中性刺激花。实验 1 关注在做出反应前的视觉搜索阶段，威胁性刺激是否先于非威胁性刺激被优先注视到？实验修改呈现范式，去除了中间位置刺激，采用经典的蛇～花配对作为刺激材料，记录被试眼动数据，验证儿童、成人对威胁性刺激是否有更快的、更短的注视。此外，还分析了当蛇作为干扰物（目标物为花）时，是否也会有更快的视觉搜索。

2.1　被试

24 名 4~6 岁的儿童。由于实验过程中不专心、动作幅度过大（眼动采样率均小于 75%），8 名儿童数据没有进入分析，有效被试 16 人（女 7 人），平均年龄为 5.1 岁（SD=0.6）；成人被试选自某师范大学本科生 22 人（女 11 人），平均年龄 19.6 岁 (SD=1.2)。所有被试视力或者矫正视力正常，无色盲、色弱。

关于蛇的先验知识问卷参照 LoBue 等人的研究自己编制 (LoBue, 2010b; LoBue & DeLoache, 2008)。问卷采取是否式计分（选"是"或"消极"计 1 分，"否"或"积极"记 0 分，最高分 4 分，具体见附录 1）。由于已有研究发现儿童记忆不准确 (Braun, Ellis, & Loftus, 2002; Cole & Loftus, 1987; Loftus & Davies, 1984)，以及幼儿言语和表达能力发展不够完善，所以参照以往研究方式，儿童关于蛇的先验知识问卷由父母填写。实验 1 父母报告的儿童先验知识平均得分 1.4（SD=1.2），成人自我报告的平均得分 2.4（SD=1.0），成人的经验显著高于儿童，$t(36) =2.70$, $p<0.05$, Cohen's $d=0.9$。此外，实验 1 与实验 2 的成人在先验知识问卷上无显著差异，$t(43)=-1.74$，$p>0.05$, $d=0.5$；实验 1 与实验 2 儿童群体也无显著差异，$t(43)=-0.47$，$p>0.05$, $d=0.1$。先验问卷分析发现性别、城乡对蛇的恐惧程度没有差异 ($p_s>0.05$)，因此在两个实验结果中均没有包含这些变量。

2.2　材料

本研究所有材料图片均来自于互联网，由研究者仿照 LoBue 等人研究 (DeLoache & LoBue, 2009; LoBue, 2010b; LoBue & DeLoache, 2008) 自己制作。图片刺激类型为蛇和花，共 72 张图片，均去除图片背景。在每个实验试次中，显示器会呈现一个 3×3 的图片矩

图 1　实验 1 材料图片示例（左图目标物为蛇，右图目标物为花）

阵（见图 1），每个矩阵中心位置不呈现图片，共包括 8 个刺激物。每个刺激矩阵中包括 1 个目标物（蛇或花）和 7 个干扰物（花或蛇）。实验目标物每种类型各 8 张，不重复出现（随机呈现），干扰图片每种类型（蛇或花）各 28 张，为了保证实验效果，所有干扰物图片采取伪随机，即在与目标图片进行匹配时重复一次。用于视觉搜索的 3×3 矩阵共 16 个，其中蛇和花作为目标物各 8 个。矩阵中每张图片大小为 325×245 像素。

采取单盲方式，请20位心理学专业学生（女 10 人）对 72 张刺激图片的恐惧性程度进行 7 点（1，一点都不害怕；7，非常害怕）以及明亮度进行 5 点（1，非常暗；5，非常亮）评定。结果显示恐惧唤起程度有显著性差异 $t(70) = 56.51$, $p<0.001$, $d=13.5$，蛇的恐惧唤起更高；亮度不存在差异 $t(70) =-0.70$, $p>0.05$, $d= 0.2$。

2.3 仪器、设计与程序

实验仪器为 Tobii T120 Eye-tracker（Tobii Technology, Sweden）。双眼红外追踪，采样率 120Hz，眼睛距屏幕距离 60cm（17 英寸显示器，1024×768 分辨率）。单张图片像素大小为 200×150 像素，单张图片刺激物的水平视角约 3.3 度，垂直视角约 2.5 度；矩阵刺激图片的像素为 800×600，矩阵图片的对侧视角约 14.3 度。

实验为 2（年龄：成人、儿童）×2（目标类型：蛇、花）的混合实验设计，实验范式修改自 LoBue 等人用于儿童的搜索范式 (DeLoache & Lo Bue, 2009; LoBue,2010b; LoBue & DeLoache, 2008)。为了控制中心位置的影响，去除了 3×3 矩阵中间位置的刺激。

在正式实验进行前，被试先要完成 7 个试次的练习（练习材料均不出现在正式实验材料中），最初要求儿童依次观看两个单张的图片，第一个是目标类型，第二个是干扰类型，目的是熟悉实验材料；接下来的两个练习是同时呈现一幅目标图片和一幅干扰图片，要求儿童观看目标图片；最后的 3 个是呈现 3×3 的矩阵图片（与正式实验一致），让儿童在干扰图片中寻找目标图片（与其他刺激不同的图片），找到后口头报告，然后主试操作进入下一个试次。正式实验中每个矩阵图片最长呈现 3s，在矩阵图片呈现之前先呈现注意吸引图片（卡通米奇），当主试确认被试确实是在看图片刺激时才进入到下一个试次，共 16 个试次（呈现顺序随机）。在每一个矩阵图片呈现的间隔都会呈现一张米奇老鼠的图片以吸引被试的注意。实验过程中，一名主试陪伴在儿童身边指导其完成练习测试。成人实验程序与儿童相同，但是指导语表述方式有所更改，去掉了儿童化语言表述。实验结束后，每位被试都需要完成一份先验知识问卷（儿童的问卷由父母完成）。

2.4 结果

为了探讨目标刺激对人的注意捕获能力和干扰物对人的注意干扰能力，对刺激图片划分了两类兴趣区（Area of Interest,AOI）：目标物（蛇或花）和干扰物（花或蛇），并且把所有干扰物看作一个整体进行数据导出和分析。采用的眼动指标为：首次注视到达时间（Elapse Time of First Fixation to AOI，该指标计算的是从刺激呈现到第一次注视到目标物的时间，时间越短表明目标被越早注视到）；首次进入兴趣区之前的注视点个数（Fixation Count Before Enter

AOI，是指在被试的首个注视点进入兴趣区之前的注视点个数，次数越少说明目标物被识别越快）；首个注视点的持续时间（First Fixation Duration of AOI，指的是进入目标兴趣区的第一个注视点的注视持续时间，此指标说明了被试对目标物的首次加工时间）（见表1）。本研究剔除了三个标准差之外的极端数据（原始眼动数据），对所有眼动数据采用2（年龄）×2（目标类型）的重复测量方差分析。

2.4.1 目标物注视

在首次到达兴趣区时间上，目标类型主效应显著（$F(1,36)$=45.14，$p<0.001$，η_p^2=0.56，partial η^2，下同），事后检验（Bonferroni，下同）发现目标刺激蛇（M=638）被首次注视到时间显著快于目标刺激花（M=905）。年龄主效应显著（$F(1,36)$=86.14，$p<0.001$，η_p^2=0.71），成人（M=479）首次到达目标刺激时间显著快于儿童（M=1174）。目标类型与年龄的交互效应显著（$F(1,36)$=22.74，$p<0.001$，η_p^2=0.39），儿童首次到达目标物蛇的时间显著短于花（$F(1,36)$=56.98，$p<0.001$）。

首次进入兴趣区之前的注视次数，目标类型主效应显著（$F(1,36)$=36.42，$p<0.001$，η_p^2=0.50），到达蛇（M=2.5）之前的注视次数显著少于花（M=3.4）。年龄主效应显著（$F(1,36)$=86.73，$p<0.001$，η_p^2=0.71），成人（M=2.2）到达目标物之前的注视次数显著少于儿童（M=4.0）。目标类型与年龄的交互效应显著（$F(1,36)$=12.84，$p<0.01$，η_p^2=0.26），相对于目标物花，儿童以更少注视次数就注视到了目标物蛇（$F(1,36)$=39.94，$p<0.001$），成人也表现出相同的结果，即更少注视次数就锁定了目标物蛇（$F(1,36)$=3.57，p=0.067）。

兴趣区首个注视点持续时间，目标类型主效应显著（$F(1,36)$=12.01，$p<0.01$，η_p^2=0.25），蛇的首个注视点的注视时间（M=459）显著短于花（M=627）。年龄主效应显著（$F(1,36)$=26.15，$p<0.001$，η_p^2=0.42），成人（M=695）首个注视点持续时间显著长于儿童（M=334）。目标类型与年龄的交互效应

显著（$F(1,36)$=10.45，$p<0.01$，η_p^2=0.23），成人对蛇的首个注视点的持续时间显著短于花（$F(1,36)$=26.64，$p<0.001$）。

2.4.2 干扰物注视

在兴趣区首次注视到达时间上，干扰类型主效应显著（$F(1,36)$=22.83，$p<0.001$，η_p^2=0.39），干扰刺激蛇（M=381）被首次注视到时间显著快于干扰刺激花（M=515）。年龄主效应不显著（$F(1,36)$=1.33，$p>0.05$，η_p^2=0.04）。干扰物与年龄的交互效应显著$F(1,36)$=6.82，$p<0.05$，η_p^2=0.16），成人对蛇的首次注视到达时间显著短于花（$F(1,36)$=32.43，$p<0.001$）；在干扰刺激蛇上，成人的首次注视到达时间显著短于儿童$F(1,36)$=7.73，$p<0.01$（见表1）。

首次进入兴趣区之前的注视点个数上，干扰类型主效应显著（$F(1,36)$=27.71，$p<0.001$，η_p^2=0.44），在看到蛇（M=1.1）之前注视点个数显著少于花（M=1.6）。年龄主效应不显著（$F<1$，$p>0.05$，$\eta_p^2<0.01$）。干扰物与年龄的交互效应显著（$F(1,36)$=9.07，$p<0.01$，η_p^2=0.20），在干扰刺激蛇上，成人到达兴趣区之前的注视点个数显著少于儿童（$F(1,36)$=5.85，$p<0.05$）；相对于花，成人以较少注视次数就注视到了蛇（$F(1,36)$=40.66，$p<0.001$）。

在兴趣区的首个注视点的注视持续时间上，干扰类型（$F(1,36)$=2.33，$p>0.05$，η_p^2=0.06）和年龄主效应（$F(1,36)$=1.45，$p>0.05$，η_p^2=0.04）均不显著。干扰类型与年龄的交互效应显著（$F(1,36)$=8.73，$p<0.01$，η_p^2=0.20）。在干扰刺激花上，成人首个注视点的注视持续时间短于儿童$F(1,36)$=11.34，$p<0.01$）；成人对蛇的首次注视持续时间长于花$F(1,36)$=11.92，$p<0.01$。

2.5 讨论

当蛇作为搜索目标时，实验1的眼动结果重复且证实了威胁性刺激蛇的首次注视时间要明显快于中性刺激花，即从刺激呈现到注视点落到蛇上，被试的首次注视到达时间更短。而且，在注视到目标物蛇上所用的注视点个数更少，对蛇的首次观看持续时间也更短。这些结果都与之前行为反应的结

表 1　　儿童和成人对威胁性刺激和非威胁性刺激的眼动数据

实验	兴趣区眼动	年龄	目标物				干扰物			
			蛇		花		花		蛇	
			M	SD	M	SD	M	SD	M	SD
	首次注视到达时间 (ms)	儿童	917	206	1432	533	503	142	447	161
		成人	435	62	522	74	524	176	334	89
实验 1	首次进入前注视点个数	儿童	3.2	1.0	4.7	1.1	1.5	0.3	1.3	0.4
		成人	2.0	0.4	2.4	0.5	1.7	0.5	1.0	0.3
	首个注视点持续时间 (ms)	儿童	329	77	339	52	258	35	248	59
		成人	553	244	836	386	222	31	254	47
	首次注视到达时间 (ms)	儿童	1126	187	1314	508	485	157	445	79
		成人	725	140	756	141	357	93	317	89
实验 2	首次注视前注视点个数	儿童	3.9	0.7	4.7	1.9	1.5	0.4	1.3	0.2
		成人	3.3	0.7	3.6	0.5	1.5	0.4	1.4	0.4
	首个注视点的持续时间 (ms)	儿童	341	88	398	112	284	54	274	58
		成人	350	77	386	88	220	32	205	35
	行为反应（ms）	儿童	3116	605	3233	602	/	/	/	/
		成人	1568	286	1625	276	/	/	/	/

果相一致，即被试对蛇这种威胁性刺激觉察更快，搜索到的时间更短 (LoBue & DeLoache, 2008; LoBue & Deloache, 2011;öhman, Flykt, etal., 2001; öhman & Mineka, 2001)。这个结果也与情绪面孔眼动研究一致，负性面孔的首次注视潜伏期更短 (Reynolds, East wood, Partanen, Frischen,& Smilek,2009)。LoBue 近期对于成人研究也发现，相对于花，首次注视到蛇的潜伏期更短 (Lo Bue etal.,2014)。

儿童表现出了和成人一致的趋势，即无论是儿童还是成人对蛇的觉察都要快于花。在交互作用上，儿童也是对蛇的搜索注视要快于花，这个结果与 LoBue 和 De Loache（2008）LoBue 和 Deloache（2011）采用触屏搜索范式结果是一致的。虽然本实验所选的成人被试不是儿童的父母，成人对目标物（蛇或花）的注视仍要快于儿童，这个结果也与以往研究类似 (Lo Bue & De Loache, 2008; LoBue & Deloache, 2011)。但是，在注视点持续时间上发现成人更长。可能是由于成人具有更成熟的视觉搜索技能以及对蛇具有更多的社会经验。所以，成人的反应要明显快于儿童，并且对蛇加工也更多。有关成人的研究发现，经验可以调节被试对威胁性刺激的视觉搜索 (Gerdes et al, 2008; Peira, Golkar, Larsson, & Wiens, 2010)，通过调查问卷数据也可以发现，儿童缺少与蛇有关的后天经验。

对干扰物的蛇进行分析，眼动数据显示无论是儿童还是成人对蛇的首次注视到达时间要短于花，以较少的注视次数就注意到蛇。这个结果与蛇和花作为目标物进行分析时结果是一致的。该结果证实了预期的假设，即不管作为目标物还是干扰物，威胁性刺激的注意定向都更快。即当蛇作为干扰物出现时，其作为恐惧刺激所具有的快速觉察仍然得到了体现。但是，对干扰物分析的眼动结果并没有支持"注意脱离困难"假设 (Fox et al., 2001)。虽然蛇作为干扰物时的视觉搜索快于花，但是在首次注视持续时间上并没有发现花与蛇的区别。

3　实验 2

在实验 1 基础上，借鉴情绪面孔研究中采用线画或模式化来突出刺激的形态特征方式 (Fox et al., 2000; LoBue & Larson, 2010)，使用线画的蛇和花作为刺激材料，进一步验证蛇低水平的知觉特征（蜿蜒外形）是否会对早期的视觉加工和搜索产生影响，检验"知觉模板"假设。儿童和成人是否会对蛇的形态特征有更快的注意和更短的注视加工？以及当蛇作为干扰物时，这种特征是否仍然起作用？此外，与实验 1 数据进行对比，进一步讨论蛇的外形、颜色、纹理在其快速觉察中的作用。
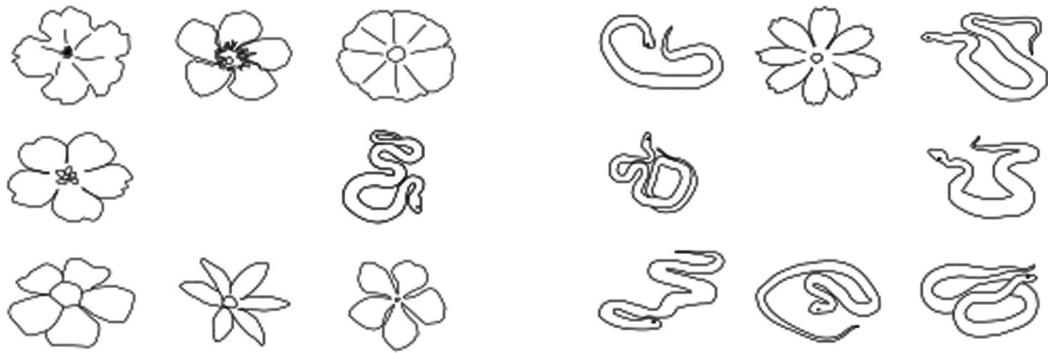
图 3　实验 2 刺激材料示例（左图目标物为蛇；右图目标物为花）

3.1　被试

22 名 4~6 岁的儿童，有 2 名儿童数据未进入分析（采样率小于 75%），有效被试为 20 人（女 5 人），平均年龄为 5.1 岁（$SD=0.5$）；成人被试选取本科生 23 人（女 10 人），平均年龄 20.0 岁 ($SD=1.3$)，所有被试视力或者矫正视力正常，无色盲、色弱。实验 2 中父母报告儿童关于蛇的先验知识平均得分 1.58（$SD=0.51$），成人平均得分 2.87（$SD=0.51$），成人经验显著高于儿童，$t(40)=6.34$，$p<0.001$，$d=2.0$。

3.2　材料

本研究作者参照实验 1 中图片，手工画制了所有线画图，将图片类型变量中的真实蛇和花转换成了线画蛇和花（见图 3 及附录 4）。所有实验材料均参照实验 1 中彩色图片绘制。18 位不知实验目的心理学本科生（女 9 人）对这些线画图进行了评定（评定方式和计分同实验 1），结果发现恐惧唤起程度有显著性差异 $t(70)=28.16$，$p<0.001$，$d=6.7$，恐惧刺激蛇的唤起度显著高于花；亮度不存在差异 $t(70)=-1.06$，$p>0.05$，$d=0.3$（数值见附录 3）。

3.3　仪器、设计与程序

仪器同实验 1；实验设计同实验 1。实验基本流程同实验 1。不同的是，在实验 2 中记录了被试的行为反应时。由于儿童被试年龄较小且无法精确操作鼠标，实验要求被试在口头报告发现目标物后，由主试及时点击鼠标（鼠标图标不在屏幕上出现），记录从刺激呈现到点击鼠标时间作为被试的行为反应时间。主试点击鼠标后自动进入下一实验序列。成人采用相同方式记录反应时间。

3.4　实验 2 结果

3.4.1　目标物注视

在兴趣区首次注视到达时间上发现，目标主效应显著（$F(1,41)=4.34$，$p<0.05$，$\eta_p^2=0.10$），目标物蛇（$M=911$）被首次注视到时间短于花（$M=1016$）。年龄主效应显著（$F(1,41)=50.56$，$p<0.001$，$\eta_p^2=0.55$），成人（$M=740$）快于儿童（$M=1220$）。目标类型与年龄的交互效应不显著（$F(1,41)=2.22$，$p>0.05$，$\eta_p^2=0.05$）。

到达兴趣区之前的注视次数，目标主效应显著（$F(1,41)=5.98$，$p<0.05$，$\eta_p^2=0.13$），蛇（$M=3.6$）在首次被注视到之前注视次数显著少于花（$M=4.1$）。年龄主效应显著（$F(1,41)=12.08$，$p<0.01$，$\eta_p^2=0.23$），成人（$M=3.4$）在到达目标刺激前注视次数显著少于儿童（$M=4.3$）。目标类型与年龄的交互效应不显著（$F(1,41)=1.12$，$p>0.05$，$\eta_p^2=0.03$）。

兴趣区首次注视持续时间，目标主效应显著（$F(1,41)=9.94$，$p<0.01$，$\eta_p^2=0.19$），蛇（$M=346$）的首次注视持续时间显著短于花（$M=392$）。年龄主效应不显著（$F<1$，$p>0.05$，$\eta_p^2<0.01$）。目标类型与年龄的交互效应不显著（$F<1$，$p>0.05$，$\eta_p^2=0.01$）。

3.4.2 干扰物注视

兴趣区首次注视到达时间上，干扰类型主效应显著（$F(1,41)=4.63$，$p<0.05$，$\eta_p^2=0.10$），干扰刺激蛇（$M=376$）被首次注视到时间显著短于干扰刺激花（$M=417$）。年龄主效应显著（$F(1, 41)=22.17$，$p<0.001$，$\eta_p^2=0.35$），成人（$M=337$）显著短于儿童（$M=465$）。干扰类型与年龄的交互效应不显著（$F<1$，$p>0.05$，$\eta_p^2<0.001$）。

首次进入兴趣区之前的注视点个数上，干扰类型主效应边缘显著（$F(1, 41)=3.52$，$p=0.068$，$\eta_p^2=0.08$），蛇（$M=1.4$）注视次数少于花（$M=1.5$）。年龄主效应（$F<1$，$p>0.05$，$\eta_p^2<0.01$）以及干扰类型与年龄的交互效应（$F<1$，$p>0.05$，$\eta_p^2<0.01$）均不显著。

首个注视点的注视持续时间上，干扰类型主效应边缘显著（$F(1,41)=3.31$，$p=0.076$，$\eta_p^2=0.08$），蛇（$M=237$）的首个注视点持续时间短于花（$M=249$）。年龄主效应显著（$F(1,41)=30.08$，$p<0.001$，$\eta_p^2=0.42$），成人（$M=212$）的首次加工时间显著短于儿童（$M=279$）。干扰类型与年龄的交互效应不显著（$F<1$，$p>0.05$，$\eta_p^2<0.01$）。

3.4.3 行为反应

目标类型主效应边缘显著（$F(1,41)=3.78$，$p=0.059$，$\eta_p^2=0.08$），目标物蛇（$M=2288$）的反应快于花（$M=2373$）。年龄主效应显著（$F(1,41)=140.31$，$p<0.001$，$\eta_p^2=0.77$），成人（$M=1596$）反应快于儿童（$M=3174$）。目标类型与年龄的交互效应不显著（$F<1$，$p>0.05$，$\eta_p^2=0.01$）。

对目标物蛇的首次注视到达时间和行为反应进行配对样本 t 检验，发现眼睛首次到达时间显著短于行为反应，$t(42) = -12.32$, $p < 0.001$, $d = 1.9$。目标物花的对比也发现，眼睛首次到达时间要显著快于行为反应，$t(42) =-11.73$, $p < 0.001$, $d =1.8$。

3.5 实验1和实验2对蛇的注视比较

为了进一步揭示蛇的外形在快速觉察中的作用，对实验1和实验2的数据进行了对比分析。探讨去掉了颜色和纹理后，实验2中的蛇是否被注视的更早。

由于实验2中增加了被试报告及要求主试点击鼠标记录被试的反应时间，为了确保实验1和实验2数据对比的有效性，计算了儿童和成人对目标物蛇的首次到达时间和首个注视点持续时间均值之和（实验1儿童：917+329=1246 ms；实验1成人：435+553=988 ms；实验2儿童：1126+341=1467 ms；实验2成人：725+350=1075 ms），发现数值远低于实验2中儿童反应时均值3116 ms和成人反应时均值1568 ms。由于研究关注的是眼动的初期反应指标（首次注视到达时间、首次注视前注视点个数、首次注视点持续时间），而不考虑后期指标（比如：平均注视时间、总注视时间、总注视次数）。所以，实验2中记录行为反应时过程并不会对3个眼动指标产生影响。因为儿童和成人的反应时均值都滞后于对目标物蛇首次到达时间和首个注视点持续时间均值之和，说明被试必须先搜索到目标刺激才能做出反应。此外，在最近采用成人被试眼动研究中，发现成人被试从首次注视结束到做出按键反应（反应时）的延迟时间为748 ms(恐惧刺激)和826 ms(非恐惧刺激)(LoBue et al., 2014, p.820)，这也说明是在首次注视结束后做出行为反应，而不影响眼动初期反应指标。数据采用了2（蛇的特征：真实、彩色的蛇，蛇的线画图）×2（年龄：成人，儿童）的被试间分析。其中，真实、彩色的蛇的数据来源于实验1，蛇的线画图数据来自实验2（见表1）。

3.5.1 蛇作为目标物

首次注视到达时间上，特征主效应显著（$F(1,77)=52.49$，$p<0.001$，$\eta_p^2=0.41$），实验1中彩色、真实的蛇（$M=638$）首次注视到达时间显著短于实验2中线画的蛇（$M=911$）。年龄主效应显著（$F(1,77)=164.24$，$p<0.001$，$\eta_p^2=0.68$），成人（$M=583$）快于儿童（$M=1033$）。特征与年龄的交互效应不显著（$F(1, 77)=1.37$，$p>0.05$，$\eta_p^2=0.02$）。

首次注视到蛇之前的注视次数，特征主效应显著（$F(1, 77)=39.35$，$p<0.001$，$\eta_p^2=0.34$），彩色、真实的蛇（$M=2.5$）注视次数显著少于线画的蛇（$M=3.6$）。年龄主效应显著（$F(1, 77)=33.70$，

$p<0.001$，$\eta_p^2=0.30$），成人（$M=2.7$）少于儿童（$M=3.6$）。特征与年龄的交互效应不显著（$F(1, 77)=2.78$，$p>0.05$，$\eta_p^2=0.04$）。

首个注视点持续时间上，特征主效应显著（$F(1, 77)=8.63$，$p<0.01$，$\eta_p^2=0.10$），实验1中彩色、真实的蛇（$M=459$）首次注视时间显著长于实验2中线画的蛇（$M=346$）。年龄主效应显著（$F(1, 77)=12.92$，$p<0.01$，$\eta_p^2=0.14$），成人（$M=449$）长于儿童（$M=335$）。特征与年龄的交互效应显著（$F(1, 77)=10.94$，$p<0.01$，$\eta_p^2=0.12$），成人对真实、彩色的蛇（实验1）首个注视点时间长于线画的蛇（实验2）（$F(1, 77)=21.60$，$p<0.001$）。

3.5.2 蛇作为干扰物

首次注视到达时间上，年龄主效应显著（$F(1,77)=26.19$，$p<0.001$，$\eta_p^2=0.25$），成人（$M=325$）快于儿童（$M=446$）。特征主效应（$F<1$，$p>0.05$，$\eta_p^2<0.01$）、特征与年龄的交互效应（$F<1$，$p>0.05$，$\eta_p^2<0.01$）均不显著。

首次注视到干扰物前注视次数，特征主效应显著（$F(1,77)=8.03$，$p<0.01$，$\eta_p^2=0.09$），彩色、真实的蛇（$M=1.1$）注视次数显著少于线画的蛇（$M=1.3$）。年龄主效应不显著（$F(1,77)=2.59$，$p>0.05$，$\eta_p^2=0.03$）。特征与年龄的交互效应边缘显著（$F(1,77)=3.86$，$p=0.053$，$\eta_p^2=0.05$）。成人首次注视到真实、彩色蛇前注视次数少于线画蛇（$F(1,77)=12.88$，$p<0.01$）。

首个注视点持续时间上，特征主效应不显著（$F(1,77)=1.07$，$p>0.05$，$\eta_p^2=0.01$）。年龄主效应显著（$F(1,77)=7.79$，$p<0.01$，$\eta_p^2=0.09$），成人（$M=229$）短于儿童（$M=262$）。特征与年龄的交互效应显著（$F(1,77)=10.94$，$p<0.01$，$\eta_p^2=0.13$），成人对真实彩色蛇的首次加工时间长于线画蛇（$F(1,77)=11.20$，$p<0.01$）。

3.6 讨论

实验2结果与实验1类似，即相对于非威胁性的花，蛇被首次注视到时间更短，所用注视次数更少，且首次加工的持续时间更短。这个结果证实了预期

假设，并且与采用其他刺激物对比的结论存在一致性(LoBue & Deloache, 2011)，为儿童和成人对蛇的快速加工是由于低水平知觉特征提供了直接证据。这种对蛇的轮廓外形效应，在其他研究中也有类似验证。比如：Blanchette (2006) 的研究发现，被试对威胁性刺激的符号表征（蛇的卡通图片）搜索反应时显著得快于中性刺激。

对干扰物的分析仍然发现了与实验1类似结果。即虽然作为干扰物，蛇仍然表现出更短的首次注视时间、更少注视次数和更短的首次注视持续时间。这不仅进一步证明了威胁性刺激蛇确实具有更短的觉察时间，同时也说明当保留蛇蜿蜒的形态特征后，作为干扰物蛇的反应仍然更快。实验2记录了被试觉察目标的行为反应速度，虽然过程不够精细，但是结果仍然重复了已有研究结论，发现蛇的反应要快于花、成人要快于儿童 (LoBue & DeLoache, 2008; LoBue & DeLoache, 2011)，也说明了结论的有效性。但是，从统计显著性水平来看，相对于真实、彩色的蛇，线画的蛇导致的反应强度还是降低了，即彩色、真实蛇具有更快的觉察效应。这个结论可以在实验1和实验2对比中得到验证。

对比行为反应和首次注视到达时间可以发现，首次到达时间要明显快于反应时，但是结果趋势具有一致性。说明行为反应的过程需要经过视觉搜索、信息传递、神经中枢决策、肌肉反应等过程，而视觉搜索和定向更加快速。也证明了眼动在威胁性刺激快速觉察研究中可以提供更加直接和有效的反应指标。此外，也有研究者认为，眼动记录视觉搜索（比如：首次注视到达时间）反应了被试对于威胁性刺激觉察的自下而上的加工过程，而行为反应则涉及到自上而下的加工过程(LoBue et al., 2014)。从客观上而言，眼动自下而上的加工过程能够更好地反应人们对蛇的快速觉察。

在实验2中线画蛇相对于真实的彩色蛇去除了许多额外信息（如：颜色、背景、突出的头部特征、纹理等）而保留了低水平知觉特征（蜿蜒的形状）。与实验1数据对比分析发现，当蛇作为目标物进行搜

索时，实验1中真实、彩色蛇的被注视到时间明显快于线画的蛇。这个结论证实了我们的推测，说明蛇所具有的颜色、纹理等额外信息还是促进了对蛇的识别。该结论提示我们，蛇作为一种特殊爬行动物，其独有的特征不仅仅包括其弯曲的外形，其颜色和纹理还是突出蛇的特殊性，使其被快速识别和觉察(Isbell, 2006, 2009)。在 LoBue 和 DeLoache(2011) 研究的一个实验中，研究者控制了刺激呈现的颜色（采用黑白图片）探讨蛇特殊的色彩和外表对变化刺激觉察影响，发现无论成人还是儿童对蛇的觉察仍然快于青蛙。她们的结论认为颜色并不影响对蛇的快速觉察。但是，在改变威胁刺激材料呈现方式后，我们发现色彩仍具有一定促进作用。当蛇作为干扰物时，这种色彩和纹理的效应明显下降，可能是由于当蛇作为干扰物时，被试不是主动去识别蛇。此外，相关研究发现情绪唤醒有助于视觉知觉(Phelps, Ling, & Carrasco, 2006; Vaish, Grossmann, & Woodward, 2008)。实验2的线画刺激相比于真实图片，情绪唤醒度可能较低，可能导致其觉察效应下降。

4　总讨论

本研究使用了眼动仪来采集被试对威胁性刺激的反应数据。根据以往的婴幼儿视觉搜索研究 (DeNicola, Holt, Lambert, & Cashon, 2013; Koster, Crombez, Van Damme, Verschuere, & De Houwer, 2004; Lipp & Waters, 2007; Peltola, Leppänen, Vogel-Farley, Hietanen, & Nelson, 2009)，可以把被试的首次注视到达时间和首次进入兴趣区前注视点个数界定为目标物注意定向（attention orienting），即被试对目标刺激（蛇或花）的注意定向速度，时间越短或注视次数越少，说明被试更早或更快地就锁定了目标刺激；把首个注视点持续时间界定为目标物注意维持（attention holding），即对目标物的注视维持时间越短，说明被试对目标加工更少。实验1重复和扩展了 LoBue 等人的实验(LoBue & DeLoache, 2008, 2010)，发现无论是成人还是儿童对威胁性刺激蛇具有更快的注意定向，更短的注意维持。

实验2在去除了蛇的颜色等外部特征后也发现了相同的结论。这些基于眼动的研究结果，也为理解威胁性刺激优先觉察提供了更加客观的实证数据。这说明无论是有经验的成人被试，还是较少经验的3~5岁儿童，他们都表现出对蛇更快的觉察。

恐惧研究是一个很广泛的领域，人们不仅对蛇、蜘蛛、蜥蜴这类威胁性动物产生注意偏向或快速反应，同时也对恐惧面孔 (LoBue & Larson, 2010; öhman, Lundqvist, & Esteves, 2001)、威胁性武器枪 (Fox, Griggs, & Mouchlianitis, 2007)、刀和注射器 (Blanchette, 2006; LoBue, 2010b) 等产生注意偏向。研究者把对蛇和蜘蛛这类刺激称为进化相关的威胁（evolutionary relevant threats）或种系相关的(phylogenetical)刺激(Fox et al., 2007; LoBue et al., 2010)。通过以上的两个实验发现，虽然儿童对蛇具有更少的经验，但是他们仍然表现出了与成人一样的注视模式，对蛇表现出了更快的觉察。这就让我们进一步思考一个重要的问题，即我们为何会对蛇觉察更快，是因为我们恐惧蛇，还是蛇特殊的外形。öhman 等人提出了"先天倾向"假说来解释蛇的快速觉察，他们假设人类认知系统中存在一个恐惧模块（evolved fear module），它是一个相对独立的行为、心理和神经系统，某些特定的刺激会自动化激活的神经系统，这些神经系统会对一些特定的威胁刺激（蛇）有选择性地、自动地激活，并产生防卫性反应。他们认为这个恐惧模块是进化适应的结果，在大部分哺乳动物中都存在 (öhman & Mineka, 2001, 2003; öhman, Soares, Juth, Lindström, & Esteves, 2012)。虽然先天倾向的进化模块可以解释现实情境中的一些恐惧刺激觉察现象，但是却不能很好地解释人类对一些低水平特征的威胁性刺激的快速觉察现象。即这个恐惧模块很难解释我们对一些特定威胁刺激的注意偏向或快速觉察机制。

LoBue 等人在恐惧模块基础上使用了低水平知觉偏向和知觉模板假设来解释婴、幼儿对蛇的觉察 (LoBue, 2013; LoBue et al., 2010)。该假说认为，对蛇具有较少经验的婴、幼儿所表现出来的快速注意偏向可能是由于知觉模板在起作用。知觉模板主要是

基于威胁刺激的基本特征或混合的低水平的图示化特征，例如，蛇的知觉模板就是其连续的曲线轮廓以及尾部弯曲为一团；蜘蛛的视觉模板是居于中心的身体及其发散到四周的曲线 (LoBue, 2013; LoBue et al., 2010)。实验 2 中结果支持了蛇弯曲的外形在快速觉察中的作用，与知觉模板假设相符合。因为经过线画处理后，蛇本身的曲线特征更加突出，可以说蛇特殊的轮廓特征导致了其快速注意定向和更短的注意维持。但是，仍然需要深入探讨的一个问题是，这种觉察是否是真的由于其特殊的外形特征，还是人类本身就对特殊的形状反应更快。因为在人类基本视觉搜索中，弯曲的线相较于直线确实更容易引起人们的视觉注意和反应 (Treisman & Gormican, 1988; Wolfe, Yee, & Friedman–Hill, 1992)。对于这个问题解释，LoBue 等人也对于知觉模板假设持有比较谨慎的态度 (LoBue, 2013; LoBue et al., 2010)，后续研究仍然需要对这一问题进行深入探讨。

蛇作为爬行动物，其所具有的特殊外形、头部特征、颜色和纹理是其区别于其他爬行动物重要特征 (Isbell, 2009)。对比实验 1 和实验 2 数据发现，儿童和成人对于实验 1 中真实、彩色蛇的反应明显要快于实验 2 中去掉这些特征点的线画蛇。这说明蛇作为一种特殊的爬行动物，之所以被大部分哺乳动物和人类恐惧和快速觉察，是因为其所具有的颜色、花纹、特殊外形、头部特征作为一个综合体仍然起重要作用。比如，相对于同为爬行动物的蜥蜴，虽然两者在外形一些外部特征上具有相似性，但是研究发现蛇的觉察仍然要快于蜥蜴 (Penkunas & Coss, 2013b)。这也可能是 LoBue 等人前期研究中使用蛇的真实图片作为刺激材料的原因，即现实生活环境中，无论是在动物园还是在野外环境中见到的蛇都是带有色彩的、有纹理的、弯曲的。

研究对于干扰刺激注视情况分析证实了研究的预期，即虽然作为搜索刺激的干扰物，威胁性刺激蛇仍然被更早和更快的注意定向，并且有更短的注意维持时间。这进一步间接证实了恐惧性刺激快速知觉的普遍性。同时，这个结果确实不同于已有研究

认为干扰物属性不影响对目标刺激觉察的结论，但是也不能完全支持"注意脱离困难"假设 (Fox et al., 2001)。比如：Penkunas 和 Coss(2013a) 将蛇和蜥蜴配对，同样被作为威胁性刺激快速觉察的蜥蜴在遇到蛇时反应变慢了，他们认为蛇作为干扰物可能会使得非威胁性刺激或低威胁性刺激觉察变慢。但是，对于可以支持这个"注意脱离困难"的首次注视加工时间来说，在实验 1 和实验 2 的分析中都没有发现蛇和花之间有显著差异。即虽然蛇作为干扰物的时候会被更快觉察，但是并没有发现在首次加工上会存在注意脱离困难。类似地，LoBue 等人 (2014) 采用成人被试得出眼动结果也不支持"注意脱离困难"假设。作为固定时间分析来说，干扰物分析结论仍然需要谨慎，这也提示后续研究探讨干扰物属性的影响作用。

本研究由于幼儿被试限制，在数据采集、实验刺激上仍然存在改进地方，未来研究应该关注以下几个问题：第一，针对幼儿被试无法精确操作键盘和鼠标的情况，未来研究需要利用声音反应盒、特殊按键等设备精确记录幼儿的行为反应时间；在幼儿被试可以接受范围内，增大实验的试次。第二，恐惧性刺激诱发的情绪在其中扮演的作用 (Peira et al., 2010)。如果对蛇的觉察归结为其特殊外形，那么这种快速觉察就归结为基本的模式差别导致的反应差异。那么，蛇作为一类"恐惧动物"，它在儿童的快速觉察中有没有引发恐惧情绪？如果诱发了恐惧情绪，那么蛇的外形是否还起作用？第三，对蛇的快速觉察是否可以概括化到蜘蛛、蟑螂、狮子等其他动物恐惧中 (LoBue, 2010a; Penkunas & Coss, 2013a)。比如：以往很多研究都发现人类对蜘蛛这种特有的生物具有与蛇类似的注意偏向 (Blanchette, 2006; Öhman & Mineka, 2001)。第四，儿童经验对恐惧性刺激刺激觉察的影响。未来研究可以尝试选取更加广泛的年龄段探讨经验对蛇快速觉察的影响。第五，关于蛇的特殊外形导致了它的注意偏向和快速觉察。以往研究对蛇的颜色、生命性、体态特征、纹理等这些因素都没有很好控制，未来研究仍然需要进一步严格控制变量进行深入探讨。

参考文献

Berger, M. (2010). 'It's the sight not the bite': A model and reinterpretation of visually-based developmental fears. *Clinical Psychology Review, 30*(6), 779–793.

Blanchette, I. (2006). Snakes, spiders, guns, and syringes: How specific are evolutionary constraints on the detection of threatening stimuli? *The Quarterly Journal of Experimental Psychology, 59*(8), 1484–1504.

Bornstein, M. H., Mash, C., & Arterberry, M. E. (2011). Perception of object–context relations: Eye-movement analyses in infants and adults. *Developmental Psychology, 47*(2), 364–375.

Braun, K. A., Ellis, R., & Loftus, E. F. (2002). Make my memory: How advertising can change our memories of the past. *Psychology and Marketing, 19*(1), 1–23.

Brosch, T., & Sharma, D. (2005). The role of fear-relevant stimuli in visual search: A comparison of phylogenetic and ontogenetic stimuli. *Emotion, 5*(3), 360–364.

Cole, C., & Loftus, E. (1987). The memory of children. In S. Ceci, M. Toglia & D. Ross (Eds.), *Children's Eyewitness Memory* (pp. 178–208): Springer US.

DeLoache, J. S., & LoBue, V. (2009). The narrow fellow in the grass: Human infants associate snakes and fear. *Developmental Science, 12*(1), 201–207.

DeNicola, C. A., Holt, N. A., Lambert, A. J., & Cashon, C. H. (2013). Attention-orienting and attention-holding effects of faces on 4- to 8-month-old infants. *International Journal of Behavioral Development, 37*(2), 143–147.

Feiring, C., Lewis, M., & Starr, M. D. (1984). Indirect effects and infants' reaction to strangers. *Developmental Psychology, 20*(3), 485–491.

Feng, G. (2011). Eye tracking: A brief guide for developmental researchers. *Journal of Cognition and Development, 12*(1), 1–11.

Flykt, A., & Caldara, R. (2006). Tracking fear in snake and spider fearful participants during visual search: A multi-response domain study. *Cognition and Emotion, 20(8),* 1075–1091.

Fox, E., Griggs, L., & Mouchlianitis, E. (2007). The detection of fear-relevant stimuli: *Are guns noticed as quickly as snakes? Emotion, 7*(4), 691–696.

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, K. (2000). Facial expressions of emotion: Are angry faces detected more efficiently? *Cognition and Emotion, 14*(1), 61–92.

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General, 130*(4), 681–700.

Fredrikson, M., Annas, P., & Wik, G. (1997). Parental history, aversive exposure and the development of snake and spider phobia in women. *Behaviour Research and Therapy, 35*(1), 23–28.

Gerdes, A. B. M., Alpers, G. W., & Pauli, P. (2008). When spiders appear suddenly: Spider-phobic patients are distracted by task-irrelevant spiders. *Behaviour Research and Therapy, 46*(2), 174–187.

Gredeback, G., Johnson, S., & von Hofsten, C. (2010). Eye tracking in infancy research. *Developmental Neuropsychology, 35*(1), 1–19.

Isbell, L. A. (2006). Snakes as agents of evolutionary change in primate brains. *Journal of Human Evolution, 51*(1), 1–35.

Isbell, L. A. (2009). The fruit, *the tree, and the serpent: Why we see so well*. USA: Harvard University Press.

Keil, M. F., Briassoulis, G., Nesterova, M., Miraftab, N., Gokarn, N., Wu, T. J., et al. (2013). Threat bias in mice with inactivating mutations of Prkar1a. *Neuroscience, 241,* 206–214.

Koster, E. H. W., Crombez, G., Van Damme, S., Verschuere, B., & De Houwer, J. (2004). Does

imminent threat capture and hold attention? *Emotion, 4(3)*, 312 – 317.

Lipp, O. V., & Waters, A. M. (2007). When danger lurks in the background: Attentional capture by animal fear–relevant distractors is specific and selectively enhanced by animal fear. *Emotion, 7(1)*, 192 – 200.

LoBue, V. (2010a). And along came a spider: An attentional bias for the detection of spiders in young children and adults. *Journal of Experimental Child Psychology, 107(1)*, 59 – 66.

LoBue, V. (2010b). What's so scary about needles and knives? Examining the role of experience in threat detection. *Cognition & Emotion, 24(1)*, 180 – 187.

LoBue, V. (2013). What are we so afraid of? How early attention shapes our most common fears. *Child Development Perspectives, 7(1)*, 38 – 42.

LoBue, V., & DeLoache, J. S. (2008). Detecting the snake in the grass: Attention to fear–relevant stimuli by adults and young children. *Psychological Science, 19(3)*, 284 – 289.

LoBue, V., & DeLoache, J. S. (2010). Superior detection of threat–relevant stimuli in infancy. *Developmental Science, 13(1)*, 221 – 228.

LoBue, V., & Deloache, J. S. (2011). What's so special about slithering serpents? Children and adults rapidly detect snakes based on their simple features. *Visual Cognition, 19*(1), 129 – 143.

LoBue, V., & Larson, C. L. (2010). What makes an angry face look so?…?angry? Examining visual attention to the shape of threat in children and adults. *Visual Cognition, 18*(8), 1165 – 1178.

LoBue, V., Matthews, K., Harvey, T., & Stark, S. L. (2014). What accounts for the rapid detection of threat? Evidence for an advantage in perceptual and behavioral responding from eye movements. *Emotion, 14*(4), 816 – 823.

LoBue, V., Rakison, D. H., & DeLoache, J. S. (2010). Threat perception across the life span: Evidence for multiple converging pathways. *Current Directions in Psychological Science, 19*(6), 375 – 379.

Loftus, E. F., & Davies, G. M. (1984). Distortions in the memory of children. *Journal of Social Issues, 40*(2), 51 – 67.

Murray, E. J., & Foote, F. (1979). The origins of fear of snakes. *Behaviour Research and Therapy, 17*(5), 489 – 493.

Oakes, L. M. (2012). Advances in eye tracking in infancy research. *Infancy, 17*(1), 1 – 8.

Öhman, A. (2009). Of snakes and faces: An evolutionary perspective on the psychology of fear. *Scandinavian Journal of Psychology, 50*(6), 543 – 552.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. Journal of Experimental Psychology: *General, 130*(3), 466 – 478.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80*(3), 381 – 396.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*(3), 483 – 522.

Öhman, A., & Mineka, S. (2003). The malicious serpent: Snakes as a prototypical stimulus for an evolved module of fear. *Current Directions in Psychological Science, 12*(1), 5 – 9.

Öhman, A., Soares, S. C., Juth, P., Lindstr?m, B., & Esteves, F. (2012). Evolutionary derived modulations of attention to two common fear stimuli: Serpents and hostile humans. *Journal of Cognitive Psychology, 24*(1), 17 – 32.

Peira, N., Golkar, A., Larsson, M., & Wiens, S. (2010). What you fear will appear: Detection of schematic spiders in spider fear. *Experimental Psychology, 57*(6),

470－475.

Peltola, M. J., Lepp?nen, J. M., Vogel–Farley, V. K., Hietanen, J. K., & Nelson, C. A. (2009). Fearful faces but not fearful eyes alone delay attention disengagement in 7–month–old infants. *Emotion, 9*(4), 560－565.

Penkunas, M. J., & Coss, R. G. (2013a). A comparison of rural and urban Indian children's visual detection of threatening and nonthreatening animals. *Developmental Science, 16*(3), 463－475.

Penkunas, M. J., & Coss, R. G. (2013b). Rapid detection of visually provocative animals by preschool children and adults. *Journal of Experimental Child Psychology, 114*(4), 522－536.

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological Science, 17*(4), 292－299.

Purkis, H. M., & Lipp, O. V. (2009). Are snakes and spiders special? Acquisition of negative valence and modified attentional processing by non–fear–relevant animal stimuli. *Cognition & Emotion, 23*(3), 430－452.

Quinlan, P. (2013). The visual detection of threat: A cautionary tale. *Psychonomic Bulletin & Review, 20*(6), 1080－1101.

Rakison, D. H., & Derringer, J. (2008). Do infants possess an evolved spider–detection mechanism? *Cognition, 107*(1), 381－393.

Reynolds, M. G., Eastwood, J. D., Partanen, M., Frischen, A., & Smilek, D. (2009). Monitoring eye movements while searching for affective faces. *Visual Cognition, 17*(3), 318－333.

Shibasaki, M., & Kawai, N. (2009). Rapid detection of snakes by Japanese monkeys (Macaca fuscata): An evolutionarily predisposed visual system. *Journal of Comparative Psychology, 123*(2), 131－135.

Soares, S. C., Esteves, F., Lundqvist, D., & ?hman, A. (2009). Some animal specific fears are more specific than others: Evidence from attention and emotion measures. *Behaviour Research and Therapy, 47*(12), 1032－1042.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*(1), 15－48.

Vaish, A., Grossmann, T., & Woodward, A. (2008). Not all emotions are created equal: The negativity bias in social–emotional development. *Psychological Bulletin, 134*(3), 383－403.

Waters, A. M., Lipp, O., & Randhawa, R. (2011). Visual search with animal fear–relevant stimuli: A tale of two procedures. *Motivation and Emotion, 35*(1), 23－32.

Waters, A. M., Lipp, O., & Spence, S. H. (2008). Visual search for animal fear–relevant stimuli in children. *Australian Journal of Psychology, 60*(2), 112－125.

Wolfe, J. M., Yee, A., & Friedman–Hill, S. R. (1992). Curvature is a basic feature for visual search tasks. *Perception, 21(4)*, 465－480.

Children's attention detection to snakes: Evidence from eye movements

WANG Fuxing[1]; LI Wenjing[1];YAN Zhiqiang[2]; DUAN Zhaohui[1]; LI Hui[3]

([1] School of Psychology, Central China Normal University, Wuhan 430079 China)

([2] Department of Psychology, Peking University, Beijing 100871 China)

([3] Academy of Psychology and Behavior, Tianjin Normal University, Tianjin 300074 China)

Abstract   Previous research shows that preschool children detect snakes quickly than non−threating stimuli (e.g. flowers). In this study, we used eye tracking technology to provide direct evidences about the superior detection about threat−relevant stimuli. Two experiments were designed to testify whether the snakes would be fixated faster and quickly by preschool children and adults. In addition, we also used line drawing snakes and flowers as stimuli to control the shape of snakes and to testify the perceptual template hypothesis. In experiment 1, sixteen 4− to 6−year−old preschool children and 22 undergraduates were recruited as participants. A revised 3 × 3 matrices of color photographs of threat−relevant (snakes) and threat−irrelevant (flowers) stimuli were presented to both preschool children and adults. All participants were asked to find the threat target (snake) among seven non−threat distractors (flowers) and vice versa. Sixteen matrices with 8 pictures (1 target and 7 distractors) were presented to the participants. We changed the standard visual search task that did not present stimuli in the middle of the 3 × 3 matrices to control the central location effect and make the procedure appropriate for eye tracking calibration. It's a 2 (age: children, adults)× 2 (target: snake, flower) mixed design, and age was the between subject variable. In experiment 2, we improved the stimuli with line drawings to pop out the continuous curvilinear contour of snakes. The design, presentation method was the same as experiment 1. In two experiments, Tobii T120 Eye tracker was used to record the viewing behaviors of adults and children. The results of experiment 1 indicated that both the preschool children and adults fixated snakes faster and with less fixation counts than flowers, and their first fixation duration was shorter on snakes than flowers. Adults performed faster fixation, much less fixation counts than children. As distractors (flower was target), snakes were also fixated quickly than flowers (snake was target). In experiment 2, the same results were found that both children and adults located line drawing snakes quicker than line drawing flowers. And the first fixation duration to snakes were much shorter than flowers. Adults still fixated faster than children. For the snakes as distractors, we found the same results as experiment 1. Compared to the fixations of the line drawing snakes without color and pattern in experiment 2, real and colorful snakes in experiment 1 were fixated faster. The real and colorful snakes were detected faster and with less fixation counts before they were located as distractors. In conclusion, even preschool children who have little snake experience also show faster attention orienting and shorter attention holding. Based on the eye movements evidences, the continuous curvilinear shape plays an

important role in the snake relevant threat detection. The eye fixations of line drawing snakes provide direct evidence to the perceptual template theory. To be a special reptile, the color, patterns can boost the quick detection.

邻近效应对多媒体学习中图文整合的影响：线索的作用 *

王福兴，段朝辉，周宗奎，陈　珺

（青少年网络心理与行为教育部重点实验室；华中师范大学心理学院，武汉 430079）

摘　要　在多媒体学习中图文排列的邻近或远离会影响学习的效果。本研究操纵了多媒体呈现的图文排列位置，采用颜色作为线索引导注意，探讨了线索有、无和图片－文字的空间位置在多媒体学习图文整合中的影响。结果发现：图文邻近呈现的保持测验和迁移测验的成绩好于图文远离呈现，图文邻近呈现时学习者对关键兴趣区的进出次数更多，体现以出文字为导向的阅读模式。线索并没有单独影响学习成绩和注意分配；但结合图文位置后，发现有线索时图文邻近学习材料的注视次数、进出关键兴趣区的次数显著多于图文远离。研究认为：图文的邻近效应不仅提高学习者对知识的识记和理解，同时也影响了学习者的注意分配过程，图文邻近呈现有助于图片和文本信息的注意联接和知识整合；线索能使邻近效应得到凸显，更好地促进图文信息的注视加工，但没有改进学习的效果。

关键词　邻近效应；线索；眼动；图文整合

分类号　B849;G44

1　问题提出

多媒体学习 (multimedia learning) 是指对言语（口头语言和文字）和图片表征材料（图片、声音、视频等）的心理建构和加工 (Mayer, 2005)。随着科技的不断发展，多媒体学习材料在教学及其他领域也得到了普及。但是丰富多样的多媒体学习材料也同时存在信息量过大，学习者注意力被任务无关的信息捕获而不利于学习效果的情况 (Lowe, 1999, 2003)，所以如何正确地引导学习者的注意力分配，以更好地促进多

媒体学习中的图文整合效果是研究关心的重点。

1.1　邻近效应对图文整合的影响

已有研究发现，良好的图文编排方式能够有效地引导学习者的注意，并提高学习者的学习效果 (Schnotz, 2005)。为了防止图文结合呈现时学习过程中的注意分离（split–attention），研究者将图片信息和文本信息邻近排列而提高学习成绩的现象称为邻近效应（contiguity effect) (Ayres & Sweller, 2005; Mayer & Moreno, 1998; Sweller, Van Merrienboer, & Paas, 1998)。根据具体操作因素的不同，它又分为

两大类。一种是在空间位置上，图片信息和文本信息邻近排列而促进学习的现象，称为空间邻近效应（spatial contiguity effect）。另一种是在时间因素上，图片信息和文本信息同步编排而促进学习的现象，称为时间临近效应（temporal contiguity principle）(Mayer, 2001)。在研究中仅探讨空间邻近效应。

空间邻近效应最先由 Tarmizi 和 Sweller(1992) 提出，他们发现在数学问题解决的过程中，将材料信息结合呈现能减少加工材料的认知负荷，并提高学习的正确率。在 Mayer(2001) 的多媒体学习原则中，称其为空间邻近原则（spatial contiguity principle），认为文本与图片空间邻近呈现时，学习者可以减少搜索相关信息的时间，同时也能将搜索到的信息更多地保存在短时记忆中，这样的呈现效果比隔开呈现更能促进学生的学习 (Ayres & Sweller, 2005; Cierniak, Scheiter, & Gerjets, 2009; Mayer & Moreno, 1998)。Ginns(2006) 对 50 项有关多媒体学习邻近原则研究的元分析发现，邻近效应确实有利于学习成绩的提高或改进学习的效果 (d=0.85)。空间邻近效应在动画教学 (Kester, Kirschner, & Merriënboer, 2005)、书本阅读 (Mayer, Steinhoff, Bower, & Mars, 1995) 和报纸阅读 (Holsanova, Holmberg, & Holmqvist, 2009) 等研究领域中证明了其独特的优势。

在多媒体学习中，图文的空间位置变化确实会影响学习者的学习效果，但是其具体的加工过程是什么，却较少被研究所探讨。以往的一些研究中，研究者更多从认知负荷角度解释邻近效应。认为邻近排列的文字和图示减轻了学习者的认知负荷，从而有利于学习 (Cierniak et al., 2009; Moreno & Mayer, 1999)。Schmidt-Weigand, Kohnert 和 Glowalla(2010) 采用眼动技术探讨了通道效应和邻近效应的认知加工过程，眼动数据发现被试在学习有图 – 文内容的闪电原理时注视文字内容更多；对图片的注视时间在图文邻近情况下长于图文远离；注视图片时间与学习结果呈现正相关。Johnson 和 Mayer(2012) 研究发现，图文邻近呈现组在图片和文字间的注视点转换比图文远离时更加频繁。基于这些研究，本研究推测也许是图 –

文位置的邻近促进了学习者的视觉搜索，进而促进了图片知识和文本知识的整合，促进了学习的结果。这就需要记录学习者学习过程中的眼动数据，进一步验证是否图 – 文邻近有利于学习者的注意加工过程。

1.2 线索对图文整合的影响

在多媒体学习中，除了调整图文位置来加强图文整合外，在学习材料中适当地加入线索（cueing）也是引导注意力分配的一个常用方法。这里的线索是指通过操作指导材料视觉空间上的特征，从而达到帮助学习者注意相关信息以更好地把新知识与已有知识组织和整合成统一图式的目的 (综述见：De Koning, Tabbers, Rikers, & Paas, 2009)。线索的类型繁多，包括列举大纲 (topic shifts)、写标题 (outlines or headings)、箭头 (arrows)、颜色 (color) 等 (中文综述见：王福兴, 段朝辉, & 周宗奎, 2013)。综述以往研究发现，对多媒体学习注意引导最有效的线索是颜色 (Kalyuga, Chandler, & Sweller, 1999; Ozcelik, Arslan-Ari, & Cagiltay, 2010; Ozcelik, Karakus, Kursun, & Cagiltay, 2009) 和对比度变化 (De Koning, Tabbers, Rikers, & Paas, 2007, 2010)。并且考虑到教学设置的操作性和实用性，研究采用颜色作为线索，通过对学习材料中的关键信息进行颜色凸显以区别于其他信息的方式来引导注意加工、促进学习 (不同类型线索有效性评价见：王福兴等, 2013. p1431)。

综述以往研究发现，线索影响多媒体学习效果主要是通过三步进行。首先是引导学习者注意到线索区，从而引导学习者对知识的组织，最后帮助学习者进行新旧知识的整合 (De Koning et al., 2009; Mautone & Mayer, 2001)。虽然在对多媒体学习的研究中，颜色线索在学习钢琴运行原理 (Boucheix & Lowe, 2010)、飞机引擎的工作原理 (Ozcelik et al., 2010)、飞机的起飞原理 (Mautone & Mayer, 2001)、心血管循环系统 (De Koning et al., 2007) 和齿轮结构 (Boucheix & Guignard, 2005) 等知识学习上已经得到正面的验证，发现颜色线索确实能提高学习者对知识的理解，从而在理解或保持测验上得分更高。研究进一步采用眼动技术探讨颜色线索和空间邻近对多媒体学习注

意引导和学习效果的影响。

在多媒体学习中，线索即便有引导学习者注意分配的作用，但是能否有效地改善对知识的识记和理解却存在分歧。对线索研究的回顾发现当文本信息不与图片同时出现时，无论其作为启动信号出现在动画或图片信息的前面 (Crooks, Cheon, Inan, Ari, & Flores, 2012; De Koning et al., 2010; Fischer & Schwan, 2010; Kriz & Hegarty, 2007)，还是作为补充信息出现在动画或图片信息的后面 (Fischer & Schwan, 2010)，图片上设置的视觉线索都没有达到线索预期的效果。这也许是因为文本信息与图片信息的时间不同步影响了知识的组织和加工，使得线索即使达到了引导注意的目的，但是引导知识组织和整合的作用却没有显现 (参见：De Koning et al., 2009)。相反的，当文本线索与图片信息同步出现时，文本线索和视觉线索引导注意、组织和整合知识的功能都能很好地实现，学习者的学习成绩也得到了提升 (Boucheix & Guignard, 2005; Boucheix & Lowe, 2010; Mautone & Mayer, 2001; Ozcelik et al., 2010)。由此推测，线索对多媒体学习效果的影响很大程度上与图文位置的呈现有很大关系。因此，本研究以颜色作为线索进一步探讨线索是否在图文邻近效应的加工中起到调节作用。此外，将线索效应与邻近效应结合也有助于探索线索在什么条件下发挥作用，为有效使用线索提供依据。

基于以上论述，研究以多媒体学习中经典的"闪电的形成原理"作为学习材料 (Mayer, 2005; Mayer & Moreno, 1998; Moreno & Mayer, 1999)，采用文字与图片同时呈现的动画作为呈现形式，同时在文本和图形中加入颜色线索，来考查颜色线索和邻近效应在多媒体学习中对图文知识识记及其理解的影响。此外，借助于即时 (moment to moment) 记录和反应个体视觉认知加工和注意分配的眼动技术 (多媒体学习眼动研究评述见：Hyönä, 2010; Mayer, 2010; Scheiter & Van Gog, 2009; Van Gog & Scheiter, 2010)，探讨邻近效应和线索效应在图文整合中学习者的注意分配问题。以往关于空间邻近效应研究大部分都证明了

空间邻近要好于空间远离 (Cierniak et al., 2009; Ginns, 2006; Kester, Kirschner, & van Merriënboer, 2005; Mayer & Moreno, 1998)，所以研究预期假设认为：图文位置邻近排列更有利于学习，并且会促进学习者对学习材料的视觉搜索，即邻近排列的图文会让被试更多注视线索信息或文字信息。对线索综述也可以发现，颜色线索在很多研究中都能有效引导注意、促进知识整合 (Kalyuga et al., 1999; Ozcelik et al., 2010; Ozcelik et al., 2009)，研究预期学习材料中加入线索会促进学习效果，有线索的测验成绩要好于无线索。根据线索可以引导注意假设 De Koning et al., 2009; Mautone & Mayer, 2001，线索加入会凸显关键信息，被试对有颜色线索的信息注视会更多。对于线索和图文的物理排列位置相互作用，根据近期的眼动研究 (Johnson & Mayer, 2012; Schmidt–Weigand et al., 2010)，研究推测线索会更加凸显邻近效应，即位置邻近且有线索时最有利于学习者视觉搜索，从而促进学习的效果。

2 方法

实验设计为线索（无线索、有线索）和图文位置（远离、邻近）的两因素被试间设计，具体设计参见图 1。

2.1 被试

被试为某大学在校大学生，所有被试都经过经验水平前测问卷进行筛选。实验共招募被试 57 人，删除眼动数据异常的 6 人。最后有效被试一共有 51 人，随机分配到 4 种条件下。其中远离无线索组 13 人，远离有线索组 12 人，邻近无线索组 14 人，邻近有线索组 12 人。男生 14 人，女生 37 人，平均年龄为 22 岁 (SD=2.4)。所有被试的视力或者矫正视力正常，无色弱或者色盲。在实验结束后，每个被试得到一份小礼品或者课堂学分作为回馈。

因变量包括：（1）测验成绩：学习后的保持测验和迁移测验。（2）主观评定：动画难度评定和问卷难度评定。（3）眼动：兴趣区眼动注视指标。

2.2 材料

2.2.1 学习材料

实验学习材料来自 Mayer 在多媒体学习中使用的经典材料"闪电的形成原理"(Mayer, 2005; Mayer & Moreno, 1998; Moreno & Mayer, 1999)，具体刺激材料样式见下图1。动画文字部分内容从英文翻译成中文，为了确保文字翻译的准确性，研究者又请母语为英语、第二语言为汉语的美国留学生翻译回英文，翻译后的一致率为 85%。研究者将翻译不一致的地方与翻译者探讨后最终确定文字材料部分的中文内容。研究将学习材料制作成一段动态的动画呈现，动画素材格式为 800×600 像素的 AVI，由 16 幅图片素材和 16 段文字内容构成，整个动画持续时间为 155 秒。其中图片与文本描述在动画中一一对应，由电脑控制使图文同步呈现。线索设置采用红色作为标记 (Ozcelik et al., 2010)。其中线索组按照任务相关与任务无关 (Canham & Hegarty, 2010; Hegarty, Canham, & Fabrikant, 2010) 的划分标准，把文字区和图片区的相关兴趣区划为线索区，经评定，线索区的信息与保持测验的答案一致率达到 95% 以上。

2.2.2 前测问卷

气象知识前测问卷参照 Mayer 等人的研究编制 (Mayer & Moreno, 1998; Moreno & Mayer, 1999)，用来控制被试的经验水平，共 8 题：1）我定期阅览报纸或者网上的天气预报；2）我知道气相中冷锋的含义；3）我能区分积云与雨云的不同；4）我知道低气压系统是如何运行的；5）我知道刮风是如何形成的；6）我知道这个符号（冷锋标志）的含义；7）我知道这个符号（暖锋标志）的含义。其中每题按自己对知识的了解情况分"很少、较少、一般、较多、很多"五个维度计分，共 5 分。

2.2.3 保持测验和迁移测验

保持测验和迁移测验参照已有研究，并参照刺激材料文字部分对问题表述和评分点进行了翻译和回翻，确保准确性和一致性 (Mayer, 2005; Moreno & Mayer, 1999)。保持测验要求学生写出闪电的形成过
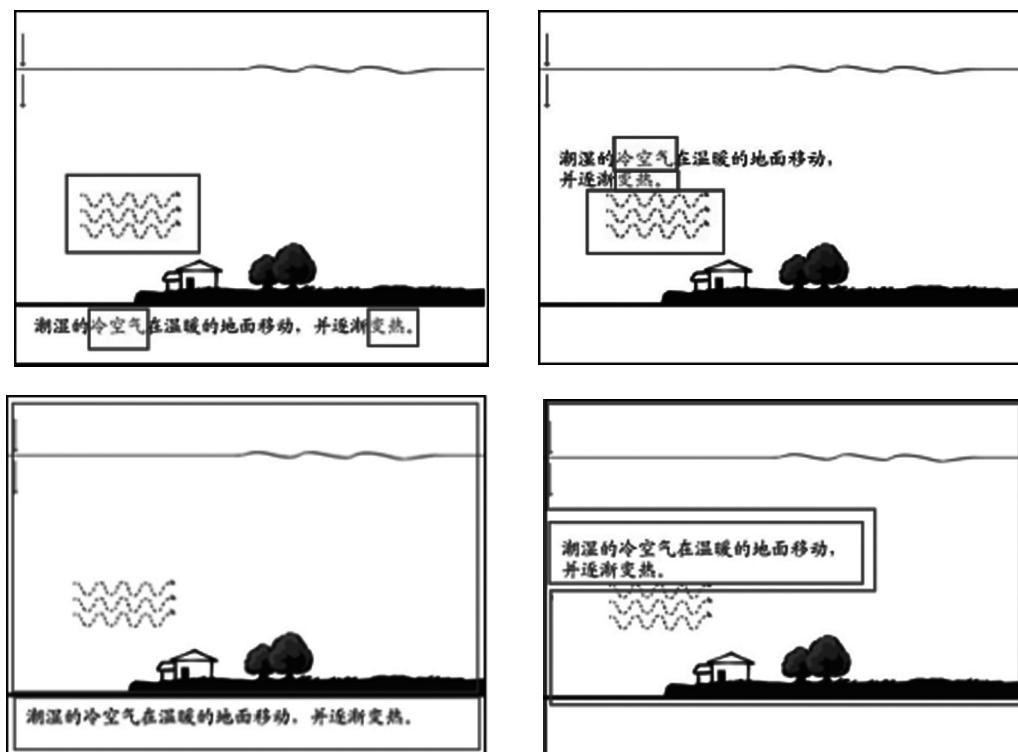
图 1　实验用动画材料截图

注：其中左上为有线索－远离，右上为有线索－邻近，左下为无线索－远离，右下为无线索－邻近；红框为眼动兴趣区标示，正式实验无红色边框，实际标记与此示意图有一定区别，仅供参考。

程，评分点共 19 个，满分 19 分。保持测验目的在于考察学习者对知识的识记能力。迁移测验由 4 个问题组成，包括"如何才能降低闪电的强度？"；"假设你看见天空中的云层却没有出现闪电，这是为什么？"；"气温和闪电是什么关系？"；"什么导致了闪电？"。其中每个问题有两个知识点，满分 8 分，迁移测验目的是测查把知识迁移并应用到新的情景中解决问题的能力。两个测验评分由 2 名经过培训的评分者完成，最后得分取 2 人的平均分。计算两个独立评分者的 Pearson 积差相关，保持测验为 0.94，迁移测验为 0.92（$p_s< 0.001$）。

2.2.4 主观评定量表

动画主观难度（1 表示非常简单，9 表示非常难）和测验难度（1 表示最简单，9 表示最难）评定均采用利克特 9 点评定（见附录 4 和附录 5）。

2.3 仪器与程序

眼动数据采集为 EyeLink 1000 Desktop 型眼动仪 (SR Research, Canada)，采样率 250 Hz，单眼瞳孔－角膜反射记录，19 英寸显示器（1024×768 分辨率，屏幕比例 5:4，眼睛距屏幕 75cm）。刺激材料的可视区域的水平视角 28.7 度，垂直视角 22.9 度。

实验开始前，对所有被试进行气象学知识前测筛选。筛选结束后向被试呈现指导语，告知实验内容及要求。被试明白后进行眼睛校准，然后进入正式实验并学习动画视频材料。实验结束后分别在规定时间内完成两份问卷：第一份问卷包括动画难度评定和保持测验，规定时间为 6 分钟；随后呈现第二份问卷，包括迁移测验和问卷难度评定，共 10 分钟。

3 结果

在前测知识问卷上，根据前人研究 (Mayer & Moreno, 1998; Moreno & Mayer, 1999)，所有参加实验被试的气象知识得分均未超过 3.8 分。各组的性别、年龄和知识经验得分，方差分析后结果如下（见表 1）。在性别上，线索（$F<1$, $p > 0.05$）和图文位置（$F<1$, $p>0.05$）及其交互作用（$F(1, 47) =1.70$, $p>0.05$, $\eta_p^2=0.04$，其中 η_p^2 为 partial η^2，下同）差异均不显著。在年龄上，线索（$F=1.09$, $p > 0.05$, $\eta_p^2=0.02$）、图文位置（$F<1$, $p > 0.05$）及其交互作用（$F<1$, $p > 0.05$）差异不显著。在知识经验得分上，线索（$F(1, 47) =1.20$, $p > 0.05$, $\eta_p^2=0.03$）、图文位置（$F(1, 47)=1.71$, $p > 0.05$, $\eta_p^2= 0.04$）及其交互作用（$F<1$, $p>0.05$）差异不显著。

3.1 保持测验和迁移测验得分

为了检验学生在不同的材料呈现形式及线索有无条件下对学习材料的识记和理解、应用，研究采用了保持测验和迁移测验，并以动画感知难度评定量表和问卷感知难度量表为补充，较为全面地对学习效果进行了检测（见表 1）。

对保持测验、迁移测验、动画感知难度和测验感知难度进行 2（线索：有线索；无线索）× 2（图文位置：远离；邻近）两因素被试间方差分析。发现在

表 1 各实验条件的问卷得分和线索区眼动数据

组别及变量		有线索				无线索			
		邻近		远离		邻近		远离	
		M	SD	M	SD	M	SD	M	SD
	性别	1.42	0.52	1.25	0.45	1.14	0.36	1.31	0.48
	年龄	22.33	2.19	21.92	2.11	21.50	2.95	21.31	2.39
	知识经验	2.63	0.74	2.47	0.58	2.51	0.75	2.17	0.59
	保持测验	8.83	2.04	7.75	2.56	9.07	2.02	6.85	2.30
	迁移测验	3.17	1.03	2.83	1.53	3.5	1.56	2.23	1.17
	动画感知难度	5.25	1.6	5.33	1.44	5.07	1.54	5.69	1.75
	测验感知难度	7.00	1.21	6.75	0.97	6.64	1.28	7.46	1.05
线索	DTP	0.09	0.02	0.10	0.02	0.09	0.01	0.09	0.01
	FCP	0.09	0.01	0.09	0.02	0.08	0.01	0.09	0.01
区眼动	RC	1.79	0.30	1.47	0.15	1.65	0.21	1.57	0.14

注：DTP 为兴趣区停留时间比例；FCP 为兴趣区的注视点个数比例；RC 为进出兴趣区的次数

保持测验上：图文位置的主效应显著（$F(1,47)=6.96$, $p < 0.05$, $\eta_p^2 = 0.13$），图文位置邻近的保持测验成绩高于图文位置远离的成绩（$MD=1.65$, $SD=0.63$）（注：MD 为 MeanDifference，下同）；线索的主效应（$F<1$, $p > 0.05$）和线索与位置交互作用（$F<1$, $p> 0.05$）均不显著。

在迁移测验上：图文位置的主效应显著 $F(1, 47)=4.50$, $p < 0.05$, $\eta_p^2=0.09$，图文位置邻近的迁移测验成绩高于图文位置远离的成绩（$MD=0.80$, $SD= 0.38$）；线索的主效应不显著 $F<1$, $p > 0.05$；图文位置和线索的交互作用不显著 $F(1, 47) = 1.54$, $p > 0.05$, $\eta_p^2 = 0.03$。

主观评定：在动画感知难度上，图文位置的主效应不显著 $F<1$, $p > 0.05$；线索的主效应不显著 $F<1$, $p > 0.05$；图文位置和线索的交互作用也不显著 $F<1$, $p > 0.05$。在问卷感知难度上，图文位置的主效应不显著 $F<1$, $p > 0.05$；线索的主效应不显著 $F<1$, $p > 0.05$；图文位置和线索的交互作用也不显著 $F(1, 47)=2.81$, $p>0.05$, $\eta_p^2= 0.06$。

3.2 图片和文字的兴趣区眼动数据

由于学习材料为动态视频，所以眼动数据处理时根据 16 幅图片素材将动画视频切分为 16 个片段，在每个片段上划兴趣区（图片、文字、线索），使用 EyeLink 自带的数据分析软件 Data Viewer 导出 16 个片段的眼动数据，最后再将 16 个片段平均后注视数据作为每个被试对兴趣区注视数据进入统计分析。为了探讨不同的线索及图文空间位置是否会影响学习者对学习材料中图片和文本的注意分配，及其相应的认知加工过程，本研究对四组动画材料进行了图片兴趣

区和文字兴趣区的划分，并确保不同图文位置材料的文字区面积大小和图片区面积大小差异不显著。选取的眼动指标有兴趣区停留时间百分比（Dwell Time Percent of AOI，简称 DTP，指注视停留在图片或文字区的时间与停留在整个动画上时间的比值，数值越大说明对图或文的加工越长）、兴趣区注视点个数（Fixation Count of AOI, 简称 FC，是指被试注视图片或文字区域的注视点数量，数值越大说明注视次数越多）以及在兴趣区的首次注视时间（First Fixation Time of AOI, 简称 FFT，指的是对图片区或文字区第一次注视加工的时间，反映了被试对图或文字的第一次加工时间）(有关眼动指标含义参见：闫国利等，2013)。这三个指标都可以说明学习者对兴趣区（图或文字）的关注程度，一般而言，值越大就代表对兴趣区的注意加工越多。由于 100 ms 是作为区分注视与其他眼动活动的界限 (Manor & Gordon, 2003)，所以研究把注视时间 100 ms 及以上的注视点纳入结果分析。对这些眼动数据进行了 2（线索：无线索、有线索）×2（图文位置：远离、邻近）×2（兴趣区：图片区、文字区）三因素混合重复测量方差分析（数据见表 2）。

在兴趣区停留时间比例（DTP）上，图文主效应显著 $F(1, 47) =141.47$, $p < 0.001$, $\eta_p^2=0.75$，文字比图片停留时间比更大（$MD = 0.34$, $SD=0.03$）。图文和位置的交互作用显著 $F(1,47) = 9.08$, $p < 0.01$, $\eta_p^2=0.16$，图文邻近时对文字注视的停留时间比高于远离 $F(1, 49) =8.84$, $p < 0.01$；图文远离时对图片的注视停留时间比例高于邻近时 $F(1,49)=7.63$, $p < 0.01$；图文和线索的交互作用不显著 $F(1,47)=2.16$, p

表 2　各实验条件下对图片 – 文字区的注视

兴趣区	图文位置	线索	DTP		FC		FFT (ms)	
			M	SD	M	SD	M	SD
文字	远离	有线索	0.58	0.13	19.08	3.59	76585	2509
		无线索	0.66	0.08	21.36	2.76	76644	2622
	邻近	有线索	0.71	0.11	22.93	3.14	72916	2951
		无线索	0.71	0.08	21.22	1.92	73090	2998
图片	远离	有线索	0.41	0.12	10.93	3.17	76462	2628
		无线索	0.32	0.09	7.80	1.24	77889	2861
	邻近	有线索	0.28	0.11	7.17	2.48	76676	6093
		无线索	0.28	0.09	7.13	2.00	71905	2843

注：DTP 为兴趣区停留时间比例；FC 为兴趣区的注视点个数；FFT 为兴趣区首次注视时长

$> 0.05, \eta_p^2 = 0.04$。图文、线索和位置三因素交互作用不显著 $F(1,47) = 2.68, p > 0.05, \eta_p^2 = 0.05$。位置的主效应不显著 $F(1,47) = 2.23, p > 0.05, \eta_p^2 = 0.05$。线索的主效应不显著 $F(1,47) = 1.52, p > 0.05, \eta_p^2 = 0.03$。位置和线索的交互作用不显著 $F < 1, p > 0.05$。

在兴趣区注视点个数（FC）上，图文主效应显著 $F(1,47) = 378.25, p < 0.001, \eta_p^2 = 0.89$，文字多于图片（$MD = 12.89, SD = 0.66$）。线索的主效应显著 $F(1,47) = 4.45, p < 0.05, \eta_p^2 = 0.09$，有线索比无线索多（$MD = 0.65, SD = 0.31$）。图文和位置的交互作用显著 $F(1,47) = 9.44, p < 0.01, \eta_p^2 = 0.17$，图文邻近时对文字注视多于图文远离 $F(1,49) = 4.32, p < 0.05$；图文远离时对图片注视多于图文邻近 $F(1,49) = 9.36, p < 0.01$。图文、线索和位置三因素交互作用显著 $F(1,47) = 7.15, p < 0.05, \eta_p^2 = 0.13$，线索和位置在文字下作用显著 $F(1,49) = 5.08, p < 0.05$，有线索的图文邻近对文字注视要多于远离；线索和位置在图片下边缘显著 $F(1,49) = 3.52, p = 0.065$，图文远离时有线索对图片注视多于无线索。位置的主效应不显著 $F < 1, p > 0.05$。位置和线索的交互作用不显著 $F < 1, p > 0.05$。图文和线索的交互作用不显著 $F(1, 47) = 1.99, p > 0.05, \eta_p^2 = 0.04$。

在兴趣区上的首次注视时长（FFT）上，图文主效应显著 $F(1,47) = 1.64, p < 0.05, \eta_p^2 = 0.03$，图片比文字长（$MD = 860, SD = 642$）。位置的主效应显著 $F(1,47) = 23.54, p < 0.001, \eta_p^2 = 0.33$，远离比邻近时间更长（$MD = 3210, SD = 662$）。位置和线索的交互作用显著 $F(1,47) = 5.63, p < 0.05, \eta_p^2 = 0.11$，位置邻近时有线索组比无线索组首次注视时间更长 $F(1,47) = 5.98, p < 0.05$。图文、线索和位置三因素交互作用显著 $F(1, 47) = 4.88, p < 0.05, \eta_p^2 = 0.09$，位置和线索在图片上的作用显著 $F(1, 49) = 8.17, p < 0.01$。线索的主效应不显著 $F(1, 47) = 1.56, p > 0.05, \eta_p^2 = 0.03$。图文和位置的交互作用不显著 $F < 1, p > 0.05$。图文和线索的交互作用不显著 $F(1, 47) = 1.49, p > 0.05, \eta_p^2 = 0.03$。

3.3 线索区的眼动数据

由上述结果看到，图文位置会影响学习者对图文信息的加工，并且线索也会在一定程度上影响这种图文整合。所以为了进一步探讨线索的影响机制，研究把材料中的线索区域作为兴趣区，对比分析邻近和远离情况下，有线索组和无线索组的眼动差异。选取的眼动指标有兴趣区停留时间百分比（DTP，含义同前）、兴趣区注视点个数百分比（Fixation Count percent of AOI，简称 FCP，是指对线索注视点个数占对整个动画注视点个数比例，值越大说明对线索注视和加工越多）和进出兴趣区次数（Run Count of AOI，简写为 RC，是指被试频繁的在线索和其他区域间来回注视次数，值越大说明对兴趣区关注越多）共 3 个指标，并进行 2(线索：有线索；无线索)×2(图文位置：远离；邻近)被试间分析（结果见表 1）。

在线索区停留时间百分比（DTP）上：线索的主效应不显著 $F < 1, p > 0.05$；图文位置的主效应不显著 $F < 1, p > 0.05$；图文位置和线索的交互作用不显著 $F < 1, p > 0.05$。

在线索区注视点个数百分比（FCP）上：线索的主效应不显著 $F < 1, p > 0.05$；图文位置的主效应不显著 $F < 1, p > 0.05$；图文位置和线索的交互作用不显著 $F < 1, p > 0.05$。

在进出线索区次数（RC）上：线索的主效应不显著 $F < 1, p > 0.05$；图文位置的主效应显著 $F(1, 47) = 11.74, p < 0.01, \eta_p^2 = 0.20$，图文邻近组多于图文远离组（$MD = 0.20, SD = 0.06$）；图文位置和线索的交互作用显著 $F(1, 47) = 4.16, p < 0.05, \eta_p^2 = 0.08$，有线索的情况下图文邻近（$M = 1.79, SD = 0.06$）多于图文远离（$M = 1.47, SD = 0.06$）。

4 讨论

本研究旨在探讨图文的空间邻近效应和线索对多媒体学习图文加工的影响，并探索颜色线索在空间邻近效应中所具有的作用。结果发现：图文位置会影响学习者对学习材料注意，图文邻近比图文远离在材料上的停留时间更长且注视次数

更多，在学习后的保持测验和迁移测验上的学习成绩也更好，这个结果与研究的预期假设一致。对图片和文字的眼动分析发现，线索只是影响了早期的注意引导，有线索的兴趣区注视点个数比无线索时更多，但线索条件在保持和迁移测验上却没有差异，没有验证预期的假设。当把图片不同位置进行线索标记后，结果发现图文邻近在有线索时得到更多的注视次数和更长的首次停留时间，在关键兴趣区的进出次数也更多，但是这种差异却没有体现在测验成绩上。

研究发现图片和文本邻近呈现的测验成绩要高于图文远离，这与前人的研究一致 (Ayres & Sweller, 2005; Mayer & Moreno, 1998; Moreno & Mayer, 1999; Tarmizi & Sweller, 1988)，即图片、文本在视觉空间和时间空间上的邻近呈现会有利于图－文信息加工，从而促进对知识理解和记忆。对眼动分析仅发现图文邻近排列的首次注视时长更长。仅对图文位置主效应分析看似与预期假设以及空间邻近效应解释不一致。但是，当把图片和文字兴趣区分析纳入后发现，图文位置排列影响到了对文字和图片注视分配，即当图文邻近排列时学习者对文字注视时间比例更高、注视点个数更多，而图文远离排列时候对图片注视时间和次数更高。空间邻近经典假设认为邻近排列的图－文将会减少学习者在图片和文本之间进行信息搜寻的时间，减少额外信息的干扰、节省认知资源，利于在图片和文本之间建立联结，从而提高学习效果 (Johnson & Mayer, 2012; Mayer, 2001; Tarmizi & Sweller, 1988)。从本研究结果看，图文的空间位置邻近效应主要是影响了对文字和图片注意分配导致了其学习结果差异。

这种在图文呈现方式的阅读模式中，对文本的更多关注被称为以文本为导向的阅读 (Holsanova et al., 2009)。这种多媒体学习中图文整合的文字导向的加工在 Schmidt-Weigand 等人 (2010) 研究中得到证实。这说明以文本为导向的阅读模式不仅在报纸阅读 (Holsanova et al., 2009) 上，儿童科学文本阅读（Hannus & Hyönä，1999）以及图文广告 (Rayner,

Rotello, Stewart, Keir, & Duffy, 2001) 中被证实，也出现在动态呈现的多媒体学习中，所以图文整合中的文本阅读模式对今后多媒体教学材料的个性化设计可能会有所启发。

此外，研究还发现图文邻近时，学习者首次停留在关键兴趣区的时间不仅更长，而且注意转换（进出兴趣区次数）也更多，这个结果与已有研究一致 (Johnson & Mayer, 2012)。进一步证实了图片和文本位置邻近确实影响学习者对信息的搜索和注意的分配，邻近呈现会减少信息之间的干扰，节省认知资源 (Johnson & Mayer, 2012)。Schmidt-Weigand 等人 (2010) 虽然发现被试对图片和文字注视会存在权衡，却没有发现对文本注视时间在图文邻近上有差异。本研究中不仅发现了保持和迁移测验上的空间邻近效应，在眼动数据上，也为空间邻近的假设（图－文邻近可能有利于学习者视觉搜索过程，进而提高了学习成绩）提供了直接数据支持。

研究并没有发现预期显著的线索作用，既没有发现学习测验上显著线索主效应，也没有发现在眼动指标上有显著主效应，仅在兴趣区注视点个数上发现有线索时候注视图文更多。De Koning 等人 (2009) 认为线索只有在实现注意导向（guiding attention）功能后，才能进一步实现它结构组织（organization of instruction）和知识元素整合（integrating elements into a functional model）的功能。线索这种自下而上影响认知加工的方式在本研究中没有被验证。综述以往研究可以发现，对于线索是否有效确实存在一定争议 (参见：王福兴，段朝辉，周宗奎，2013)，一部分研究确实没有发现线索有效引导注意并促进学习 (e.g. De Koning, Tabbers, Rikers, & Paas, 2011)；但是，也有一部分研究发现线索能够引导注意 (e.g. Boucheix & Lowe, 2010; De Koning et al., 2010; Ozcelik et al., 2010)。

意外的是，如果把图文位置和图片、文字兴趣区纳入后分析，发现线索可以使邻近效应更加凸显。即有线索时，学习者进出关键信息区的次数在位置邻近时会更多，对整个学习材料的注视次数也更多，

有线索的图文邻近对文字注视要多于远离，图文远离时有线索对图片注视多于无线索。虽然线索对邻近的突显作用不是在所有眼动指标上都有体现，但是，可以推测线索可能加速了学习者对关键信息的搜索，减少了额外的认知负荷，凸显了邻近效应。当然，这种促进作用并没有在学习测验上体现出差异，所以，仍然需要进一步实验来研究邻近效应对学习者的注意引导通过哪些因素起到调节作用。

对于线索在空间邻近上的作用，也需要再思考线索起作用的边界条件（boundary condition）(Mayer, 2010)。线索可能只有在学习材料结构非常复杂，学习者空间感知能力很低，并且先验知识经验很少的情况下才会发挥其积极作用 (De Koning et al., 2009)。本实验中线索之所以在学习效果上主效应不显著在眼动上部分交互显著，可能受到学习材料限制，因而使得线索的加入没有必要，也就没有起到帮助学习者对知识进行心理表征的作用 (Hegarty & Just, 1993)。另一个原因可能是学习者没有注意到动画材料中线索与关键信息的因果关系，因而对线索没有进行过多的关注，也就是说学习者即使知觉到了线索的外在表征，但是其内在的知识组织和整合的作用却没有体现出来。今后探讨线索的作用时，应该设法让学习者知觉到线索的存在，并提及线索在认知加工上的深层次功能。

本研究通过对图文邻近效应和线索的探讨，发现了线索作为材料设计的方式之一，对材料的结构及其复杂性有一定的要求。并且要想有效地实现线索的注意导向、结构组织、知识元素整合的功能，对学习者的空间整合能力及其先前知识经验也是有特定范围约束的。邻近效应作为图文空间变量，也具有注意导向和知识整合的功能。并且两者之间在特定情况下会共同对学习者进行影响。所以，未来研究可能需要考察学习者本身的知识经验等特征的影响；在被试的选取上遵循了以往研究范式，对被试经验进行了控制 (e.g. Boucheix & Lowe, 2010; Fischer & Schwan, 2010)。所以，未来研究需要特别关注被试已经具备的经验知识。

5 结论

最后，研究得出以下结论：（1）多媒体中的图文邻近呈现有助于图片和文本信息的注意联接和知识整合，从而促进基于多媒体的知识学习。（2）在图文整合的多媒体学习中，虽然图片占的面积更大，但是学习者表现出了文字导向的阅读。（3）当空间邻近中加入线索后，线索可能使图 – 文邻近的排列得到凸显从而促进图 – 文信息的搜索。

参考文献

Ayres, P., & Sweller, J. (2005). The split−attention Ayres, P., & Sweller, J. (2005). The split−attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 135–146). New York: Cambridge University Press.

Boucheix, J.−M., & Guignard, H. (2005). What animated illustrations conditions can improve technical document comprehension in young students? Format, signaling and control of the presentation. *European Journal of Psychology of Education, 20*(4), 369–388.

Boucheix, J.−M., & Lowe, R. K. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction, 20*(2), 123–135.

Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction, 20*(2), 155–166.

Cierniak, G., Scheiter, K., & Gerjets, P. (2009). Explaining the split−attention effect: Is the reduction of extraneous cognitive load accompanied by an increase in germane cognitive load? *Computers in Human Behavior, 25*(2), 315–324.

Crooks, S. M., Cheon, J., Inan, F., Ari, F., & Flores, R. (2012). Modality and cueing in multimedia learning:

Examining cognitive and perceptual explanations for the modality effect. *Computers in Human Behavior, 28*(3), 1063–1071.

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2007). Attention cueing as a means to enhance learning from an animation. *Applied Cognitive Psychology, 21*(6), 731–746.

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2009). Towards a framework for attention cueing in instructional animations: Guidelines for research and design. *Educational Psychology Review, 21*(2), 113–140.

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from a complex animation: Seeing is understanding? *Learning and Instruction, 20*(2), 111–122.

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2011). Attention cueing in an instructional animation: The role of presentation speed. *Computers in Human Behavior, 27*(1), 41–45.

Fischer, S., & Schwan, S. (2010). Comprehending animations: Effects of spatial cueing versus temporal scaling. *Learning and Instruction, 20*(6), 465–475.

Ginns, P. (2006). Integrating information: A meta–analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*(6), 511–525.

Hegarty, M., Canham, M. S., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. Journal of Experimental Psychology: *Learning, Memory, and Cognition, 36*(1), 37.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*(6), 717–742.

Holsanova, J., Holmberg, N., & Holmqvist, K. (2009). Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology, 23*(9), 1215–1226.

Hy?n?, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*(2), 172–176.

Johnson, C. I., & Mayer, R. E. (2012). An eye movement analysis of the spatial contiguity effect in multimedia learning. *Journal of Experimental Psychology: Applied, 18*(2), 178–191.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split–attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*(4), 351–371.

Kester, L., Kirschner, P. A., & Merri?nboer, J. J. (2005). The management of cognitive load during complex cognitive skill acquisition by means of computer–simulated problem solving. *British Journal of Educational Psychology, 75*(1), 71–85.

Kester, L., Kirschner, P. A., & van Merri?nboer, J. J. G. (2005). The management of cognitive load during complex cognitive skill acquisition by means of computer–simulated problem solving. *British Journal of Educational Psychology, 75*(1), 71–85.

Kriz, S., & Hegarty, M. (2007). Top–down and bottom–up influences on learning from animations. *International Journal of Human–Computer Studies, 65*(11), 911–930.

Lowe, R. K. (1999). Extracting information from an animation during complex visual learning. *European Journal of Psychology of Education, 14*(2), 225–244.

Lowe, R. K. (2003). Animation and learning: Selective processing of information in dynamic graphics. *Learning and Instruction, 13*(2), 157–176.

Mautone, P. D., & Mayer, R. E. (2001). Signaling as a cognitive guide in multimedia learning. *Journal of Educational Psychology, 93*(2), 377–389.

Mayer, R. E. (2001). *Multimedia learning*. New York:Cambridge University Press.

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20*(2), 167–171.

Mayer, R. E. (Ed.). (2005). *The Cambridge handbook of multimedia learning*. Cambridge: Cambridge University Press.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312–320.

Mayer, R. E., Steinhoff, K., Bower, G., & Mars, R. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development, 43*(1), 31–41.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358–368.

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior, 26*(1), 110–117.

Ozcelik, E., Karakus, T., Kursun, E., & Cagiltay, K. (2009). An eye-tracking study of how color coding affects multimedia learning. *Computers & Education, 53*(2), 445–453.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied, 7*(3), 219–226.

Scheiter, K., & Van Gog, T. (2009). Using eye tracking in applied research to study and stimulate the processing of information from multi-representational sources. *Applied Cognitive Psychology, 23*(9), 1209–1214.

Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). Explaining the modality and contiguity effects: New insights from investigating students' viewing behaviour. *Applied Cognitive Psychology, 24*(2), 226–237.

Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 49–69). New York: Cambridge University Press.

Sweller, J., Van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*(3), 251–296.

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80*(4), 424.

Van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*(2), 95–99.

Wang, F. X., Duan, Z. H., & Zhou, Z. K. (2013). Attention guidance in multimedia learning: The role of cueing. *Advances in Psychological Science, 21* (8), 1430–1440.

[王福兴 , 段朝辉 , 周宗奎 . (2013). 线索在多媒体学习中的作用 . *心理科学进展 , 21* (8), 1430–1440.]

Yan, G. L., Xiong, J. P., Zang, C. L., Yu, L. L., Cui, L., & Bai, X. J. (2013). Review of eye-movement measures in reading research. *Advances in Psychological Science, 21* (4), 589–605.

[闫国利 , 熊建萍 , 臧传丽 , 余莉莉 , 崔磊 , 白学军 . (2013). 阅读研究中的主要眼动指标评述 . *心理科学进展 , 21* (4), 589–605.]

The spatial contiguity effect in multimedia learning: The role of cueing

WANG Fuxing; DUAN Zhaohui; ZHOU Zongkui; CHEN Jun

(Key Laboratory of Adolescent Cyberpsychology and Behavior, Ministry of Education;

School of Psychology, Central China Normal University, Wuhan 430079, China)

Abstract　Text and illustrations integrated in spatial distribution could be helpful for learners' performance during multimedia learning. In addition, recent studies showed that cues, e.g. highlighting with color, arrows, bold typeface, could guide learner's attention and improve their learning outcomes. The researchers argued that the picture and text close to each other can shorten the visual search time and reduce the cognitive load, thereby enhancing the learning results. Previous studies also showed that adding cues to the learning materials could guide the learners' attention, promoted the organization and integration of the new knowledge. But what was the specific process of the contiguity effect? Whether the changes of the picture-text's location and adding cues would affect the allocation of attention? In this study, we expected that the contiguity effects and cueing would affect the learners' attention allocation. The integrated text and pictures with cues would have more fixation counts and longer dwell time on the task related area, and higher scores in the retention test and transfer test. In this study, fifty one college students were recruited as the participants. And a computer-generated animation depicting the process of lightning formation was used as the experiment material. Highlighting red color on text and pictures were manipulated as cues. First of all, a demographic questionnaire including a prior knowledge quiz would be sent to all of the prospective participants who want to participate in the experiment. The student who could be the participants had been measured by the prior knowledge quiz, to ensure they knew little about the lightning knowledge. After that they were randomized into four groups. The four groups were as follows: the integrated group, in which the texts were close to the pictures; the split group, in which the texts were away from the pictures; the cued group, in which the key information areas were highlighting; the un-cued group, in which there had no highlighting areas. They participants would watch the animation and learn this material. Eyelink 1000 was used to record their eye data. After that, they would have retention test and transfer test to test their learning results. The results showed that the students who learned the text pictures integrated had higher scores than the text picture split group on the retention test and transfer test. And the integrated group had more run counts and longer first fixation time on the task related area (highlighting cues) than the split group. Besides that, as to the text, the integrated group also had longer dwell time and more fixation counts, manifested a text-oriented reading mode. Although the main effect of cues were not so significant on the retention and transfer tests and eye-movement data, the interaction effect between cues and picture-text location were significant. The integrated group had more fixation counts and more run counts on the task related area when the text and pictures were cued with red color. In addition, the four groups have no difference

over their gender, age, the prior knowledge. In sum, these results indicated that the contiguity effects can not only improve learners' memorizing and understanding, but also help to allocate the attention during multimedia learning. The text close to the picture can benefit to integrate the information between pictures and text. What's more, cues can make that process much shorter and easier when the text and illustrations are integrated or contiguous in space.

装饰图片影响多媒体学习的眼动研究 *

龚少英 [1,2]，段　婷 [1,2]，王福兴 [1,2]，周宗奎 [1,2]，卢春晓 [1,2]

（[1] 青少年网络心理与行为教育部重点实验室 武汉 430079）

（[2] 华中师范大学心理学院　武汉 430079）

摘　要　为探讨装饰图片对多媒体学习效果和认知加工过程的影响，本研究采用眼动仪追踪 30 名低知识经验大学生在有装饰图片和无装饰图片条件下学习多媒体课件的视觉注意过程。结果发现：（1）有装饰图片组的保持和迁移成绩显著低于无装饰图片组；（2）有装饰图片组被试在认知兴趣图的注视次数、文本与认知兴趣图之间的注意转换次数显著少于无装饰图片组；（3）在装饰图片组内，80% 学习者报告被装饰图片吸引，并回忆出与装饰图片有关的先前知识经验。这些结果表明，装饰图片干扰学习者对主要学习内容的记忆与理解；装饰图片可能主要通过干扰学习者对主要学习内容的一致性理解以及激发不恰当的先前知识经验而阻碍学习。

关键词　多媒体学习；装饰图片；眼动

1　问题提出

装饰材料 (seductive details) 是指学习材料中插入的兴趣水平高、与学习主题有关联，但与学习目标无关的内容 (Garner, Gillingham, & White, 1989)。相对主要学习内容而言，装饰材料容易唤起学习者的兴趣（高兴趣水平），但它提供了与当前学习目标无关的信息，是在当前目标学习中不重要的内容（低重要性）(Lehman, Schraw, McCrudden, & Hartley, 2007)。

以往研究在关于装饰材料是否影响学习效果上存在很大争议。多数研究发现装饰材料干扰了学习 (Garner, et al., 1989; Harp & Mayer, 1998; Lehman, et al., 2007; Peshkam, Mensink, Putnam, & Rapp, 2011;Sanchez, Wiley, 2006；Myayer, Heiser & Lonn, 2001)，有装饰材料组学生在主要内容的回忆成绩和理解成绩上显著低于无装饰材料组。然而，也有一些研究 (Goetz & Sadoski, 1995; Sadoski, Goetz, & Fritz, 1993; Sanchez, Wiley, 2006) 发现装饰材料促进了学习，装饰图片能吸引学生更多地关注主要学习内容，进而加强对主要学习内容的记忆与理解。同时，也有研究发现，装饰材料对学习成绩并无显著影响 (Lusk, 2008)。关于装饰材料影响学习成绩的研究存在争议的主要原因可能是：（1）装饰材料的界定不一致。大部分研究 (Lehman, et al., 2007; Wade, Schraw, Buxton, & Hayes,

1993) 根据装饰材料的兴趣水平和重要性水平两个显著特征界定装饰材料，发现装饰材料干扰了学习。也有研究 (Lusk, 2008) 未按照装饰材料的特征严格界定装饰材料，发现装饰材料对学习无阻碍作用。（2）装饰材料的类型和呈现方式不一致。装饰材料干扰学习的情形主要为：装饰文本插入解释性文本或动画中 (Lehman, et al., 2007)、装饰文本和装饰图片同时插入解释性文本中 (Harp & Mayer, 1998)；而装饰材料促进学习或者不影响学习的主要情形为：装饰文本插入描述性的文本中 (Peshkam, et al., 2011)、装饰声音加入动画教学材料中 (Thalheimer, 2004)。（3）学习者的个体差异。工作记忆容量小或知识经验少的学习者其多媒体学习效果更易受装饰性材料干扰（Sanchez & Wiley, 2006；Magner, Schwonke, Aleven, Popescu & Renkl, 2012)。

多媒体学习的认知理论为解释装饰材料对多媒体学习效果的影响提供了理论基础。首先，根据多媒体学习认知理论的通道容量有限假设，每个通道在同一时间加工的信息数量有限。如果呈现的不同类型材料（如文字和图片）占用同一通道，可能导致容量超限，对学习产生干扰作用（Mayer, 2001）。与此同时如果学习过程中呈现与学习目标无关的内容，或者学习材料的呈现方式不当，将导致外部认知负荷增加，可用于加工学习内容和促进知识整合的相关认知负荷减少，从而对学习产生阻碍作用 (Mayer, 2001)。前述发现装饰材料干扰了文本学习的研究中，研究者将装饰文本或装饰文本加装饰图片插入解释性文本 (Lehman, et al., 2007；Harp & Mayer, 1998)，这些装饰材料干扰多媒体学习的一个可能原因是装饰文本或装饰文本加装饰图片和要学习的目标文本都占用视觉通道，导致通道容量超载，干扰了学习。此外，工作记忆容量小的学习者其学习容易受到装饰性材料干扰（Sanchez, Wiley, 2006)，可能也与装饰材料占用了其有限的资源，导致相关认知负荷减小有关。

其次，多媒体学习的认知理论提出有意义的学习包括选择、组织和整合三个主动加工阶段 (Mayer, Heiser, & Lonn, 2001)，为研究者探查装饰材料影响

多媒体学习的过程提供了理论基础。当学习内容以多媒体方式呈现时，学习者需要进行主动加工，包括选择相关的信息、将所选择的信息组织成一致的心理表征，最后将心理表征和已有知识经验整合成一个整体。为了探讨装饰材料具体影响学习的哪个阶段，研究者 (Harp & Mayer, 1998; Lehman, et al., 2007) 提出了相对应的三种假设：（1）注意减少假设 (reduced attention hypothesis)，是指装饰材料吸引学生更多注意，减少对主要学习内容的注意和阅读，从而导致对主要内容的记忆成绩下降；（2）一致性中断假设 (coherence break hypothesis) 是指装饰材料的加入导致学习者对学习材料的一致性理解中断，因而导致理解成绩下降；（3）注意转移假设 (diversion hypothesis)，后又被称为不恰当图式激活假设 (inappropriate schema hypothesis)，是指装饰材料会激活与当前的主要学习内容不太匹配的先前知识，阻碍学习者在新旧知识间建立恰当的联系，影响学习者对学习内容的全面加工，从而干扰记忆与理解成绩。关于装饰材料影响多媒体学习的上述三个假设都得到了研究证据的支持。Harp 和 Mayer(1998) 的研究主要采用间接测量的方法，通过学习成绩来推测装饰材料对学习中的认知加工过程的影响，其结果支持注意转移假设。Mayer 等人 (2008) 的研究通过探讨不同兴趣水平的装饰材料对学习成绩的影响间接探讨装饰材料对多媒体学习认知加工阶段的影响，研究结果支持一致性中断假设。为了更直接地检验三个研究假设，Lehman 等人 (2007) 在 Harp 和 Mayer(1998) 研究的基础上加入了测量每个句子的阅读时间这一客观的指标，结果发现，有装饰材料组对主要内容的阅读时间显著少于无装饰材料组，且对主要内容的回忆成绩和理解成绩均显著低于无装饰材料组，结果支持注意减少假设和一致性中断假设。

前述关于装饰材料影响多媒体的研究主要采用间接测量法探查了装饰材料对多媒体学习的影响，但对于装饰材料如何影响多媒体学习的过程还没有得到明确的结论，本研究欲采用直接的眼动记录法和间接的学习效果测量法，探查装饰材料对多媒体

学习过程及效果的影响 (Hyona, 2010)，揭示装饰材料影响多媒体学习效果的机制。眼动追踪研究为检验多媒体学习的认知理论提供了独特的方法 (Mayer, 2010)，可以为研究者揭示多媒体学习中的认知过程提供直接的证据。如总的眼睛注视时间反映被试对多媒体学习中的各部分的注意（Hyona, 2010），注视次数反映学习材料的认知加工负荷（闫国利等，2013），在文本和图片之间的眼动转换次数反映正在整合和理解所阅读的内容 (Holsanova, Holmberg, & Holmqvist, 2009)，使用被试眼动轨迹作为线索要求被试回忆当时所思所想来探查眼动背后的认知加工过程 (De Koning, Tabbers, Rikers, & Paas, 2010; Scheiter, Gerjets & van Gog, 2010) 等等。本研究欲将学习者在多媒体学习中的眼动轨迹、眼动轨迹回放报告和学习结束后的学习效果测量结合起来，揭示装饰图片对多媒体学习过程和学习效果的影响。本研究采用的装饰材料严格按其操作定义进行界定，满足高兴趣水平和低重要性两个标准。其次，装饰材料只使用单一的装饰图片，而不是混合的装饰材料。这主要是考虑到在多媒体学习中，采用图片作为装饰材料是一种典型的情况。因此，本研究先从单一的装饰图片类型入手进行研究。本研究中的眼动指标包括总的注视时间、注视次数、注视点比率、文本和认知兴趣图之间的眼跳次数等。本研究将从两方面对已有研究进行扩展：一是将直接的眼动追踪和间接的学习结果测量结合起来，既可以揭示在一定条件下装饰材料是否影响多媒体学习的效果，更可以揭示装饰材料是如何通过影响多媒体学习过程，从而影响最终学习效果的；二是使用眼动轨迹回放报告可以为解释眼动过程背后的认知机制提供证据。

根据前人研究和多媒体学习的认知理论，装饰材料虽然可以激发学习者的兴趣，但可能增加低知识经验者的外部认知负荷，影响其多媒体学习加工过程，从而影响多媒体学习的效果，因此，本研究提出如下假设：（1）装饰图片组在主要内容（文本区、认知兴趣图区）上的总注视时间、注视次数和注视比率以及在文本区与认知兴趣图间的眼跳次数将显著低于无装饰图片组；（2）在装饰图片组内，大部分被试将在学习后报告回忆出与装饰图片有关的先前知识经验。本研究采用眼动轨迹回放报告 (cued retrospective reporting) 的方式来验证此假设；（3）无装饰图片组的学习效果将好于有装饰图片组。

2　方法

2.1　被试

被试为某师范大学学生 30 名，其中男生 13 人，女生 17 人，平均年龄为 20.60 ± 1.67。被试的视力或矫正视力正常，无色盲色弱。根据前测（闪电形成原理的经验测验）结果，选择先前知识经验少（前测少于 10 分，满分 35 分）的被试，随机分配到两种实验条件下：无装饰图片组，有装饰图片组。

2.2　材料

先前知识经验测验是修订 Harp 和 Mayer(1997) 研究中的测验内容，共包括 4 个关于气象学知识熟悉度的主观评定题和 1 个关于闪电形成原理的问答题。其中，主观评定采用 0（完全不符合）到 4（完全符合）点评价，满分 16 分；问答题为："请写下你所知道的或想到的关于闪电形成的原理（即闪电是如何形成的），19 个计分点，答对一个计 1 分"。

学习材料：翻译并修订 Harp 和 Mayer(1997) 研究材料"闪电的形成过程与原理"，制成 PPT 课件的形式，共 5 页。其中，无装饰图片组学习材料仅有文本和认知兴趣图（重要性和兴趣水平都高的图片），有装饰图片组学习材料包括文本、认知兴趣图和装饰图片（兴趣水平高重要性低的图片）。学习时间为 5 分钟。

后测内容依次为：算术题、心理努力评价、认知负荷评价、再认测验、保持测验、迁移测验、眼动轨迹回放报告。保持测验和迁移测验题目是通过翻译并修订 Harp 和 Mayer(1997) 研究中采用的内容。心理努力和认知负荷采用 9 点计分量表进行测试，要求被试报告在学习课件时的努力程度，或者报告学习
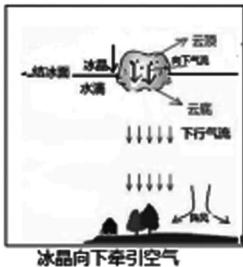
<p style="text-align:center">a（无装饰图片）　　　　　　b（有装饰图片）</p>
<p style="text-align:center">图 1　学习材料示例（图的上部为文本区，左下图为认知兴趣图，右下图为装饰图）</p>

内容的难度，分值越高，表示越努力或认知负荷越大。再认测验是 7 个多选题，答案完全来自所学课件；保持测验是让学生写出所记住的与闪电有关的所有内容；迁移测验是 4 个需要应用闪电原理来解决的问题，例如，如何才能降低闪电的强度。

前测和后测的计分：根据以往研究 (Harp & Mayer, 1998; Schmidt-Weigand, Kohnert, & Glowalla, 2010) 的评分标准进行修订后，保持测验内容为学习材料中的重要知识点，共有 19 个知识点，每回忆出 1 个知识点计 1 分。迁移成绩的评分则根据学习者回答的正确数目计分，每答对 1 点得 2 分，答错不扣分。再认测验的评分标准按学习者选择了多少正确答案计分，每选对一个选项计 1 分。所有成绩的评分均由两位经过严格培训并对评分标准非常熟悉的心理学研究生担任，评分者的一致性系数均在 0.95 以上。

为确保装饰图片符合装饰材料的操作定义，以及所选的学习材料难度适中，在正式实验之前，本研究另招募了 20 名被试进行预实验，要求学生在多媒体学习后对各部分学习材料（文本、认知兴趣图、装饰图片）的重要性（从 1~9 表示从一点也不重要至非常重要）和兴趣（从 1~9 表示从非常枯燥至非常有趣）进行评定。评定结果为：被试对装饰图片的兴趣程度显著高于文本 (p<0.001) 和认知兴趣图
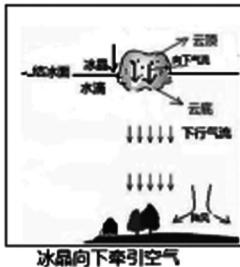
(p<0.05)，对认知兴趣图的兴趣程度显著高于文本区的兴趣程度 (p<0.001)；而在重要性程度上装饰图片显著低于文本 (p<0.001) 和认知兴趣图 (p<0.001)，而文本区和认知兴趣图在重要性上无显著差异 (p>0.05)，表明所选装饰图片满足装饰材料的操作定义。此外，预实验发现学习者学习每页 PPT 的时间约为 60 秒，因此在正式实验中将学习每页 PPT 的时间定为 1 分钟，以确保所有学生均能阅读完学习内容。

2.3 设计

本实验为单因素实验设计，自变量为装饰图片的类型（无装饰图片、有装饰图片），因变量为心理努力程度、认知负荷、再认测验成绩、保持测验成绩、迁移测验成绩以及眼动指标。其中具体的眼动指标为：兴趣区的总注视时间、注视次数、注视比率、兴趣区间眼跳次数与次序。

2.4 仪器与程序

本实验采用的仪器为 EyeLink 1000（SR Research, Canada) 眼动仪，采样率为 250 Hz。19 英寸显示器，分辨率为 1280×1024。实验材料的水平视角为 28.7 度，垂直视角 15.3 度，被试眼睛与屏幕之间的距离为 75cm。

实验开始前，对被试进行 9 点校准。然后，呈现实验指导语。在确认被试已熟悉实验任务并做好了

表 1　有无装饰图片组在后测成绩上的差异分析

	无装饰图片组 (n=13)		有装饰图片组 (n=15)		t	d
	M	SD	M	SD		
再认成绩	12.46	1.33	11.53	2.42	1.23	0.47
保持成绩	10.73	3.01	6.73	1.51	4.53**	1.68
迁移成绩	8.38	2.48	6.43	1.72	2.44*	0.91
心理努力	7.92	1.12	7.53	1.46	0.8	0.30
难度	4.15	1.41	4.80	1.66	−1.1	0.42

注：* $p<0.05$，** $p<0.01$，*** $p<0.001$，下同。

实验准备后，让被试开始学习 PPT 课件。学习结束后，先进行 5 分钟算术题测验，以排除短时记忆对学习成绩的影响。然后依次进行心理努力与认知负荷评价、再认测验、保持测验、迁移测验，最后要求有装饰图片组的被试根据回放的眼动轨迹报告学习过程中的思维内容。

2.5　结果

2.5.1　后测成绩分析

正式实验中有两名被试由于实验过程中头部移动程度较大，导致眼动追踪中断，剔除了这两名被试的所有数据。有、无装饰图片组的后测成绩差异分析结果（表 1）发现，无装饰图片组的保持成绩和迁移成绩显著高于有装饰图片组，但两组被试的再认成绩无显著差异。

2.5.2　心理努力和认知负荷

有、无装饰图片组被试的心理努力程度不存在显著差异，且两组被试的心理努力平均值大于 7 分，表明两组被试的心理努力均较高；有、无装饰图片组在学习难度上不存在显著差异，两组被试在任务难度的均值在 4–5 分之间，表明学习内容难度中等偏低。

2.5.3　有无装饰图片组在眼动指标的差异分析

对眼动数据进行分析前，先将每一页学习内容划分为三个兴趣区，分别是文本区、认知兴趣图、装饰图片。眼动数据分析结果见表 2。有、无装饰图片组在文本区和认知兴趣图的注视时间以及注视比率上均不存在显著差异。无装饰图片组在认知兴趣图的注视次数显著多于有装饰图片组，但在文本区的注视次数无显著差异。无装饰图片组在从文本区到认知兴趣图的眼跳次数 [$t(26)=3.38$，$p<0.01$，$d=1.27$] 以及从认知兴趣图到文本区的眼跳次数上 [$t(26)=3.46$，$p<0.01$, $d=1.30$] 都显著高于有装饰图片组。

2.5.4　眼动轨迹回溯报告结果分析

根据已有研究 (Harp & Mayer, 1998; Lehman, et al., 2007) 中的注意转移假设（或不恰当图式激活假设）：装饰材料激活了学生先前知识经验中与之有关的内容而阻碍学习。本研究将学习时被试回忆出与装饰图片有关而与闪电形成原理无关内容的现象界定为发生了注意转移，反之则为无注意转移。正式实验中，有装饰图片组的 15 名被试均参与了眼动轨迹回溯报告，其中有 12 人 (80%) 报告被装饰图片吸引，并回忆出与装饰图片有关的先前知识经验，有 3 人 (20%) 报告装饰图片与文字没什么关联，并未进行过多思考。

表 2　有、无装饰图片组在眼动指标上的差异分析结果

兴趣区	眼动指标	无装饰图片组		有装饰图片组		t	d
		M	SD	M	SD		
文本区	总注视时间 (ms)	37295	5250	37045	4950	0.13	0.05
	注视次数	155.83	24.38	145.83	26.40	1.04	0.39
	注视点比率	0.73	0.10	0.72	0.08	0.54	0.11
	文本向认知兴趣图的眼跳次数	5.28	2.24	2.69	1.80	3.38**	1.27
认知兴趣图	总注视时间 (ms)	12349	4332	10525	4322	1.12	0.42
	注视次数	54.85	22.81	38.91	15.26	2.20**	0.82
	注视点比率	0.25	0.10	0.19	0.07	1.83	0.69
	认知兴趣图向文本的眼跳次数	6.00	2.20	3.33	1.88	3.46**	1.30

3 讨论

3.1 装饰图片对多媒体学习效果的影响

本研究发现，有装饰图片组在保持成绩和迁移成绩上显著低于无装饰图片组，但在再认成绩上没有显著差异。两组被试在保持的差异效应量最大，在再认的差异效应量最小。这些结果部分验证了装饰图片对学习有阻碍作用的预期。这与以往的部分研究结果一致 (Harp & Mayer, 1997, 1998; Lehman, et al., 2007; Moreno & Mayer, 2000)，支持装饰图片对学习有干扰作用的假设。装饰图片不仅阻碍了学习者对主要学习内容的保持记忆，同时也阻碍了学习者对重要内容的理解和应用。装饰图片对多媒体学习的不同方面干扰作用不同，表明装饰图片很可能对简单的再认任务没有干扰作用，而对需要理解和应用的任务有显著干扰。相对来说，再认任务只需从四个答案中选择在文本中出现的信息，不需要深度加工，而保持任务需要被试在没有线索的情况下回忆出学习过的内容，迁移任务则需要被试将学习过的原理应用于新情境，需要将已学原理整合到已有的图式中，并提取出来加以运用，两者都需要学习者对所学内容进行深度加工。这些结果表明装饰图片可能主要影响学习者对内容的理解、保持和应用，而不是简单的再认学习。

但本研究结果与 Park 等人（2011）按照高兴趣水平和低重要性两个特征定义的装饰材料影响多媒体学习的研究结果不一致。在 Park 等人的研究中，装饰材料为装饰文本加装饰图片，装饰材料只在 11 页多媒体课件中的 4 页出现，被试评定的认知负荷为中等，学习效果的测量任务是保持、迁移、问题解决任务的混合，结果发现在主要学习内容为文本加认知兴趣图条件下有无装饰材料组的学习结果无显著差异。本研究和 Park 等的研究在装饰材料类型、装饰材料的呈现频率（是否在每页出现）和学习结果测量方式不同可能是造成这种差异的一个原因。其次，不管有无装饰材料，Park 等人研究中的被试在解说加认知兴趣图条件下的认知负荷显著高于文本条件

（两种条件下的认知负荷皆为中等水平），而且在解说和认知兴趣图加装饰材料条件下的学习效果显著好于其他三种条件（文本和认知兴趣图加装饰材料、文本和认知兴趣图、解说和认知兴趣图），出现了装饰材料逆转效应。这一方面表明解说和认知兴趣图加装饰材料条件下的积极学习效果伴随着更高水平的认知负荷，意味着认知负荷的适度增加并不必然对学习效果产生干扰作用；另一方面从不同通道呈现主要学习内容和装饰材料可能会减少装饰材料的消极作用，甚至起到促进作用。综合本研究和前人研究，装饰材料对多媒体学习的影响可能受到任务难度、装饰材料类型、装饰材料的呈现次数安排、主要内容和装饰材料的呈现通道以及个体因素等多方面因素的影响。装饰材料因素如何和其他因素相互作用影响多媒体学习的过程和效果还有待进一步深入系统的研究探查。

3.2 装饰图片对多媒体学习过程的影响

本研究采用眼动追踪技术实时记录学习者进行多媒体学习时的认知加工模式，结果发现，有、无装饰图片组在文本区的总注视时间、注视次数和注视比率以及在认知兴趣图的注视时间和注视比率上均不存在显著差异，仅在认知兴趣图的注视次数上，无装饰图片组显著高于有装饰图片组。因此，本研究结果并不完全支持注意减少假设。可能装饰图片本身只需很少的注意努力就能使学习者在短时内获得整体理解 (Shirey & Reynolds, 1988)，因而并不会占用过多的阅读时间和注视次数。学习者在有、无装饰图片条件下的心理努力和难度评价都没有显著差异，表明装饰图片的呈现没有给学习者增加显著的认知负荷，也为上述装饰图片可能只需要很少注意这一推论提供了额外支持。这一研究结果与 Mayer 等 人 (Harp & Mayer, 1998; Mayer, Griffith, Jurkowitz, & Rothman, 2008) 的结果一致，即装饰材料并不是通过减少对主要内容的注意而干扰学习；但与 Lehman 等人 (2007) 的研究结果（有装饰文本组对主要内容的阅读时间显著少于无装饰文本组）矛盾。主要原因可能是 Lehman 等人 (2007) 的研究

采用的是装饰文本，发现学习有装饰文本的学习者对基本文本的注意和阅读时间减少。而与文字不同的是，图片在获取信息上具备独特的高效性和丰富性 (Dansereau & Simpson, 2009; 沈德立 & 陶云，2001)，装饰图片相对于装饰文本而言在短时间内容易理解且可以获得整体信息而不需要花费过多注意来精加工，因而装饰图片并未导致对主要内容的注视时间显著减少。

虽然两组被试在文本区的注视时间、注视次数和注视点比率无显著差异，但和认知兴趣图相比，两组被试都在文本区有更多的注视次数、注视时间和注视点比率。这表明学习者优先注意文本而不是图片，与以往研究发现的文本优先趋势是一致的（Schmidt – Weigand, Kohnert, & Glowalla, 2010）。

眼动数据分析发现无装饰图片组在文本区与认知兴趣图之间的眼跳次数显著高于有装饰图片组。被试在文本与图片之间的眼跳（又称为注意转换）反应了被试试图在语义上连接文本和图形信息形成整合性理解的程度 (Holsanova, et al., 2009)。这一结果表明在无装饰图片的条件下，学习者可能更多地在文本与认知兴趣图上进行组织和深加工，而当遇到装饰图片时，学习者对主要学习内容的整合加工次数显著减少，从而导致保持和理解成绩下降，这一发现支持一致性中断假设。

眼动轨迹回放报告分析结果表明，有 80% 的学习者（12 名被试）在看到装饰图片时想起了与图片有关的内容（例如：有的学习者看到闪电中的自由女神像，当时的心理活动是"自由女神像在哪个国家？建筑师是谁？"）。研究发现，仅有 20% 的学习者（3 名被试）报告"仅仅是瞟了一眼，发现有趣但与学习任务无关，于是没有详细浏览"，即并没有激活与装饰图片有关的先前知识经验。这一方面说明，装饰图片是否对多媒体学习产生影响与学习者的抑制干扰能力有关。装饰图片对学习的影响程度存在个体差异，学习者在学习过程中的自主调节学习影响了装饰图片对学习的作用方向与程度。更重要的是，在外界材料足够吸引学习者的情境下，大部分

学习者被装饰图片所吸引，并回忆起了与装饰图片有关的内容。可见，总的来看，装饰图片可能激发了学习者先前知识经验中相关的内容而干扰学习。眼动轨迹回放报告的结果支持不恰当图式激活假设。综上，本研究结果主要支持一致性中断假设与不恰当图式激活假设，表明装饰图片对多媒体学习的干扰主要发生在组织和整合阶段。对于装饰图片是否影响多媒体学习中的选择过程、什么样的学习者（不）会产生注意转移，以及认知负荷增加对学习效果的影响受到哪些因素的调节等问题还有待未来进一步的研究探查。

本研究将直接的眼动轨迹和间接的学习表现结合起来探查装饰图片对多媒体学习过程和效果的影响，揭示了装饰图片影响多媒体学习过程中的组织和整合阶段，表明装饰图片对多媒体学习的不同阶段产生阻碍作用，为多媒体学习认知理论解释装饰材料效应提供了基于过程和结果的证据，这是本研究的一个特色。另一个特色是本研究将实时的眼动轨迹指标和事后的眼动轨迹回放报告结合起来，初步揭示了在包含有装饰图片的多媒体学习中眼动背后的认知加工过程。但是，本研究也存在一些局限。首先，本研究的结果来自于先前知识经验较少的被试，学习材料的难度中等，对于具有较多先前知识经验的学习者，或者学习不同难度的内容，情况是否也如此，还有待进一步探查。其次，虽然本研究揭示了对于低先前知识经验的学习者来说，在学习内容中插入装饰图片对其学习过程和效果产生了干扰，但未来还有待进一步探查不同类型装饰材料（如文本、声音、图片、动画）或不同类型装饰材料的组合对多媒体学习的影响，深入系统地揭示装饰材料影响多媒体学习的机制。

4 结论

（1）本研究发现在以科学主题为学习内容的多媒体学习情境中，装饰图片阻碍学习者对主要学习内容的记忆与理解；

（2）装饰图片主要通过干扰低知识经验学习者对主要学习内容的一致性理解以及激发学习者不恰当的先前知识经验而阻碍学习者对主要学习内容的记忆与整合性理解。

参考文献

沈德立 , & 陶云 . (2001). 初中生有无插图课文的眼动过程研究 . *心理科学 , 24*(4), 385 – 389.

闫国利，熊建萍，臧传丽，余莉莉，崔磊，白学军（2013）。阅读研究中的主要眼动指标评述。*心理科学进展，21*（4）， 589–605

Dansereau, D. F., & Simpson, D. D. (2009). A picture is worth a thousand words: The case for graphic representations. *Professional Psychology: Research and Practice, 40*(1), 104 – 110.

De Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Attention guidance in learning from complex animation: Seeing is understanding? *Learning and Instruction, 20*(2), 111 – 122.

Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of "seductive details" on macroprocessing and microprocessing in adults and children. *Cognition and Instruction, 6*, 41 – 57.

Goetz, E. T., & Sadoski. (1995). The perils of seduction: Distracting details or incomprehensible abstractions? *Reading Research Quarterly, 30*(3), 500 – 511.

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional and cognitive interest. *Journal of Educational Psychology, 89*, 92 – 102.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*(3), 414 – 434.

Holsanova, J., Holmqvist, K., & Holmberg, N. (2009). Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology, 23*(9), 1215 – 1226.

Hyona, J. (2010). The use of eye movements in the study of mutimedia learning. *Learning and Instruction, 20,*(2), 172–176

Lehman, S., Schraw, G., McCrudden, M. T., & Hartley, K. (2007). Processing and recall of seductive details in scientific text. *Contemporary Educational Psychology, 32*(4), 569 – 587.

Lusk, D. L. (2008). The effects of seductive details and segmentation on interest, recall and transfer in a multimedia learning environment. Doctoral *Dissertation. Virginia Polytechnic Institute and State University.*

Magner, U. I. E, Schwonke, R., Aleven, V., Popescu, O., Renkl, A. (2014). Triggering situational interest by decorative illustrations both fosters and hinders learning in computer–based learning enviorments. *Learning and Instruction, 29*, 141 – 152

Mayer, R. E. (2001). Multimedia learning. New York: *Cambridge University Press.*

Mayer, R. E., Griffith, E., Jurkowitz, I. T. N., & Rothman, D. (2008). Increased interestingness of extraneous details in a multimedia science presentation leads to decreased learning. *Journal of Experimental Psychology: Applied, 14*(4), 329 – 339.

Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187 – 198.

Mayer, R. E.(2010). Unique contributions of eye–tracking research to the study of learning with graphics. *Learning and Instruction, 20*, 167–171

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*(1),

117–125.

Paas, F., Renkl, A., Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*(1), 1–4

Park, B., Moreno, R., Seufert, T., & Brünken, R. (2011). Does cognitive load moderate the seductive details effect? A multimedia study. *Computers in Human Behavior, 27*(1), 5–10.

Peshkam, A., Mensink, M. C., Putnam, A. L., & Rapp, D. N. (2011). Warning readers to avoid irrelevant information: When being vague might be valuable. *Contemporary Educational Psychology, 36*(3), 219–231.

Sadoski, M., Goetz, E. T., & Fritz, J. B. (1993). Impact of concreteness on comprehensibility, interest, and memory for text: Implications for dual coding theory and text design. *Journal of Educational Psychology, 85*(2), 291–304.

Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & cognition, 34*(2), 344–355.

Schmidt – Weigand, F., Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system and self–paced instruction in multimedia learning. *Learning and Instruction, 20*(2), 100–110.

Shirey, L. L., & Reynolds, R. E. (1988). Effect of interest on attention and learning. *Journal of Educational Psychology, 80*(2), 159–166.

Thalheimer, W. (2004). Bells, whistles, neon, and purple prose: When interesting words, sounds, and visuals hurt learning and performance, Retrieved June 10, 2011, *from http://www.oktopusz.hu/domain2019/ files/ modules/module*2015/28283C28732CAE28682.pdf.

Wade, S. E., Schraw, G., Buxton, W. M., & Hayes, M. T. (1993). Seduction of the strategic reader: Effects of interest on strategies and recall. *Reading Research Quarterly, 28*(2), 93–114.

The effects of seductive illustrations on multimedia learning: An eye movement study

GONG Shaoying[1,2] ; DUAN Ting[1,2] ; WANG Fuxing[1,2] ; ZHOU Zongkui [1,2] ; LU Chunxiao [1,2]

（ [1] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan, 430079 ）

（ [2] School of Psychology, Central China Normal University, Wuhan 430079 ）

Abstract This study aimed to explore the effects of seductive illustrations on performance and cognitive processes by using eye movement to track thirty low prior experience learners' process of visual attention during multimedia learning with or without seductive illustrations. Findings revealed that relative to learners in the non−seductive illustration condition, learners in the seductive illustration condition (1) recalled significantly fewer main ideas and generated fewer problem− solving solutions, (2) got fewer number of fixations in cognitive interest illustrations and also fewer attention switches between text and cognitive interest illustrations. Moreover, in the seductive illustration condition, 80% learners reported attracted to seductive illustrations, and recalled prior knowledge related to seductive illustrations. The result indicates that seductive illustrations may interfere students' memory and understanding of the main learning content by disrupting the coherent understanding of the materials and by priming inappropriate prior knowledge around seductive illustrations.

Keywords multimedia learning; seductive illustrations; eye−movement

实习教师课堂情绪与教师效能感的关系
——情绪调节的中介作用

龚少英，张扩滕，张盼盼，李佳蔚

摘　要　以华中师范大学214名实习教师为研究对象，通过问卷调查法探究了实习教师课堂情绪与效能感之间的关系，以及情绪调节策略对这一关系的影响。研究结果表明，教师在实习期间的积极课堂情绪体验对于教师效能感有促进作用，消极的课堂情绪体验对于教师效能感有阻碍作用。认知重评在 课堂情绪体验与教师效能感的关系中起中介作用，表达抑制中介作用不显著。

关键词　积极情绪；消极情绪；认知重评；表达抑制；教师效能感

1　问题提出

教学不仅仅是传授知识的认知性的活动，还是一种情绪性的活动。比如，当课堂教学进展顺利或者学生能够理解教学内容时，教师就会感到高兴；当学生没有理解所教内容时，教师就会感到挫败；当教师无力应对来自家长以及上级的各种要求时，就会产生焦虑。

Bandura 指出，情绪是自我效能感的来源之一，即个体会依据情绪对自我效能感进行判断。积极情绪就意味着个体有能力应对当前情境，这有助于提高个体自我效能感；消极情绪会减弱个体的自我效能感。Kavanagh 与 Bower 的研究为 Bandura 的观点提供了支持。那么，教师在课堂教学中的情绪是否也会影响其效能感，目前还很少有研究关注这一问题。

情绪不会自发地影响教师效能感，为了使教学任务顺利完成、保持"好教师"形象，教师会使用不同的策略来调节他们的情绪。在众多情绪调节策略中，认知重评 (cognitive reappraisal) 和表达抑制 (expressive suppression) 受到研究者更多的关注。认知重评是指个体通过改变对客体或事件的评价来调节他们的情绪，比如，有的学生上课分心，教师就会想，这个同学可能是身体不舒服，这样教师就不会产生愤怒的情绪；表达抑制是指个体为了实现一定的目标而抑制情绪表达行为，比如，当学生上课出现违纪行为，教师感到十分生气，但是为了教学任务的顺利完成以及不干扰其他同学，教师就会抑制住愤怒的情绪。已有研究表明，相对于表达抑制，认知重评是一种更加有效的情绪调节策略。当个体对消极情绪进行调节时，认知重评不仅能够减少情绪表达行为，而且还能减弱个体的情绪体验；表达抑制仅仅能够减少个体的消极情绪表达行为，但是对于个体的消极情绪体验没有影响；当个体对积极情绪进行调节时，认知重评能够根据情境的需要增强或减弱个体的主观体验、生理反应以及表达行为；表达抑制不仅能够减少个体的积极情绪表达行为，由于个体在抑制过程中，需要付出一定的心理努力，这就导致个体的生理反应增强，以及积极情绪体验的减弱。

在教育领域，还没有研究者来探究实习教师课堂情绪、教师效能感以及习惯采用的情绪调节策略之间的关系。虽然已有的研究表明，积极情绪能够促进自我效能感，而消极情绪对自我效能感具有阻碍作用，但是这些研究并未考虑到个体习惯采用的情绪调节方式的影响作用。由于认知重评对个体情绪体验产生影响，而表达抑制仅仅改变个体的表达行为，对个体情绪体验的影响较小，由此可以推论，认知重评在实习教师课堂情绪体验与教师效能感之间可能起中介作用，而表达抑制在两者之间中介作用可能不显著。因此，本研究以实习教师为研究对象，通过问卷调查法，来探究教师课堂情绪体验与效能感之间的关系，以及情绪调节方式对于这一关系的影响作用。

基于以上论述，本研究提出以下假设：第一，课堂积极情绪体验促进实习教师效能感；第二，课堂消极情绪体验会阻碍实习教师效能感；第三，认知重评在实习教师课堂情绪体验与效能感之间的关系中起中介作用，而表达抑制在这二者之间中介作用不显著。

2 研究方法

2.1 调查对象

本研究选取华中师范大学实习教师214名，其中男生96名；年龄分布在18~28岁之间，平均年龄为21.9岁；所有被试均在高中实习，其中在重点学校实习的有186人，非重点学校实习的有26人，2人没有填写实习学校信息；183名被试在实习期间担任过班主任；实习教师所教科目主要包括数学、英语、地理、历史、信息技术等。

2.2 研究工具

2.2.1 教师课堂情绪体验量表

在参照已有文献的基础上，本研究选取了快乐、喜爱、惊讶、生气、焦虑、紧张、失望、挫败、害怕、伤心、厌恶、无助12种情绪，分为两个因素，即前3个项目测的是积极情绪，后9个项目测的是消极情绪。本量表采用5点计分法，分数越高，表示体验到该情绪的频率越高。在本研究中，整个量表的内部一致性系数为0.81。

2.2.2 教师效能感量表

该量表是由Tschannen-Moran和Hoy编制的，原量表有24个项目，包含学生投入、教学策略、班级管理这三个维度。该量表采用9点计分法，分数越高，表示教师效能感就越高。国内研究者龙君伟等人对其进行了修订。考虑到原量表的某些项目不适用于中国文化，因此根据量表编制者的建议删除了一个项目。在以职前教师为被试的研究中，Tschannen-Moran和Hoy发现，该量表测的是一个因素，即教师效能感。因此在本研究中，也将本量表视为单因素结构，该量表的内部一致性系数为0.95。

2.2.3 情绪调节方式问卷

该问卷是由王力等人修订的，该量表共包含14个项目，测量认知重评和表达抑制的项目各7个。该量表采用7点计分法，分数越高，表示越倾向采用该策略调节情绪。在本研究中，测量认知重评和表达抑制的项目的内部一致性系数为0.88和0.80。

3 结果

3.1 差异检验

3.1.1 实习教师积极情绪与消极情绪体验的差异

对实习教师在课堂教学中体验到的积极情绪与消极情绪平均得分进行配对样本t检验，发现实习教师在课堂教学过程中体验到积极情绪的频率显著高于体验到消极情绪的频率 ($t=20.99$, $p<0.01$)。

3.1.2 实习教师情绪调节策略差异

对实习教师在认知重评与表达抑制这两种情绪调节策略的平均得分进行配对样本t检验，发现相对于表达抑制，实习教师更习惯采用认知重评策略调节情绪 ($t=11.12$, $p<0.01$)。

3.1.3 不同性别实习教师在各变量上的差异（见表1）

对不同性别实习教师在各变量上的均数差异进

表 1　不同性别实习教师在各变量上的均值差异检验

变量	性别	M(SD)	T
积极情绪	男	3.54(0.53)	−0.54
	女	3.58(0.51)	
消极情绪	男	2.30(0.66)	0.06
	女	2.29(0.62)	
认知重评	男	4.90(0.91)	−1.71
	女	4.92(0.87)	
表达抑制	男	4.39(0.93)	3.12**
	女	4.00(0.90)	
教师效能	男	6.51(1.03)	−0.50
	女	6.58(0.99)	

注：* $p<0.05$，** $p<0.01$. 下同

行独立样本 t 检验，发现仅在表达抑制这一情绪调节方式上差异显著，男实习教师比女实习教师更多地使用表达抑制策略（$t=3.12$, $p<0.01$）。

3.1.4　不同班主任经验实习教师在各变量上的均值差异比较（见表 2）

对担任过与没有担任过班主任的实习教师在各变量上的均数差异进行独立样本 t 检验，发现担任过班主任的实习教师在课堂教学中较多地体验到积极情绪（$t=2.16$, $p<0.05$），并且报告出更高水平的效能感

表 2　不同班主任经验实习教师在各变量上的均值差异检验

变量	班主任	M(SD)	T
积极情绪	是	3.59(0.50)	2.16*
	否	3.38(0.55)	
消极情绪	是	2.26(0.66)	−1.73
	否	2.48(0.47)	
认知重评	是	4.95(0.89)	1.39
	否	4.71(0.84)	
表达抑制	是	4.18(0.95)	0.51
	否	4.09(0.83)	
教师效能	是	6.62(0.97)	2.43*
	否	6.15(1.13)	

注：* $p<0.05$，** $p<0.01$. 下同

（$t=2.43$, $p<0.05$）。

3.2 课堂情绪体验、情绪调节方式与教师效能感的相关分析

对实习教师课堂情绪体验、情绪调节方式、教师效能感进行皮尔逊积差相关分析，结果见表 3。实习教师效能感水平高于中位数 5，这与 Hoy 和 Spero

的研究结果一致；积极情绪、认知重评与表达抑制与教师效能感之间呈显著正相关，而消极情绪与教师效能感之间呈显著负相关；认知重评与积极情绪呈显著正相关，与消极情绪呈显著负相关；表达抑制

表 3　各变量间相关分析结果

变量	M(SD)	1	2	3	4	5
1. 积极情绪	3.56(0.51)	1				
2. 消极情绪	2.29(0.64)	−0.16*	1			
3. 认知重评	4.91(0.88)	0.24**	−0.27**	1		
4. 表达抑制	4.17(0.93)	0.22**	0.00	0.42**	1	
5. 教师效能	6.55(1.01)	0.25**	−0.33**	0.54**	0.21**	1

与积极情绪呈显著正相关，与消极情绪相关不显著。

3.3 回归分析

本研究采用层级回归分析的方法，以教师效能感为因变量，检验教师课堂情绪体验、情绪调节方式对教师效能感的影响。首先将性别、年龄、所教科目、是否担任班主任、学校类型这些人口统计学变量作为第一层预测变量，教师课堂情绪体验作为第二层预测变量，教师情绪调节方式作为第三层预测变量。回归分析结果如表 4 所示。

由表 4 可以得出，在第一层回归分析中，人口学变量共解释实习教师效能感变异量的 5%，只有是否担任班主任这一维度对实习教师效能感有显著的影响（$\beta =-0.16$, $p<0.01$）；在第二层回归分析中，加入实习教师课堂情绪体验这一变量之后，课堂情绪体验共解释实习教师效能感变异量的 13%，是否担任班主任这一变量对实习教师效能感的影响作用不再显著，积极情绪对实习教师效能感产生正向影响，

表 4　以教师效能感为因变量的层级回归分析结果

变量	第一层（β）	第二层（β）	第三层（β）
性别	0.01	0.01	0.00
年龄	−0.13	−0.11	−0.10
学科	−0.02	−0.05	0.00
班主任	−0.16**	0.09	−0.09
学校	−0.05	0.04	−0.04
积极情绪		0.18**	0.09
消极情绪		−0.29**	−0.18**
认知重评			0.46**
表达抑制			0.00
ΔR^2	0.05	0.13	0.19
ΔF	2.00	15.92**	29.85**

消极情绪对实习教师效能感产生负向影响 (β =0.18, $p<0.01$；β =−0.29, $p<0.01$)；在第三层回归分析中，加入教师情绪调节方式变量之后，情绪调节方式共解释实习教师效能感变异量的19%，积极情绪对实习教师效能感的影响作用不再显著，消极情绪体验对实习教师效能感依然有显著性的影响 (β =−0.18, $p<0.01$)，认知重评对实习教师效能感产生积极影响，表达抑制的影响作用不显著 (β =0.46, $p<0.01$；β =0.00, $p>0.05$)。

3.4 情绪调节方式在课堂情绪与效能感之间的中介作用检验

由于在加入认知重评与表达抑制以后，积极情绪与消极情绪对教师效能感的影响作用减弱，因此，情绪调节方式可能在教师课堂情绪体验与教师效能感的关系中起中介作用。

本研究采用温忠麟等人提出的中介作用检验程序，采用回归分析的方法来验证情绪调节方式的中介作用（见表5）。

由表5可以得出，当未加入中介变量时，积极的课堂情绪与消极的课堂情绪对于实习教师效能感有显著的影响 (β =0.18, $p<0.01$; β =−0.29, $p<0.01$)。当加入认知重评变量后，积极情绪对于教师效能感的影响作用不再显著 (β =0.09，$p>0.05$)，消极情绪对于教师效能感仍然有显著的影响 (β =−0.18, $p<0.01$)，因此，认知重评在课堂积极情绪与教师效能感的关系中起完全中介作用，在课堂消极情绪与教师效能感的关系中起部分中介作用；当加入表达抑制变量后，由

表5　情绪调节方式在课堂情绪体验与教师效能感之间的中介

路径	B（SE）	β
认知重评 a		
C	0.35(0.13)/−0.46(0.10)	0.18**/−0.29**b
A	0.33(0.12)/−0.34(0.09)	0.19**/−0.25**
B	0.53(0.08)	0.46**
C'	0.17(0.12)/−0.28(0.09)	0.09/−0.18**
表达抑制 c		
C	0.35(0.13)/−0.46(0.10)	0.18**/−0.29**
A	0.45(0.12)/0.06(0.10)	0.25**/0.04
B	0.00(0.07)	0.00
C'	0.17(0.12)/−0.28(0.09)	0.09/−0.18**

于路径系数a与b均不显著,因此需要进行Sobel检验。Sobel 检验结果表明，表达抑制在课堂情绪与教师效能感之间中介作用不显著 ($z=0$, $p>0.05$; $z=0$, $p>0.05$)。

4　讨论

4.1　实习教师课堂情绪体验特点

在本研究中，实习教师在课堂教学中体验到更高频率的积极情绪，这主要是由于对于大多数实习教师而言，这是他们第一次将所学的理论知识用于教学实践活动，第一次与学生接触，这种新鲜感与热情可能会使他们更经常地体验到积极情绪。

4.2　实习教师效能感特点

在本研究中，实习教师效能感水平比较高，这与前人的研究是一致的。这主要是由于在实习过程中，实习教师能够随时获得指导老师的帮助和引导，这可能使他们更容易地应对课堂教学中的一些困难和挑战，这些成功经验的获得促进了他们的教师效能感。

4.3　实习教师情绪调节的特点

在本研究中，实习男教师更经常地使用表达抑制策略来调节他们的情绪，这与前人的研究结果是一致的。这主要是由于社会规范以及社会行为准则的影响作用。在社会化的过程中，男性和女性逐渐获得自己的性别角色以及相应的行为准则。在大多数文化中，男性一般被认为是理性化的，而女性被认为是情绪化的，因此男性不能像女性那样随意表露自己的情绪，因而更多的采用抑制的方式来掩饰他们的情绪。

在本研究中，与表达抑制策略相比，实习教师更习惯于使用认知重评策略。这可能是由于与表达抑制策略相比，认知重评是一种更加有效的策略。Gross 和 John 的研究表明，与习惯采用认知重评策略的个体相比，经常采用表达抑制策略调节情绪的个体在人际交往中更少与他人分享他们的情绪体验，并且不愿意与别人建立亲密关系，体验到较高水平的抑郁、较低的生活满意度、自尊及社会支持，尤

其是情绪上的支持。因此，随着生活经验的积累个体会更多地使用认知重评策略来调节他们的情绪。

4.4 班主任经验对实习教师课堂情绪与教师效能感的影响

在实习过程中，担任过班主任的教师比没有担任过班主任的教师在课堂教学中更经常地体验到积极情绪以及报告出更高水平的效能感。这可以从两个方面进行解释：第一，如果实习教师能够担任班主任这一角色，他们就会有更多的机会参与班级管理，与指导教师有更多的接触，从而获得更多的班级管理经验以及教学经验，这就使他们能更好地应对课堂教学过程中遇到的问题，进而更经常地体验到积极情绪以及较高水平的效能感；第二，如果实习教师担任班主任这一角色，他们就会有更多的机会和学生接触，从而建立起比较亲密的师生关系，促进学生良好的课堂行为，这也会对其效能感产生促进作用。

4.5 课堂情绪、情绪调节方式与教师效能感之间的关系

4.5.1 课堂情绪与情绪调节方式之间的关系

认知重评策略的使用与积极情绪呈正相关、与消极情绪呈负相关，这与Gross等人的研究结果是一致的。这是由于认知重评发生在情绪反应之前，能够影响情绪产生过程，即该策略不仅能够影响情绪的外部表达，而且还影响情绪的内部体验。对于表达抑制而言，使用该策略与积极情绪呈正相关、与消极情绪相关不显著，这与Gross等人的研究结果是不一致的。在Gross等人的研究中，习惯于使用表达抑制策略的被试体验到更少的积极情绪以及更多的消极情绪。然而这可能是由于东、西方的文化差异造成的。已有研究表明，在集体主义文化背景下，人们更多使用表达抑制的策略，相对于西方个体主义文化背景下的个体，该策略频繁使用并不会引起个体消极情绪体验的增强。虽然对情绪进行抑制需要消耗一定的认知资源，并且产生情绪失调，但是这有助于个体实现目的（比如，使课堂教学顺利进行、达成教学目标），这可能反而促进个体产生积极的情绪体验。

4.5.2 课堂情绪与教师效能感的关系

本研究发现，实习教师在课堂教学中的积极情绪体验对其效能感具有促进作用，而消极情绪体验对其效能感具有阻碍作用。这可以从以下两个方面来理解。第一，在积极情绪状态下，教师的思维以及行为会更加灵活，进而能够更有效地应对在课堂教学过程中的一些困难和挑战，这些成功经验的获得有助于提高实习教师效能感。

第二，课堂教学对于实习教师来说是比较有挑战性的任务。在这种情况下，个体不仅依赖其认知，而且还依赖情绪提供一些信息，来对个体的表现进行判断。如果实习教师更经常地体验到积极情绪，那么这就意味着个体能成功应对这一情境，这些直接经验的获得会增强其效能感；相反，如果实习教师更经常地体验到消极情绪，这就表明个体无力应对课堂教学情境，可能会降低其教师效能感。

4.5.3 认知重评在课堂情绪与教师效能感之间的中介作用

研究发现，认知重评在课堂情绪体验与教师效能感之间起中介作用，而表达抑制的中介作用不显著。这主要是由于经常使用认知重评策略能够使个体体验到更多的积极情绪以及更少的消极情绪，而情绪体验能够对个体的自我效能感产生影响；对于表达抑制而言，由于该策略主要是调节个体的情绪表达，对于个体的情绪体验没有显著影响，它对教师效能感影响不显著。

4.6 建议和启示

本研究对师范生教育和教师培训有重要启示。首先，在师范生教育以及教师培训的过程中，有必要将有关情绪和情绪调节的知识作为教师培训的一项重要内容，帮助教师认知自己的情绪，了解情绪调节的各种方法，进而促进教师的效能感；其次，要促进师范生和在职教师尤其是新手教师情绪调节技能的提高，鼓励他们更多地使用认知重评这种更具建设性的策略，更少地使用表达抑制这种可能对身心健康产生不利影响的策略，从而提升教师的情绪幸福感；最后，学校领导以及教育政策制定者要为教

师尤其是新手教师提供适当支持，以促进他们效能
感的提高或者使他们的效能感保持在较高的水平上，
从而促进学生学习。

参考文献

程立、袁加锦、何媛媛、李红 . (2009). 情绪调节策略：
认知重评优于表达意抑制 . *心理科学进展，17*(4).

龙君伟、 曹科岩 . (2006). 教师组织公民行为与教学
效能感关系的研究 . *心理科学，29*(4).

王力、 陆一萍、 李中权 . (2007). 情绪调节量表在
青少年人群中的适用 . *Chinese Journal of Clinical
Psychology, 15*.

温忠麟、 张雷、 侯杰泰、 刘红云 . (2004). 中介效应
检验程序及其应用 . *心理学报，36*(5).

Bandura, A. (1977). Self-efficacy: Toward a Unifying
Theory of Behavioral Change. *Psychological Review ,
84*.

Butler, E. A., Lee, T. L., & Gross, J. J. (2007). Emotion
Regulation and Culture: Are the Social Consequences
of Emotion Suppression Culture-Specific. *Emotion,* 7.

Fredrickson, B. L. (2001). The Role of Positive Emotions
in Positive Psychology. *American Psychologist, 56.*

Giuliani, N. R., McRae, K., & Gross, J. J. (2008). The Up-
and Down-Regulation of Amusement: Experiential,
Behavioral and Autonomic Consequences. *Emotion,* 8.

Gross, J. J. (1998). Antecedent-and Response-Focused
Emotion Regulation: Divergent Consequences for
Experience, Expression, and Physiology. *Journal of
Personality and Social Psychology, 74.*

Gross, J. J., John, O. P. (2003). Individual Differences in
Two Emotion Regulation Processes: Implications for
Affect, Relationships, and Well-Being. *Journal of
Personality and Social Psychology, 85.*

Gross, J. J., Levenson, R. W. (1997). Hiding feelings:
The Acute Effects of Inhibiting Negative and Positive
Emotion. *Journal of Abnormal Psychology, 106.*

Hostani, R., Imai-Matsumura, K. (2011). Emotional
experience, expression and regulation of high-quality
Japanese elementary school teachers. *Teaching and
Teacher Education, 27.*

Hoy, A. W., Spero, R. B. (2005). Changes in teacher
efficacy during the early years of teaching: A
comparison of four measures. *Teaching and Teacher
Education, 21.*

Kavanagh, D. J., Bower, G.H. (1985). Mood and Self-
Efficacy: Impact of Joy and Sadness on Perceived
Capabilities. *Cognitive Therapy and Research, 9.*

Sutton, R. E. (2004). Emotional regulation goals and
strategies of teachers. *Social Psychology of Education,
7.*

Tschannen-Moran, M., Hoy, A. W. (2001).
Teacherefficacy: capturing an elusive construct.
Teaching and Teacher Education, 17.

The relationship between student teachers' classroom emotion and teacher efficacy
——the mediation effect of emotion regulation

Gong Shaoying[1,2,3], Zhang Kuoteng[2] , Zhang Panpan[2], Li Jiawei[2]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior(CCNU), Ministry of Education, Wuhan 430079, China)

([2] School of Psychology, Central China Normal University)

([3] Laboratory of Human Development and Mental Health of Hubei Province, Wuhan 430079, China)

Abstract This study explored the mediation effect of emotion regulation in the relationship between student teachers' classroom emotion and their teacher efficacy in two-hundred and fourteen student teachers. The results indicated that student teachers' positive emotion promoted their teacher efficacy, however, negative emotion weakened their teacher efficacy. Cognitive reappraisal mediated the relationship between emotion and teacher efficacy, and expression suppression didn't.

Keywords Positive emotion; negative emotion; cognitive reappraisal; expression suppression

心理健康双因素模型指标的再探讨及稳定性研究 *

海　曼 [1,2,3]，熊俊梅 [1,2,3,**]，龚少英 [1,2,3]，秦　屹 [4]，高苗苗 [1,2,3]

（[1] 青少年网络心理与行为教育部重点实验室，武汉，430079）

（[2] 湖北省人的发展与心理健康重点实验室，武汉，430079）

（[3] 华中师范大学心理学院，武汉，430079）

（[4] 中国地质大学外国语学院，武汉，430074）

摘　要　采用整群抽样以 1293 名中学生为被试，对其中 531 名进行追踪，探讨以生活满意度和抑郁构建的心理健康双因素模型 (DFM) 的可行性、稳定性及影响因素。结果：(1) 该 DFM 在中国中学生中适用；(2) 完全心理健康组的学业和社会功能最好，疾患组最差；(3) 完全心理健康组稳定性最高，有症状但满足组最低；(4) 管理消极情绪效能感和学业情绪显著预测心理健康。结论：简化指标 DFM 是双因素模型新的研究视角；管理消极情绪效能感和学业情绪是中学生心理健康的重要影响因素。

关键词　心理健康双因素模型；心理健康稳定性；情绪调节自我效能感；学业情绪

1　问题提出

在积极心理学的影响下，心理健康双因素模型 (Dual-Factor Model of Mental Health, DFM) 突破了以往以精神病理学指标作为心理健康的唯一衡量标准，融入了主观幸福感 (Greenspoon & Saklofske, 2001)。以主观幸福感为积极指标，精神病理学为消极指标，逐渐发展出四分说 (Suldo & Shaffer, 2008) 和六分说 (Keyes, 2002)。四分说的积极指标是主观幸福感，六分说的积极指标包含情感幸福感、心理幸福感和社会幸福感；四分说的消极指标包含内化、外化问题，六分说的消极指标是内化问题，这为 DFM 采用简化

指标提供了依据。四分说将人群分为：完全健康者，无心理疾病且主观幸福感高；易感者，无心理疾病但主观幸福感水平较低；有症状但满足者，虽有心理疾病但仍有高主观幸福感；疾患者，有心理疾病且主观幸福感低 (王鑫强，张大均，2011; Suldo & Shaffer, 2008)。Keyes 将心理健康分为六种类型：完全健康、部分心理健康 I、部分心理健康 II、完全病态、部分心理疾患 I 和部分心理疾患 II (Keyes, 2002)。因部分心理疾患 II 比例较少，故通常将其和部分心理疾患 I 合并为部分心理疾患 (Keyes, 2007)。部分心理健康 II 者与易感者心理健康状态类似，部分心理疾患者与有症状但满足者类似。DFM 的贡献之处在于筛查

出处于"风险边缘"的易感人群，并强调积极心理健康因素对人的保护作用。以四分说为基础研究青少年心理健康的效度较高 (董文婷 , 熊俊梅 , 王艳红 , 2014; Greenspoon & Saklofske, 2001; Suldo & Shaffer, 2008)，六分说常用于研究成人心理健康 (Keyes, 2002, 2007)。本研究的被试为中学生，故采用四分说的理论框架。

主观幸福感包括生活满意度、积极和消极情绪 (Diener, 2000)；精神病理学指标包括内化、外化问题 (Suldo & Shaffer, 2008)。两者在理论上均是人为合成的复合指标 (Greenspoon & Saklofske, 2001)。有研究者认为复合指标 DFM 可能存在以下问题 (王鑫强 , 张大均 , 2011; Doll, 2008)。第一，复合指标的评估敏感性不足。第二，生活满意度与内化问题关系紧密，而与外化问题的关系并不明确 (McKnight, Huebner, & Suldo, 2002)。第三，复合指标在心理干预中的可操作性不强。此外，复合指标 DFM 的各个子指标存在重复之处，如主观幸福感中的消极情感与精神病理学指标中的内化问题存在重叠。国内外研究者指出可以从简化指标的角度考虑构建双因素模型 (王鑫强 , 张大均 , 2011; Doll, 2008)。鉴于此，本研究尝试采用简化指标构建适合于中国中学生的双因素模型。

我国青少年的抑郁率高达 25.5% 到 44% (冯正直 , 张大均 , 2005)。抑郁是男性和女性青少年共有的高发心理疾病 (Achenbach & Edelbrock, 1989)，通常损害其学业和社会功能，如导致学业成绩下降，人际交往障碍，甚至自杀等行为 (Balázs et al., 2013; Maurizi, Grogan-Kaylor, Granillo, & Delva, 2013)。 本研究选取抑郁作为构建 DFM 的消极指标。采用西方文化下开发出的主观幸福感测量工具通常得出东方人的幸福感比西方人低的结果 (Diener, Suh, Smith, & Shao, 1995)，其原因在于复合指标的测量方法忽视了东方人对低激发情感 (如"平静"而不是"兴奋") 的文化偏好及在情感词汇上的谦虚表达造成情感量表的区分度下降 (吴胜涛 , 王力 , 周明洁 , 王文忠 , 张建新 ,2009)。更能反映中国人主观幸福感真实情况的是生活满意度 (孙瑞琛 , 刘文婧 , 许燕 , 2010)。由

于生活满意度比积极和消极情绪更加稳定，生活满意度已成为衡量青少年心理健康积极维度的核心指标 (Suldo & Huebner, 2006)。已有研究采用生活满意度作为 DFM 的积极指标进行量化研究 (Eklund, Dowdy, Jones, & Furlong, 2011)。此外，由于抑郁已经包含了消极情绪，为不重复测量指标，本研究拟选取生活满意度作为心理健康积极指标构建双因素模型。研究问题一探讨拟构建的 DFM 的可行性。

实证研究发现四类心理健康人群的学业和社会功能存在差异：完全心理健康组的学业动机、对学校价值的评价和对学校的态度 (Suldo & Shaffer, 2008)，以及认知参与 (Antaramian, Huebner, Hills, & Valois, 2010) 高于其他三类人群；完全心理健康组和有症状但满足组的情感参与，以及希望情绪高于易感组和疾患组 (Antaramian et al., 2010)。董文婷等人 (2014) 发现完全心理健康者与其他三类人群相比，体验到更少的压力和无助学业情绪，更多的放松情绪。情绪调节效能感通过学业效能感影响学生的亲社会行为、违纪行为和抑郁水平 (Bandura, Caprara, Barbaranelli, Gerbino, & Pastorelli, 2003)。学业自我效能和学业情绪对学业表现和心理健康有预测作用 (董妍 , 俞国良 , 2010; 熊俊梅 , 2012)。研究问题二进一步考察中学生心理健康人群的分布和功能（情绪调节效能感、学业效能感、学业情绪）差异。

目前，基于复合指标 DFM 探讨心理健康人群稳定性的纵向研究较为缺乏。Kelly 等人 (2012) 对中学生进行为期一年的追踪，发现心理健康未发生变化者占总人数的 68.5%，其中，完全心理健康组最稳定，易感组最不稳定。处于中间状态的心理健康群组最易发生转化的结论需要更多实证研究的支持。在纵向研究设计基础上，研究问题三深入考察中学生心理健康状况的稳定性及影响因素。

2　研究方法

2.1　被试

横断面研究被试来自湖北、河南、安徽的中学

生 1293(15.52 岁 ±.75) 人。其中，男生 680 人，女生 613 人；初中 624 人，高中 669 人。对横断研究中湖北省的一所高中和一所初中的 531(15.45 岁 ±.68) 人进行追踪。其中，男生 298 人，女生 233 人；初中 255 人，高中 276 人。

2.2 研究工具

采用由 Pavot 和 Diener 于 1993 年编制，2002 年 Mantak Yuen 修订的《生活满意度问卷》(岳颂华，张卫，黄红清，李董平，2006)，共 5 题，7 点计分。本研究中的 α 系数前、后测为 .85 和 .87。

选取刘贤臣和杨杰 (1997) 修订的 Achenbach《青少年自评量表》中的抑郁维度测量青少年抑郁水平，共 16 题，3 点计分。本研究中的 α 系数前、后测为 .85 和 .89。

采用 Caprara 等人的《情绪调节自我效能感量表》的国内修订版 (文书锋，汤冬玲，俞国良，2009; Caprara et al., 2008)，共 12 题，5 点计分。量表包括表达积极情绪效能感和合并后的管理消极情绪 (包括沮丧 / 痛苦和生气 / 愤怒) 效能感 (李彩娜，党健宁，何姗姗，李红梅，2013)，其同质性和合成信度分别为 .67 和 .82，合成信度 95% 置信区间为 (.805，.835) (叶宝娟，温忠麟，2012)。本研究中表达积极情绪效能感、管理消极情绪效能感的 α 系数前测分别为 0.70、0.80，后测分别为 0.78、0.83。

采用 Pintrich 和 De Groot(1990) 的《学业自我效能感量表》的国内修订版 (梁宇颂，2000) 中的学业能力效能感维度，11 道题，5 点计分。本研究中的 α 系数前测为 0.86，后测为 0.90。

采用《青少年学业情绪问卷》(董妍，俞国良，2007) 中高兴和无助维度，各有 6 个和 4 个题项，5 点评分。本研究中高兴、无助的 α 系数前测为 0.70、0.81，后测为 0.72、0.84。

2.3 施测与数据处理

在征得校领导和被试的知情同意后，以班级为单位施测。主试由心理学专业研究生担任。采取自愿参加原则。第一次施测于 2013 年 9 月，发放 1400 份问卷，回收有效问卷 1293 份。第二次施测于 2014 年

1 月，发放 600 份问卷，回收有效问卷 531 份。采用 SPSS17.0 和 AMOS17.0 进行数据处理。采用 Harman 单因素方法进行共同方法偏差检验，发现横断研究和纵向研究中，特征根大于 1 的因子各有 12 个，第一个因子的解释率分别为 17.90% 和 14.77%，远低于临界标准 40%，说明不存在严重的共同方法偏差。

3 结果

3.1 验证以生活满意度和抑郁为指标的 DFM

为控制由潜变量多个项目造成的膨胀测量，采用平衡取向中的主成分分析法对抑郁项目打包(卞冉，

表 1 拟构建 DFM 的单因素模型和双因素模型的结构方程模型指标

	x^2/df	GFI	CFI	TLI	AGFI	RMSEA
模型 A	92.38	.64	.57	.42	.41	.27
模型 B	4.12	.98	.99	.98	.97	.05

车宏生，阳辉，2007)。将负荷最高的几个项目作为每个小组的锚定项目，反方向加入次高项目组，用每个包裹内的平均分作为打包后的观测变量。生活满意度仅 5 题，存在上述问题的可能性较小，故不打包。

以 DFM 为理论基础 (Suldo & Shaffer, 2008)，参考熊俊梅 (2012) 的模型建构思路，设计出两个模型。模型 A 为单因素模型：潜变量是具有双极性的心理健康单维指标，将生活满意度和抑郁题项负荷其上；模型 B 是以生活满意度和抑郁为指标的 DFM：两个潜变量是积极、消极心理健康维度，其中生活满意度题项负荷在积极维度上，抑郁题项负荷在消极维度上。模型 B 各项指标拟合良好，优于模型 A(见表 1)。

3.2 四种心理健康类型青少年的学业和社会功能的特点

根据生活满意度的评分标准，4 分及以上表明生活满意度较高，4 分以下表示较低 (Eklund, et al., 2011)。根据《青少年自评量表》划分内化问题临界点的标准，90 百分位及以上为高内化问题组，90 百分位以下为低内化问题组 (刘贤臣，杨杰，1997)。结合以上两个划分标准，完全心理健康者占总人数

表 2　四类心理健康人群情绪调节效能感、学业效能和学业情绪的特点

	完全心理健康者（N=631）		易感者 (N=513)		有症状但自我满足者 (N=47)		疾患者 (N=102)		F 值	η 2partial
	M	SD	M	SD	M	SD	M	SD		
表达积极情绪效能	4.11	.64	3.87	.68	4.02	.66	3.95	.65	12.66***	.03
管理消极情绪效能	3.54	.68	3.25	.67	3.20	.62	2.91	.72	34.64***	.08
学业效能	3.57	.60	3.34	.59	3.35	.54	3.09	.73	24.79***	.06
高兴	4.03	.58	3.84	.58	4.02	.57	3.81	.70	11.89***	.03
无助	2.35	.78	2.57	.83	3.25	.94	3.15	.96	41.92***	.09

注：*p<.05, **p<.01, ***p<.001；下同。

(N=1293) 的 48.8%，易感者占 39.7%，有症状但满足组占 3.6%，疾患组占 7.9%。

对四类人群的学业效能感、情绪调节效能感和学业情绪进行方差分析 (见表 2)。结果表明，四类人群在各变量上差异显著。LSD 事后检验显示，完全心理健康组各方面均优于其他三组 ($M_{difference}$=0.16~0.90，SE=0.07~0.01, p<0.05)，有症状但满足组的管理消极情绪效能感、学业效能感及高兴均显著高于疾患组 ($M_{difference}$=0.29, SE=0.12, p<0.05; $M_{difference}$=0.26, SE=0.11, p<0.05; $M_{difference}$=0.21, SE=0.10, p<0.05)。

3.3 以生活满意度和抑郁为指标的 DFM 的稳定性及影响因素

3.3.1 四类心理健康人群的稳定性及转化

T1 时间点上，完全心理健康者占总人数（N=531）的 48.8%，易感者占 39.8%，有症状但满足者占 2.4%，疾患者占 9.0%；T2 时间点上，完全心理健康者占

49.5%，易感者占 37.3%，有症状但满足者占 3.8%，疾患者占 9.4%。心理健康类型比例基本一致说明基于简化指标 DFM 分类的四类心理健康人群较为稳定。其中，完全心理健康组稳定性最高，有症状但满足组稳定性最低 (见表 3)。

将心理健康的转化分为不变，变好和变差三种趋势 (Kelly, et al., 2012)。心理健康不变的人数占总样本的 62%。在心理健康发生变化的人群中，心理健康状况变好 (包括：2、3、4、变 1；4 变 2、3；3 变 2) 和变差的人数 (包括：1 变 2、3、4；2、3 变 4；2 变 3) 分别占总样本的 19.4%、18.6%。

3.3.2 心理健康稳定性及影响因素

从 T1 时间自变量对 T2 时间生活满意度的回归分析发现 (见表 4)，无变化组管理消极情绪效能感和高兴显著预测生活满意度，变好组仅无助显著预测生活满意度，变差组仅高兴显著预测生活满意度。从

表 3　基于生活满意度和抑郁指标的四类心理健康人群的转化情况

抑郁	生活满意度				
	高			低	
低	1 完全心理健康者	2 易感者			
		①	本研究	①	本研究
	无变化	85%	70.3%	2 变 1	46%	33%
	1 变 2	6%	23.2%	无变化	29%	58%
	1 变 3	9%	1.2%	2 变 3	14%	3.8%
	1 变 4	1%	5.4%	2 变 4	12%	5.2%
高	3 有症状但满足者	4 疾患者			
		①	本研究	①	本研究
	3 变 1	43%	38.5%	4 变 1	18%	14.9%
	3 变 2	7%	15.4%	4 变 2	23%	27.7%
	无变化	42%	23.1%	4 变 3	12%	12.8%
	3 变 4	7%	23.1%	无变化	47%	44.7%

注：数字 1 表示完全心理健康者，2 表示易感者，3 表示有症状但满足者，4 表示疾患者。①表示 Kelly（2012）研究中各类型人群的转化比例，其划分标准是主观幸福感和精神病理学指标。本研究中划分标准是生活满意度和抑郁。

表 4　T1 时间情绪调节效能感、学业效能、学业情绪分别对 T2 时间生活满意度、抑郁的回归

因变量	生活满意度						抑郁					
	无变化组		变好组		变差组		无变化组		变好组		变差组	
自变量	β	t	β	t	β	t	β	t	β	t	β	t
表达积极情绪效能	.09	1.70	−.02	−.18	−.19	−1.81	−.01	−.26	−.04	−.45	−.05	−.54
管理消极情绪效能	.30	5.41***	.18	1.74	−.05	−.41	−.17	−3.06**	−.35	−3.47***	−.26	−2.36*
学业效能	.03	.46	.16	1.22	−.03	−.24	.02	.27	−.19	−1.53	.05	.35
高兴	.18	2.90**	−.15	−1.32	.24	2.10*	−.08	−1.32	.10	.93	.19	1.72
无助	−.07	−1.08	−.23	−2.15*	−.03	−.22	.30	4.72***	.10	.93	.29	2.36*
R^2	.22		.18		.06		.17		.21		.11	
F	18.25***		4.21**		1.25		14.54***		6.46***		3.46**	

T1 时间自变量对 T2 时间抑郁的回归分析发现，无变化和变差组的管理消极情绪效能感和无助显著预测抑郁，在变好组中，只有管理消极情绪效能感显著预测抑郁。

4　讨论

以生活满意度和抑郁作为两维指标的验证性因素分析说明简化指标 DFM 适用于中国中学生。该研究结果扩大了 DFM 的外部效度，为心理健康干预工作提供了新的视角，并为中国人的生活满意度可以较好反映其主观幸福感的结论提供了间接支持。生活满意度和抑郁是心理健康两个独立但相关的指标。无抑郁症状并不意味着高生活满意度，低生活满意度并不意味着有抑郁症状。

国外研究中完全心理健康者居多，而其他三类人群的分布尚未有一致结论 (Eklund, et al., 2011; Suldo & Shaffer, 2008)。国内高中生中完全心理健康者比例也最高。与复合指标 DFM 相比本研究筛查出的易感者比例增幅较大。筛查出更高比例的易感人群为简化指标 DFM 敏感性更高的结论提供了部分证据，该结论是否成立有待进一步研究。以抑郁作为 DFM 的消极指标同样可以较好区分不同心理健康类型的学生，说明用抑郁作为构建 DFM 消极指标切实可行。心理健康积极指标从主观幸福感精简为生活满意度。主观幸福感中的积极和消极情绪的干预难度较大，而对认知层面生活满意度的干预更具有操作性 (Suldo & Huebner, 2006)。与复合指标 DFM 相比简化指标 DFM

的潜在优势在于：其一，可提高鉴别不同类型心理健康人群的敏感性；其二，该模型建构时选取了中学生常见的抑郁问题，对中学生心理健康干预更有针对性；其三，心理健康指标构成更简洁，使得心理筛查和干预的可操作性更强。此外，简化指标在构建上解决了复合模型中指标重复测量的问题。

在四类心理健康人群的功能差异方面，完全心理健康组最好，疾患组最差，另外两类人群居中，这与以往研究的结果一致 (董文婷等 , 2014; Antaramian et al., 2010; Suldo & Shaffer, 2008)。四类心理健康人群的学业和社会功能差异进一步证明了简化指标 DFM 的可行性。值得关注的是，有症状但满足者的表达积极效能感和高兴体验与完全健康者无异；易感组的抑郁虽不明显，但他们低的生活满意度使其积极情绪效能感和高兴体验与疾患组无异。如此看来，生活满意度可能是学业和社会功能的保护性因素。

简化指标 DFM 的前后测结果显示，四类人群的比例分布基本保持稳定，说明模型信度良好。四类心理健康人群中 62% 的青少年心理健康未发生质的变化，其中完全心理健康组稳定性最高，有症状但满足组最低，且有症状但满足组向完全心理健康组和疾患组转化的比例均居于首位。有症状但满足组心理健康变化呈现两种截然相反的发展趋势，原因可能是其积极和消极心理健康指标"双高"，表现为同时具有高生活满意度和高抑郁水平。这提醒心理健康工作者要关注心理健康处于中间状态的人群并及时进行干预。未来研究需要对有症状但满足者的心理健康的影响因素及变化规律做更深入的量化研究。

本研究发现，在情绪调节效能感中，管理消极情绪效能感对生活满意度、抑郁的影响尤为明显。以往研究发现表达积极情绪效能感高的个体的亲社会行为多，管理消极情绪效能感则对抑郁、违纪等行为具有重要影响 (Bandura, et al., 2003; Caprara, et al., 2008)，说明两种情绪调节效能感对个体发展的影响不同。在中国文化中分享积极情绪比分享消极情绪更被认可，表达消极情绪受到更多的限制 (周婷，王登峰，2012)。是否能调节好消极情绪对个体的心理健康状况尤为关键，这可能是管理消极情绪效能感对中国中学生心理健康状况影响较大的原因之一。今后可展开情绪调节效能感与心理健康之间的机制研究。此外，无变化组和变差组高兴越多，则生活满意度越高；无助越多，则抑郁水平越高。这说明学业情绪对心理健康具有良好的风向标作用。由此看来，管理消极情绪效能感和学业情绪是心理健康的重要影响因素。本研究涉及的学业情绪种类有限，其它种类的学业情绪对中学生心理健康的影响还需要进一步展开量化研究。此外，由于本研究仅选取了中国部分省份，未来的研究可扩大样本抽样范围，以增加模型的外部效度。

5 结论

（1）以生活满意度和抑郁为指标的 DFM 在中国中学生中适用，为复合指标 DFM 提供了新的研究视角。

（2）完全心理健康者的学业和社会功能表现最优，疾患组最差，其余两类人群居中。

（3）中学生心理健康状况较稳定；完全心理健康组稳定性最高，有症状但满足组稳定性最低。

（4）管理消极情绪效能感和学业情绪是中学生心理健康的重要影响因素。

参考文献

卞冉，车宏生，阳辉 . (2007). 项目组合在结构方程模型中的应用 . 心理科学进展, 15(3), 567–576.

董文婷，熊俊梅，王艳红 . (2014). 心理健康双因素模型的中国高中生实证调查 . 中国临床心理学杂志, 22(1), 88–91.

董妍，俞国良 . (2007). 青少年学业情绪问卷的编制及应用 . 心理学报, 39(5), 852–860.

冯正直，张大均 .(2005). 中学生抑郁症状的流行病学特征研究 . 中国行为医学科学, 14(2), 11–13.

李彩娜，党健宁，何姗姗，李红梅 . (2013). 羞怯与孤独感——自我效能的多重中介效应 . 心理学报, 45(11), 1251–1260.

梁宇颂 . (2000). 大学生成就目标, 归因方式与学业自我效能感的研究 . 武汉：华中师范大学 .

刘贤臣，杨杰 . (1997). Achenbach 青少年行为自评量表的信度和效度研究 . 中国心理卫生杂志, 11(4), 200–203.

孙瑞琛，刘文婧，许燕 . (2010). 不同出生年代的中国人生活满意度的变化 . 心理科学进展, 18(7), 1147–1154.

文书锋，汤冬玲，俞国良 . (2009). 情绪调节自我效能感的应用研究 . 心理科学, 32(3), 666–668.

王鑫强，张大均 . (2011). 心理健康双因素模型述评及其研究展望 . 中国特殊教育, 136(10), 68–73.

吴胜涛，王力，周明洁，王文忠，张建新 .(2009). 灾区民众的公正观与幸福感及其与非灾区的比较 . 心理科学进展, 17(3), 579－587.

熊俊梅 . (2012). 中学生的学业情绪与心理健康 . 广州：世界图书出版广东有限公司 .

叶宝娟，温忠麟 (2012). 用 Delta 法估计多维测验合成信度的置信区间 . 心理科学, 35(5), 1213–1217.

岳颂华，张卫，黄红清，李董平 . (2006). 青少年主观幸福感、心理健康及其与应对方式的关系 . 心理发展与教育, 22(3), 93–98.

周婷，王登峰 .(2012). 情绪表达抑制与心理健康的关系 . 中国临床心理学杂志 . 20(1), 65–68+64

Achenbach, T. M., & Edelbrock, C. (1989). *Manual for the Youth Self-Report and Profile.* University of Vermont

Department of Psychiatry.

Antaramian, S. P., Huebner, E. S., Hills, K. J., & Valois, R. F. (2010). A dual - factor model of mental health: Toward a more comprehensive understanding of youth functioning. *American Journal of Orthopsychiatry, 80*(4), 462–472.

Balázs, J., Miklósi, M., Keresztény, ?., Hoven, C. W., Carli, V., Wasserman, C., . . . Wasserman, D. (2013). Adolescent subthreshold–depression and anxiety: Psychopathology, functional impairment and increased suicide risk. *Journal of Child Psychology and Psychiatry, 54*(6), 670–677.

Bandura, A., Caprara, G. V., Barbaranelli, C., Gerbino, M., & Pastorelli, C. (2003). Role of affective self - regulatory efficacy in diverse spheres of psychosocial functioning. *Child Development, 74*(3), 769–782.

Caprara, G. V., Di Giunta, L., Eisenberg, N., Gerbino, M., Pastorelli, C., & Tramontano, C. (2008). Assessing regulatory emotional self–efficacy in three countries. *Psychological Assessment, 20*(3), 227–237.

Diener, E. (2000). Subjective well–being: The science of happiness and a proposal for a national index. *American Psychologist, 55*(1), 34–43.

Doll, B. (2008). The dual–factor model of mental health in youth. *School Psychology Review, 37*(1), 52–68.

Diener,E.,Suh,E.M.,Smith,H.,&Shao,L.(1995).National differences in reported subjectiv ewell–being:?Why do they occur *Social Indicators Research,34*,7–32.

Eklund, K., Dowdy, E., Jones, C., & Furlong, M. (2011). Applicability of the dual–factor model of mental health for college students. *Journal of College Student Psychotherapy, 25*(1), 79–92.

Greenspoon, P. J., & Saklofske, D. H. (2001). Toward an integration of subjective well–being and psychopathology. *Social Indicators Research, 54*(1), 81–108.

Kelly, R. M., Hills, K. J., Huebner, E. S., & McQuillin, S.

D. (2012). The longitudinal stability and dynamics of group membership in the dual–factor model of mental health: Psychosocial predictors of mental health. *Canadian Journal of School Psychology, 27*(4), 337–355.

Keyes, C. L. M. (2002). The mental health continuum: From languishing to flourishing in life. *Journal of Health and Behavior Research, 43*(2), 207–222.

Keyes, C. L. M. (2007). Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health. *American Psychologist, 62*(2), 95–108

Maurizi, L. K., Grogan–Kaylor, A., Granillo, M. T., & Delva, J. (2013). The role of social relationships in the association between adolescents' depressive symptoms and academic achievement. *Children and Youth Services Review, 35*(4), 618–625.

McKnight, C. G., Huebner, E. S., & Suldo, S. (2002). Relationships among stressful life events, temperament, problem behavior, and global life satisfaction in adolescents. *Psychology in the Schools, 39*(6), 677–687.

Pekrun, R. (2006). The control–value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341.

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and self–regulated learning components of classroom academic performance. *Journal of Educational Psychology, 82*(1), 33–40.

Suldo, S. M., & Huebner, E. S. (2006). Is extremely high life satisfaction during adolescence advantageous？ *Social Indicators Research, 78* (2), 179 — 203

Suldo, S. M., & Shaffer, E. J. (2008). Looking beyond psychopathology: The dual–factor model of mental health in youth. *School Psychology Review, 37*(1), 52–68.

A reexamination of the indicators for the dual-factor model of mental health: Stability and predictors of mental health

HAI Man[1,2,3]; XIONG Junmei[1,2,3]; GONG Shaoying[1,2,3]; QIN Yi[4]; GAO Miaomiao[1,2,3]

([1]Key Laboratory of Adolescent Cyberpsychology and Behavior, Ministry of Education, Wuhan, 430079)

([2]Key Laboratory of Human Development and Mental Health of Hubei Province, Wuhan, 430079)

([3]School of Psychology, Central China Normal University, Wuhan, 430079)

([4]Foreign Languages School, China University of Geosciences, Wuhan, 430074)

Abstract　Adual-factor model of mental health (DFM) examines the presence of wellness (i.e., subjective well-being; SWB) and psychopathology (i.e., internalizing and externalizing behavior problems) in explaining youth mental health functioning. Using a dual-factor model, previous research has yielded four groups: the complete mental health, the vulnerable, the symptomatic but content, and the troubled. Some researchers believe that DFM with integrated indicators may have some potential problems (Doll, 2008; Wang & Zhang, 2011). Affective self-regulatory efficacy is related to students' depression and anxiety. Students' academic self-efficacy and academic emotions contribute to their learning outcome and mental health. This study aimed at testing the applicability of using life satisfaction on the wellness dimension and depression on the illness dimension of DFM, comparing academic and social functions of the four mental health groups, and further investigating the stability and changes of mental health group membership.Method: Cross-sectional design and longitudinal design were applied in this study. This research took cluster sampling method to collect data from 1293 middle school students (Mage = 15.52, SD = ± .75) and among them 531 were followed as the subjects of a longitudinal study. The self-report questionnaires used in the study included affective self-regulatory efficacy scale, academic self-efficacy scale, academic emotions scale, life satisfaction scale, and depression subscale of internalizing problems scale (Youth Self-Report). The survey was conducted at two time points: September 2013 and January 2014. Structural equation modeling was applied to test the applicability of DFM model. MANOVA and Post Hoc tests (LSD) were applied to investigate the four groups' academic and social functions. Multiple regressions were applied to test the contribution of affective self-regulatory efficacy, academic self-efficacy, and academic emotions to mental health. Results: (1) DFM with life satisfaction loaded on the wellness dimension and depression loaded on the illness dimension was applicable to Chinese middle school students. (2) The complete mental health group demonstrated the best academic and social functions, and the troubled demonstrated the worst functions. The symptomatic but content demonstrated higher self-efficacy to regulate negative affect, academic self-efficacy and enjoyment than the troubled. The vulnerable did not differ from the troubled in self-efficacy to regulate positive affect or enjoyment. (3) From T1 to T2, middle school students' mental health was relatively stable. The majority of the longitudinal sample did not change in mental health status. The proportions of positive change and negative change were

close to each other. Among the change groups, the complete mental health had the highest stability, while the symptomatic but content had the lowest stability. (4) Self–efficacy for negative affect regulation, enjoyment and hopelessness predicted students' life satisfaction and depression significantly. The above results suggested that life satisfaction and depression are two independent and correlated constructs of DFM, providing an alternative research perspective to DFM with integrated indicators.

Keywords wordsdual–factor model of mental health; stability of mental health; affective self–regulatory efficacy; academic emotions

高中生数学学业情绪、学习策略与数学成绩的关系

熊俊梅，龚少英，Anne C. Frenzel

摘　要　本研究通过对 464 名高二学生问卷调查，考察了高中生学业情绪、学习策略与数学成绩的关系。结果显示: 数学成绩与数学自我概念、积极情绪、元认知学习策略成显著正相关,与消极情绪成显著负相关; 高兴、自豪与认知、元认知策略成显著正相关, 生气、焦虑、无助、厌倦与认知、元认知策略成显著负相关; 情绪、学习策略、自我概念和环境因素可以显著地预测数学成绩; 男生比女生运用更多的元认知策略,有更高的数学自我概念,更多的积极情绪, 和更少的消极情绪。影响数学成绩的因素包括认知和非认知因素; 女生在数学学习中的消极情绪可能与低数学自我概念有关。

关键词　学业情绪；学习策略；高中生；数学学习

　　了解学生认知、情绪和动机过程的内在联系是教育心理学日渐形成的焦点。情绪和动机不仅能调节注意力和认知,有时还能转化认知。个体的自我认知、成就目标、认知能力等个体因素以及班级、家庭、课堂等环境因素都会影响学业情绪,且自我认知是环境因素影响其学业情绪的一个重要中介变量。了解青少年在数学学习过程中的情绪和认知有助于教育工作者从多维角度分析学生的学业表现并帮助学生提高学习效果。

　　上世纪 50 年代初心理学界开始了对考试焦虑的广泛研究。但对积极学业情绪的关注度不够。德国教育心理学家 Pekrun 提出了社会认知控制 – 价值学业情绪理论（social cognitive control–value theory of academic emotions）。学业情绪是与学习活动或结果直接联系在一起的情绪,包括积极和消极情绪,高唤醒和低唤醒情绪。低唤醒消极情绪如厌倦对进取心和学习效果有负面影响;高唤醒消极情绪如焦虑在减少内在学习兴趣的同时加强外在进取心。中学生和大学生常体会到的学业情绪有喜悦、希望、自豪、生气、焦虑、厌倦和羞愧。学业情绪概念的相似性在跨文化背景的实证研究中已得到证实。学业情绪与人格,进取心,学习策略,认知资源,自我调控和学习效果有关。

　　Pintrich 等人认为认知和元认知学习策略包括重复演练法、整合联系法、组织法、批判性思维以及元认知自我调控;元认知学习策略的使用是反映认知水平和学习能力的重要指标之一;不同的认知策略如重复演练法、整合联系法和组织法对加强积极认知卷入和提高学业成绩有促进作用。由此可以看出学习策略在学习过程中的重要性。

　　近 20 年来, 学业情绪影响认知加工、注意力、学习动机、学习策略和成绩的研究在国内外已经广泛展开。关于数学学业情绪的研究发现, 当控制学生已有的学习水平, 女生比男生有更多的消极情绪和更少的积极情绪; 女生的情绪模式与低能力信念有关。学习环境和情绪有显著相关关系。

在前人的理论和实证研究基础上，本研究试图深入了解高中生数学学习心理特点，进一步明确学业情绪、学习策略和数学成绩之间的关系，并检验它们在跨文化背景中关系的一致性。

1 研究方法

1.1 对象

随机选取石家庄某重点中学和武汉市两所普通中学高二年级 8 个班学生为被试。发放问卷 490 份，回收 480 份，有效问卷 464 份，有效率为 94.7%。男生 237 人，女生 227 人。整个样本平均年龄 17.16 岁（$SD=0.56$）。

1.2 研究工具

1.2.1 数学学业情绪量表

该量表修订自 Pekrun 等编制的数学学业情绪量表（Academic Emotions Questionnaire, 简称 AEQ-M），1–5 级记分，共 60 个条目，分为高兴、自豪、生气、焦虑、羞愧、无助和厌倦 7 个维度。各维度内部一致性信度系数在 0.84–0.92 之间。积极学业情绪是高兴和自豪两个维度的平均分，消极学业情绪是生气、焦虑、羞愧、无助和厌倦五个维度的平均分。

1.2.2 数学学习策略量表

该量表修订自 Pintrich 等编制的学习策略问卷（Motivated Strategies for Learning Questionnaire, 简 称 MSLQ），1–7 级记分，共 31 个条目，分别测量重复演练法、整合联系法、组织法、批判性思维法和元认知学习策略。各维度内部一致性信度系数在 0.64–0.80 之间。

1.2.3 数学自我概念量表

自编量表，5 点记分，共 4 个条目。该量表信度良好，前期研究中该量表的内部一致性信度系数为 0.81。

1.2.4 学习环境问卷

该问卷测量学生感知到的学习环境。自编量表，5 点记分，共 12 个条目，分为父母控制与惩罚、同辈态度倾向和教师热忱三个维度。前期研究中该量表

的三个维度内部一致性信度系数在 0.75 和 0.76 之间。

2 结果

2.1 不同性别高中生数学学习策略、情绪及自我概念的比较

以性别为自变量，已有数学成绩为协变量，数学学习策略、学业情绪和自我概念为因变量的多元协方差分析发现在数学学习中男生比女生有更多的高兴（$F(2, 355) = 57.09***$, $p< 0.001$ 下同）、自豪（$F(2, 355) = 28.29***$）和更少的生气（$F(2, 355) = 56.96***$）、焦虑（$F(2, 355) = 39.74***$）、羞愧（$F(2, 355) = 7.50***$）、无助（$F(2, 355) = 50.08***$）、厌倦（$F(2, 355) = 76.08***$），更高的数学自我概念（$F(2, 355) = 37.65***$），更频繁使用批判性思维（$F(2, 355) = 40.97***$）和元认知学习策略（$F(2, 355) = 55.60***$）。控制已有的数学成绩是必要的，因为已有的数学成绩有可能和性别一起预测学业情绪和策略等变量（结果见表 1）。

2.2 学业情绪、学习策略和学业成绩的关系

2.2.1 学业情绪、学习策略和学业成绩的相关

表 1 以性别为自变量，数学成绩为协变量，学业情绪、策略、自我概念为因变量的多元协方差分析

	男		女			
	M	SD	M	SD	F	偏 η^2
学业情绪						
高兴	3.30	0.70	3.11	0.68	57.09***	0.24
自豪	3.38	0.74	3.29	0.75	28.29***	0.14
生气	2.04	0.70	2.25	0.78	56.96***	0.24
焦虑	2.40	0.67	2.63	0.67	39.74***	0.18
羞愧	2.49	0.57	2.53	0.63	7.50***	0.04
无助	2.37	0.82	2.63	0.83	50.08***	0.22
无聊	2.16	0.79	2.35	0.92	76.08***	0.30
学习策略						
重复演练	3.19	1.05	3.43	1.12	3.59*	0.02
整合联系	3.95	1.12	3.81	1.20	18.46***	0.09
组织法	3.32	1.18	3.63	1.18	11.95***	0.06
批判思维	4.54	1.29	3.97	1.39	40.97***	0.19
元认知	4.57	0.95	4.24	0.99	55.60***	0.24
自我概念	3.01	0.72	2.75	0.62	37.65***	0.18

注：between group df=2 *P<0.05, ** P<0.01, *** P<0.001

表 2 数学学习策略、学业情绪、自我概念和数学成绩的平均数、标准差及相关值

	M(SD)	高兴	自豪	生气	焦虑	羞愧	无助	厌倦	自我	成绩
M(SD)		3.21 (.70)	3.34 (.76)	2.18 (.76)	2.54 (.68)	2.54 (.60)	2.53 (.85)	2.28 (.86)		
重复	3.35(1.14)	0.37	0.37	−0.21	−0.11	−0.002	−0.14	−0.29	0.16	0.07
整合	3.91(1.17)	0.56	0.48	−0.43	−0.35	−0.14	−0.36	−0.44	0.37	0.31
组织	3.52(1.20)	0.43	0.39	−0.31	−0.18	−0.03	−0.22	−0.35	0.24	0.19
批判	4.31(1.36)	0.50	0.44	−0.44	−0.41	−0.21	−0.38	−0.44	0.46	0.41
元认知	4.41(0,99)	0.64	0.51	−0.58	−0.49	−0.26	−0.50	−0.62	0.48	0.48
自我	2.91(0.70)	0.50	0.43	−0.40	−0.46	−0.31	−0.45	−0.37	–	–
成绩	108.52(26.26)	0.49	0.37	−0.49	−0.41	−0.20	−0.46	−0.55	0.39	–

注：|r| ≥ 0.10, $P<0.05$；|r| ≥ 0.12, $P<0.01$；|r| ≥ 0.15, $P<0 .001$。

学习策略与积极学业情绪成显著正相关关系，与消极学业情绪成显著负相关，其中重复演练法、组织法同羞愧感之间的相关系数不显著。数学自我概念与积极情绪（$0.43 \leq r \leq 0.50, p<0.001$）和学习策略（$0.16 \leq r \leq 0.48, P<0.001$）成正相关，与消极情绪成显著负相关（$−0.31 \leq r \leq −0.46, p<0.001$）。数学成绩与积极学业情绪成显著正相关（$0.37 \leq r \leq 0.49, p<0.001$），与消极情绪成显著负相关（$−0.20 \leq r \leq −0.55, p<0.001$）；成绩与除重复演练法的学习策略成正相关（$0.19 \leq r \leq 0.48$，当 $r \geq 0.12, p<0.01$ 当 $r \geq 0.15, P<0.001$）（结果见表 2）。

2.2.2 数学成绩的多元回归分析

数学成绩回归模型中的自变量包括学业情绪、策略、数学自我概念、学习环境和性别，因变量是数学成绩。回归的方差分析确认了回归模型的显著性 ($F_{(19, 338)} = 15.53, p< 0.001$)。五个显著预测指标对数学成绩方差解释水平调整后的多重相关平方值为 0.44。被列入回归方程模型的性别变量不是预测数学成绩的显著预测指标之一。父母控制、厌倦、无助感及重复演练法负向预测数学成绩；数学自我概念

表 3 数学成绩多元回归分析

	B	SEB	Beta	t 值
父母控制	−7.00	1.27	−0.26	−5.53***
厌倦	−8.44	2.59	−0.28	−3.26***
数学自我概念	4.95	2.04	0.13	2.43*
无助感	−5.82	2.50	−0.18	−2.33*
重复演练法	−3.28	1.33	−0.14	−2.46*

注：$F_{(19, 338)} = 15.53^{***}$，Adj. R2=0.44
*$P< 0.05$, ** $P<0.01$, *** $P<0.001$

正向预测数学成绩（结果见表 3）。

3 讨论

3.1 男女生数学学习心理差异带来的启示

研究结果显示，男生的数学学习成绩优于女生，男生比女生在数学学习中体验到更多的积极情绪和更少的消极情绪，更多采用批判性思维和元认知学习策略，并具有更积极的自我概念。Spelke 在关于数学性别差异一文中写到："大多数研究性别差异的专家发现，男性和女性有相等的认知能力，但具体构成因素有差异"。她建议我们超越认知能力的视角，从人的生物和社会视角去寻找男性数学学科优势的原因。Dai 认为资质具发展性，受遗传和环境因素影响，反映了先天和培育的复杂相互作用。因此，男、女生数学学习成绩的差异不能单纯认为是资质差异的结果，还需从环境因素中找原因。

和男生相比，女生自信心更低并且认为数学更难学。这一信念在较早阶段已经表现出来。这些性别差异性可能与父母看待子女学习数学的难易程度有关。学生家长和教师的态度一定程度上受数学学习性别刻板印象的影响。女性角色似乎成为了学习数学的天然障碍。数学学习性别刻板印象是解释数学成绩差异和情绪差异的可能原因。数学学习刻板印象可能是导致女生消极自我概念和低自我效能感的原因之一。因此，教育工作者和学生家长要关注积极学习环境与氛围的建设，帮助维持和加强学生的

自我概念，帮助学生成为自我调控的学习者。具体方法如：①给学生多一点学习的自主权，少一点控制；②少一点批评，多一点建设性意见；③合理评价学生的学业表现，不要过于严格和苛刻；④减少过度的重复性练习。

3.2 数学学业情绪、学习策略和数学成绩的关系

积极学业情绪、学习策略与数学成绩成显著正相关；消极学业情绪与数学成绩成显著负相关。学习策略与积极学业情绪成显著正相关关系，与消极学业情绪成显著负相关关系。数学自我概念与积极情绪和学习策略成正相关，与消极情绪成负相关。本研究的结果与前人的研究结果一致，这也证明了学业情绪、学习策略、自我概念和数学成绩间关系的跨文化一致性。

数学成绩的多元回归分析发现，数学成绩有多重预测指标，其方差可由认知和非认知等因素解释。厌倦感（$Beta=-0.28$，$t=-3.26$，$p<0.001$）、无助感（$Beta=-0.18$，$t=-2.33$，$p<0.05$）、学习环境中的父母控制与惩罚（$Beta=-0.26$，$t=-5.53$，$p<0.001$）和学习策略中的重复演练法（$Beta=-0.14$，$t=-2.46$，$p<0.05$）负向预测学习成绩；数学自我概念（$Beta=0.13$，$t=2.43$，$p<0.05$）正向预测学习成绩。因此，只从认知因素方面，如智力水平和学习策略的使用来解释数学成绩具有明显的片面性和不科学性。性别不是数学成绩的显著预测指标之一。因此，教师、学生家长和中学生要克服数学学习性别刻板印象。教师和学生家长要关注学生的情绪反应，给学生更多范围允许内的自主学习空间，适当减少成绩评估的严苛程度，指导学生使用正确的学习方法，保持和增进学生的自信心。学生要明确努力的作用，采用有效的学习策略学习数学。学生要积极地向自我调控的学习者转化。自我调控指树立现实的自我概念，控制和缓解消极的学业情绪，运用适当的学习策略和学会管理学习环境资源。

该研究是一个横断面研究，其横断面的性质决定了不能对认知与情绪变量的关系做因果判断。今后的相关研究可以采用现场实验和纵向研究的方法来解释学习心理机制中认知、情绪变量间的因果关系。该研究的其它局限性体现在采用自我报告法，女性被试有报告更多消极情绪体验的倾向。

参考文献

熊俊梅 . *学习的情绪因素：青少年数学学习自我概念、成就价值、情绪和成就动机* . 佩特朗恩出版社：法兰克福美因慈，2009.

徐先彩，龚少英 . (2009). 学业情绪及其影响因素 . *心理科学进展，17*(1).

俞国良，董妍 . (2007). 情绪对学习不良青少年选择性注意和持续性注意的影响，*心理学报，39*(4).

Eynde, P. O., & Turner, J. E. (2006). Focusing on the complexity of emotion issues in academic learning: dynamical component systems approach. *Educational Psychology Review, 18.*

Frenzel, A. C., Pekrun, R., & Goetz, T. (2007). Girls and mathematics A "hopeless" issue? A control value approach to gender differences in emotions towards mathematics. *European Journal of PsychologyofEducation, 22.*

Frenzel, A. C., Thrash, T. M., Pekrun, R., & Goetz, T. (2007). Achievement emotions in Germany and China: A cross- cultural validation of the Academic Emotions Questionnaire Mathematics. *Journal of Cross Cultural Psychology, 38,* 302– 309.

Pekrun, R., Goetz, T., & Frenzel, C. (2005). *Academic Emotions Questionnaire Mathematics (AEQ- M)*. User's manual. Department of Psychology, University of Munich.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. (2002). Academic emotions in students' self- regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37.*

Pekrun, R. (2006). The control- value theory of achievement emotions: Assumptions, corollaries, and

implications for educational research and practice. *Educational Psychology Review, 18.*

Pintrich, P. R., Smith, D. A., Garcia, T., & McKeachie, W. J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ),* Ann Arbor: UniversityofMichigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist, 60 .*

Zeidner, M. (1998). *Test anxiety: The state of the art,* Plenum, NewYork.

韦氏儿童智力量表第四版对学习障碍评估和干预的价值 *

臧 玲 [1,2]，张 微 [1,2]，刘华山 [1,2]

（[1] 青少年网络心理与行为教育部重点实验室）
（[2] 华中师范大学心理学院暨湖北省人的发展与心理健康重点实验室，武汉 430079）

摘 要 学习障碍儿童在基本的听、说、读、写、算等认知加工过程上发生困难，不同类型的学习障碍儿童在韦氏智力量表上的表现具有特异性，这是由于韦氏智力量表的内在结构与一些特定的认知加工过程有关。为加强韦氏智力量表的临床效应，第四版（WISC–IV）在结构、内容、功能等进一步优化调整，更加突出了量表在特殊学习者和情绪及行为障碍儿童上的临床辨别力不同类型的学习障碍儿童在韦氏智力量表第四版上的特定的分量表及特定的合成指数和认知加工指数上有特异性表现，这既可以成为临床辅助评估的出发点，同时也可以为不同类型的学习障碍儿童的个体化认知加工训练提供参考。

关键词 韦氏儿童智力量表第四版；学习障碍；评估；干预
分类号 G449

1 前言

一般而言，智力是指一般认知能力（g 因素），是感知、记忆、思维等各种认知能力的综合，其中以抽象思维能力为核心。Wechsler 则认为，智力是个体有目的地行动、理智地思考以及有效应对环境的综合能力，不仅如此，Wechsler 还注意到，智力不限于智力潜能，还包括一些非智力因素，这些非智力因素对智力的发展产生了极其重要的影响。

然而，早期的韦氏智力量表并没有以智力的内在维度结构为基础来设置分测验，而是借助了比奈智力测验的经典测验任务，以任务的呈现方式是言语还是非言语的，机械的将韦氏智力量表划分为言语量表和操作量表，合成的言语智商和操作智商并不等同于言语理解能力和空间知觉与推理能力，这给临床使用带来了困惑。1975 年 Kaufman 通过因素分析发现除了 WISC 修订本（WISC–R）中的标准三项分数以外，还存在第三因素，即抗干扰因素。它包括算术、背数和译码三个分测验，主要考察的是儿童的注意集中性、保持性和工作记忆能力。为强化第三因素的作用，1991 年修订的韦氏儿童智力量表第三版（WISC– Ⅲ）增加了符号搜索分测验，因素分析的结果却发现从注意力因素中分化出

工作记忆和数量能力这两个相互影响的因素。在此基础上，WISC-IV 进一步强化了这四个维度，调整为四指数结构，将抗干扰因素改为工作记忆指数（WMI），并与加工速度指数（PSI）一起成为新的分测验。

在中国，已经完成了韦氏儿童智力量表第二版和第四版的修订。本文将关注于新修订的韦氏儿童智力量表第四版（WISC-IV）在学习障碍的评估和干预上的价值。

2　WISC-IV 的新变化

WISC-IV 对量表的结构、内容做了大幅度的调整，更新了常模，增加了部分分测验，在评分标准上也做出了调整。使得测验的结果有助于特殊教育工作者做出更准确的解释和临床判断，如在分测验上，ADHD 儿童在译码和算术测验上的得分显著低于正常儿童，表现出了最大的效应量，而在鉴别阅读障碍上，效应量最大的几项分测验是：词汇理解，字母 – 数字排序、常识和算术。加工速度指数 (PSI) 在区分读写障碍和阅读障碍上的作用明显。

正如 Berninger 所说，WISC-IV 的技术特点是可用于学习障碍或其它儿童期异常的神经心理诊断，所以可以作为认知过程的监控者来确定认知优势和不足。

2.1　理论基础的充实和常模的更新

由于能力 – 成就差异和 RTI 方法在学习障碍的诊断中的不足，以 Cattell-Horn-Carroll（CHC）理论为基础的"Third Method"受到广泛关注。CHC 理论是以认知能力作为框架来确定认知加工和学业成就缺陷之间的相关，它的优势在于"使用不同的分测验综合考查和分析各项认知因子，之前没有哪种心理测量模型可以如此坚定的以数据为基础"。在吸收了 CHC 理论及其跨系列测评方法（Cross-Battry approach），使得 WISC-IV 及其各个分测验对测量目标的定义和解释变得明确而有体系，为增大测验结果的解释能力和对实践的指导价值奠定了

基础。

在常模建立上，为增强临床效用，WISC-IV 引入了特殊群体，包括天才儿童、中度精神障碍者、各种学习障碍儿童、ADHD 儿童、自闭症儿童、艾森伯格障碍儿童等 16 种特殊障碍儿童，为临床诊断提供了支持。并且，值得关注的是，在 WISC-IV 指导手册中，已经明确强调了特殊群体在 WISC-IV 测验中的一些可能的表现。

2.2　结构、内容及评分标准的变化

在结构上最大的变化是由两因素结构到四指数结构，之前版本中的言语分量表（VCI）和操作分量表（PRI）由 WISC-IV 中的言语理解指数（VCI）、知觉推理指数 (PRI)、工作记忆指数 (WMI)、加工速度指数 (PSI) 四指数代替，更加关注工作记忆和加工速度的测量，是将一般认知能力转化为测验任务上高表现的执行过程，使对学习障碍的针对性评估诊断成为可能。言语理解指数（VCI）、知觉推理指数 (PRI) 反映的是一般认知能力的综合（即一般能力指数，GAI），而工作记忆指数 (WMI)、加工速度指数 (PSI) 则凸显的是如何将一般认知能力转化为高效的任务表现的执行加工过程（即认知熟练指数，CPI），与学习障碍关系极为密切。

在分测验上，WISC-IV 增加了部分新的分测验，删除了一些过时的题目，在 15 个分测验中，类同、词汇、理解、常识、积木、填图、数字广度、译码、算术、符号搜索是 10 个保留测验，新增了图画概念、字母 – 数字排序、划消测验、矩阵推理和词语推理 5 个新测验。更关注于言语推理而不是晶体智力，更关注于流体智力而不是视觉 – 空间技能。

因此，在评价一个儿童的认知能力和个性化特点时不是仅依靠总智力（FSIQ），而是根据言语理解指数（VCI）、知觉推理指数 (PRI)、工作记忆指数 (WMI)、加工速度指数 (PSI) 的得分，更加全面，精确。正如 Saklofske 所说，总智力用来对天才儿童和障碍儿童进行分类和评价，而四指数用来更好的理解一个儿童是"跟大多数人一样，跟小部分人一样还是都不一样"。

3 不同类型学习障碍在 WISC-IV 各量表上的表现

3.1 阅读障碍

阅读障碍（Reading Disability，简称 RD）包括获得性阅读障碍和发展性阅读障碍。发展性阅读障碍是指具有正常的智力、教育及社会文化机会，没有明显的神经或器质上的损伤，而难以获得有效的阅读技能，在阅读方面表现出特殊的学习困难状态。主要不足表现在口语或拼写句子和文章中的速度和准确率上。一般包括两个水平上的缺陷：一种是解码障碍，主要指字形、语音加工方面存在的缺陷；另一种是阅读理解障碍，主要是指语义加工方面所存在的缺陷。解码障碍是基本的心理加工过程发生障碍，主要表现在汉字和词汇的识别和加工能力上，它在 WISC-IV 上的表现主要体现在排序、命名、译码等需要对字形和语音加工的分测验上；阅读理解障碍是在更高水平上的障碍，它在字词加工上没有障碍，但是理解句子意义发生困难，它在 WISC-IV 上的表现主要体现在字母广度、算术等工作记忆（WM）分测验上。解码障碍儿童在基本的字形和语音上存在障碍，他们不能正确拼读词汇，混淆字词的发音，不能把字形和读音之间建立自动化的联系。因此，他们在词语推理、快速命名等分测验上的表现就不尽如人意。而阅读理解障碍儿童因为不能理解某一单词的含义，所以不能正确读出读音，或者即使读出正确读音，但是也不能理解单词的含义。尤其汉字中有很多的多义字，如"花朵"与"花费"中的"花"的含义就不一样，阅读障碍儿童不能正确理解字的意思，阅读能力受到了限制。

阅读能力与各种认知功能和学业技能有关，因此，不同的缺陷都可能导致阅读障碍，如有单词阅读困难的儿童可能会有语音、符号、语法、流畅性、表达性和工作记忆方面的缺陷，从而导致阅读障碍。在总智商分数上，虽然阅读障碍儿童的总体智力并不存在明显的缺陷，但落后于正常儿童 10 分左右。

更为明显的是，阅读障碍儿童在不同的核心智力分数上表现出更大的差异性，阅读障碍儿童在言语理解指数（VCI）、知觉推理指数（PRI）、工作记忆指数（WMI）、加工速度指数（PSI）上都有较低的得分，言语理解指数（VCI）与词汇发展、接受和表达语言、晶体能力有关，它是预测阅读障碍重要指标。工作记忆指数（WMI）虽然在预测学习障碍上的作用有限，但它在儿童的阅读理解上发挥了重要的作用。值得注意的是，WISC-IV 增加了视觉工作记忆指数（WMI）分测验，从而使量表可以从听觉、言语和视觉三方面全面的揭示阅读障碍儿童的工作记忆缺陷的实质。当考虑到阅读理解水平而非字词解码时，Wechsler 发现，在学习障碍儿童中知觉推理指数（PRI）变异很小，说明学习障碍儿童在阅读较困难的单词或句子时，不能有效地使用问题解决技能（知觉推理指数，PRI 很少被使用），而是仅依靠他们有限的言语理解（VC）和工作记忆（WM）技能。因此，作为预测阅读能力的重要指标，知觉推理指数（PRI）是不可或缺的，这与有些学者的研究是一致的，他们发现阅读困难儿童的右半球功能会对左半球的不足进行补偿。因为阅读功能定位于左半球，左半球受损造成阅读功能缺陷，右半球会自动发展一些补偿功能，这体现了大脑的整体化。阅读障碍儿童在加工速度指数（PSI）分量表上的得分略低于正常组的儿童，虽然加工速度指数（PSI）看起来与阅读理解没有很大的相关，但它在符号加工和从长时记忆中提取单词上具有重要作用，而这些对于流畅的阅读技能是很重要的。

进一步细分阅读障碍亚类型与特定分测验的关系时，Fiorello 将阅读障碍分为四种亚型：语音、正字法、流畅性和整体阅读方面的障碍。语音障碍的特点是在常识、数字广度、算术、矩阵推理和译码测验上的得分偏低；正字法障碍的儿童在译码、符号搜索、划消和算术上有较低的得分，这与阅读障碍儿童在视觉注意、符号表征和加工速度上的缺陷是一致的；流畅性障碍的儿童在数字广度、字母–数字排序、算术和译码测验上的得分偏低；整体阅读障碍在词汇、常识、数字广度和译码测验上有较低的得分。

表明各亚型并没有反映 WISC-IV 因子的整体缺陷，而是反映了各因子之间的变异，这证明了分量表（如常识和算术）的诊断价值和预测效度。同时，为了验证 WISC-IV 的生态学一致性和治疗效度，还需要使用标准的认知和神经心理学的方法来进行综合检验。

分量表在预测阅读能力上有不同的特色，但是要具体分析阅读障碍儿童在某一任务中所需要的问题解决能力和认知能力，将这些分量表结合起来会有更好的效果，如工作记忆指数（WMI）和加工速度指数(PSI)的结合可以反映执行控制的快速命名技能，而言语理解指数 (VCI) 和知觉推理指数 (PRI) 一起则可以反映在阅读时先前知识和问题解决之间的关系。知觉推理指数 (PRI) 分测验与其它认知功能相结合可以很好的预测阅读障碍儿童流体推理能力。流体推理（fluid reasoning）是 Cattell 在晶体智力与流体智力理论中提出的概念，流体推理也就是流体智力，它是指在人们的早期生活和儿童在信息检索过程中的一般认知能力，与空间、数字、概念等学习有关。

3.2 数学障碍

数学障碍（Math Disability, 简称 MD）主要表现在数学知识、算术运算、数学的视觉－空间理解上的障碍，而这种障碍并不能归因于智力的明显缺陷以及一般学习环境和情绪的缺失。数学障碍儿童在智力上的整体得分与同龄人差异不大，但他们却表现出了特定的认知劣势，这主要是由于他们在与数学学习密切相关的认知活动的基本心理过程上缺损。同样，这些特定的认知加工缺陷也能在 WISC-IV 总分及各分量表上展现出来。

Mazzocco 等发现大多数的数学运算和推理的变异可通过言语理解指数（VCI）的晶体能力和知觉推理指数（PRI）的流体推理以及工作记忆指数（WMI）的短时记忆测验来说明，即晶体能力（常识、词汇等）、流体推理（图画、矩阵、词语）及短时记忆（算术、背数）分测验对数学障碍有最强的预测力。有研究者分别使用数字排序、倒背、数字划消、图案接龙、心算等测验来考察数学障碍儿童的数字推理、图案推理和心算能力。结果发现数学障碍儿童在这方面的成绩显著低于正常组，而它们分别与工作记忆、知觉推理和加工速度有关。而知觉推理和工作记忆过程被认为是反映大脑右半球和前额叶的非言语功能，他们在左右大脑功能的融合或者是白质功能上出现异常会导致数学能力出现障碍。当然，加工速度指数（PSI）在预测数学障碍上并不是毫无作用的，它主要在运动操作和心理运动速度上起作用，所以它影响了自动化加工从而影响数学运算速度，导致数学障碍儿童使用更多不成熟的计算策略，如必须使用手指或其他实物帮助运算。

在特定的分测验与数学障碍的关系上，数学障碍也可分为五种亚型：流体推理障碍、数量运算障碍、右半球学习障碍、数字－数量知识障碍、计算障碍－顶叶综合症。他们主要的差异体现在常识、词汇、理解、算术、词语推理、填图测验和符号搜索分量表上，而各亚型在数字广度和译码上没有显著差异，这表明，数学障碍的儿童在这两个分测验上的差异是一致的。五种亚型在个分量表上的表现分别如下：流体推理障碍在矩阵推理、图画概念和算术上有较低的得分；数量运算障碍在常识、数字广度、算术和矩阵推理得分较低；右半球学习障碍在知觉推理指数（PRI）和加工速度指数（PSI）和常识、数字广度和算术分量表上得分都偏低；数字－数量知识障碍主要在加工速度指数（PSI）和数字广度、算术分量表上得分较低；计算障碍－顶叶综合症的最低得分主要在加工速度指数（PSI）和常识、算术、积木设计和填图分量表上。通过上述内容我们可以发现数学障碍儿童不是只有非言语学习障碍，他们在言语理解上可能也会存在问题，如在应用题的分析理解上出现障碍。

3.3 注意缺陷多动障碍

注意缺陷多动障碍（Attention Deficit Hyperactivity Disorder, 简称 ADHD）是一种儿童期常见的行为障碍，发生率在 3-5% 之间，主要表现在自我管理、行为组织和目标导向行为的障碍。ADHD 儿童的核心缺陷是注意缺陷、多动和冲动，典型的行为表现是主动注意能力低下，多动冲动，角色管理失调，组织混乱，

唤醒不足。而这些症状往往与反应抑制、工作记忆等执行功能的缺陷密切相关。Barkley 等人认为 ADHD 根本原因是儿童自我控制能力的缺损，主要缺陷体现在反应抑制上，即一种在信息加工过程中压制无关信息进入工作记忆的能力。一直以来，ADHD 儿童在韦氏智力量表上的表现都得到了充分关注。早先的韦氏智力量表（第二版）中，第三因素，也称为注意力因素、抗干扰因素，它与 ADHD 关系非常紧密，相关的三个分测验（译码、背数和算术），实际上也是认知心理学中考察注意和工作记忆的经典任务，ADHD 儿童在此三项任务上表现出明显的不适应，而在第四版中，这一领域得到进一步加强，ADHD 儿童的临床敏感性进一步增加，主要表现在以下几方面。

在总体智力上，由于在某些分量表上的得分使得整体得分低于正常儿童，这与阅读障碍儿童的 IQ 得分偏低是一致的。WISC- Ⅳ的手册中，提供了 ADHD 儿童与正常儿童比较研究的数据，结果表明，ADHD 儿童的总智商低于正常儿童，但是分数大于 80。

另外，在四项主要的指数，ADHD 儿童表现出明显的不平衡性，在特定的分测验上，ADHD 儿童有着不同程度的敏感性。手册中提到，ADHD 儿童的 WISC- Ⅳ测验分数在加工速度指数（尤其是此项）、言语理解指数、工作记忆指数上的得分相对较低；在分测验上，ADHD 儿童在译码和算术测验上的得分显著低于正常儿童，表现出了最大的效应量。具体来说，ADHD 儿童在工作记忆指数（WMI）和加工速度指数 (PSI) 上的得分显著低于言语理解指数（VCI）和知觉推理指数（PRI）上的得分，而在加工速度指数 (PSI) 中译码测验得分显著高于符号搜索分测验，工作记忆指数（WMI）和加工速度指数 (PSI) 上的低得分是 ADHD 的重要表现，而且在 WISC-IV 中新增的字母 – 数字排序分测验使得算术分测验不只是评价 ADHD 儿童的数学能力，使工作记忆指数（WMI）在预测 ADHD 中更有效。Barkley 认为 ADHD 主要是行为抑制上的问题，也就是执行功能方面的缺陷。而执行功能的测验大多与工作记忆指数（WMI）和加工速度指数 (PSI) 相关，如算术、数字广度、编码等。

不少研究通过数字广度任务发现 ADHD 儿童的言语工作记忆的缺陷，当把数字广度任务分为正背数字和倒背数字任务时，结果是 ADHD 儿童在正背任务上成绩正常，但在倒背任务上落后。也有研究者通过 N–back 任务研究 ADHD 儿童在"冷"、"热"执行功能上的表现，发现 ADHD 在言语工作记忆上存在明显的缺陷。因此，有人认为用一般能力指数（GAI）（工作记忆指数（WMI）和加工速度指数（PSI）的合成分数）来评价 ADHD 就足够了，但是现在证明是不够全面的，因为人们发现 ADHD 儿童在 WISC–IV 的知觉推理指数（PRI）中新增的图画概念和矩阵推理测验中表现出了相对的视觉推理优势。可见，WISC–IV 各分量表在综合评价 ADHD 的缺陷与优势上是很有利的。

进一步考察不同亚型 ADHD 与 WISC–IV 的关系时，DSM– Ⅳ将 ADHD 分为三种亚型：注意缺陷型（ADHD-PI）、冲动 – 多动型（ADHD-PH）和混合型（ADHD-C）。ADHD-PI 的主要表现是注意力分散或不能持续性注意，ADHD-PH 的主要表现是容易分心，多动和冲动，ADHD-C 是兼有这两方面的障碍。它们在具体的分量表上的独特表现可以反映它们各自的特点，如在划消测验上，ADHD-PI 的主要表现是不能集中注意，出现漏划的现象，漏掉部分数字。而 ADHD-PH 的主要表现则是冲动、马虎，出现错划现象，可能会将 9 错当成 6 划掉。而这两种情况都有可能出现在 ADHD-C 中。分量表的不同特点显然为 ADHD 及其亚型的诊断和分类提供了有利的证据。

4 WISC–IV 对于学习障碍的干预价值

韦氏儿童智力量表在特定学习障碍上的评估发挥了重要作用，根据这种特异性诊断而进行的针对性干预也必将更有价值。

4.1 引导教师和家长形成关于学习障碍儿童正确的观念

临床工作者要充分认识到学习障碍儿童智力发

展的不平衡不是智力缺陷，这是在对家长和教师进行辅导时要注意的。这种不平衡既有先天的因素，也是环境的结果，它与学习障碍本身是交互式发展的。先天的一些缺陷，如基础性视听加工技能落后导致了阅读能力的落后，而后者又影响了个体经验的获得，从而导致了不利的智力得分。因此，临床人员应将儿童看作是独立的个体，在解释测验时将智力以外的因素考虑在内。更为重要的是家长和教师要改变对学习障碍儿童的偏见，不要将学习障碍儿童的学习困难看作是态度问题，应该无条件地接纳他们，要调整观念，了解他们的不足之处，降低期望，以鼓励、赞扬的方式为主来激发学习障碍儿童的学习动机，帮助他们提高学习能力。同时也要认识到这些儿童并不是在所有的方面都存在缺陷，他们还有很多长处需要被发现和认可。

同时要重视非智力因素－意志力，道德修养，自信，自立，自强及克服困难的勇气等在儿童的成长和发展中的重要作用，重视学生综合素质的培养，而不仅仅是智力水平。注意到智力与非智力因素的相互影响，这种影响既有补偿作用，也有阻碍作用，因此，在学习障碍儿童的干预训练过程中，既要防止不利因素的伴随发生导致更为复杂的学习障碍，又要充分利用学习障碍儿童的特长来补其缺陷，促进知识的获得。例如，利用 ADHD 儿童的兴趣来促进其知识的获得，使用强化物来训练其注意缺陷，对于数学障碍的儿童采取实物、影像等形象化教学，增进学习热情，提升儿童智力。

如果家长和教师能够做到耐心教育和具体帮助学习障碍儿童，说明他们已经充分的认识到了学习障碍儿童的看似"顽皮不认真"的糟糕表现实则为学习能力的基本缺陷而非态度或认识问题。这种观念的调整来之不易，需要在专家的指导下才能获得。但是，观念的转变带来的干预效果不仅仅限于具体的教育帮助策略，更加重要的是，家长和教师调整了对于学习障碍儿童的不合理期望，调整了对自己之前看似"徒劳无功"的努力的不认可，并且降低了情绪困扰，提升了教学效能感和家庭幸福感，促进了亲子关系和师生关系的良性发展，而这反过来对学习障碍儿童又产生了积极的作用。

4.2 根据剖面图分析进行针对性的认知加工训练

WISC–IV 的手册使用剖面图分析来解释个体的相对优势与劣势，还提供了有关分测验之间的得分差异和各指标分数间的差异是否达到统计显著性的评估数据，这不仅表明了被试在同一类别的人群中的相对位置，还为我们判断个体内在的相对认知优势和劣势提供了依据。针对性的认知加工训练对于不同类型的学习障碍来说，应当受到重视。类似的，以智力的 PASS 模型为基础的"PASS 阅读增强方案"（PASS Reading Enhancement Program）研究取得了很好的效果，它是 Das 提出的一个阅读障碍矫治方案，训练内容包括继时加工任务，包括移动矩阵、连接形状、窗口排序和连接字母；同时性加工任务，包括句子矫正、追踪、形和物体、形状设计；第三类任务牵涉两种加工任务，包括相关的记忆集合、矩阵。此外，还有一些提高理解力、获取句子和段落意义的训练任务。总之，它以认知加工策略的动态获得为主，与工作记忆指数（WMI）密切相关，对于阅读综合技能的训练，提高认知综合测验和阅读理解任务的成绩，有着重要的作用。而针对 ADHD 儿童主要在工作记忆指数（WMI）上的缺陷，Klingberg 采用计算机化的工作记忆训练 ADHD 儿童也有显著地效果，训练采用阶梯训练法，即根据被试工作记忆容量及在训练任务中的表现，逐步调整训练难度，提高工作记忆能力。还有数字广度训练、Stroop 干扰任务等扩展工作记忆的容量。训练有效地提高了 ADHD 儿童的工作记忆，同时，在改善反应抑制和推理、减少注意缺陷症状上也取得了重大的突破。同样，应该对数学障碍、书写障碍等其他类型的学习障碍儿童进行针对性的阅读、计算、注意、知觉组织以及动作协调训练等等。

4.3 为学习障碍儿童的教学干预提供指导

由于学习障碍儿童智力发展的不平衡性，教师应该采取有针对性的教学指导，如对阅读障碍儿童进行单词阅读、快速命名和流利阅读训练有助于提

高阅读成绩；对于 ADHD 儿童，使用笔记本、收音机等帮助记忆，充分的指导和及时的支持和反馈，及时的调整等都可以从中获益。对特定的认知技能的缺陷进行训练的同时，也应该发挥学习障碍的优势认知技能进行补偿，如利用 ADHD 儿童在图画推理的优势，采取多种形式的教学。另外，结合特定的学业任务进行教学干预也是有必要的，尤其是关注到特定学习障碍在特定学业领域上的独特表现，例如，在考虑到数学学习障碍儿童在应用题问题解决上存在困难时，Hale 等人认为数学障碍儿童在解决数学问题时需要特殊的指导，家长和教师应该在运算步骤、综合分析数学问题、学习和使用数学运算、有效地问题解决方案方面进行充分的指导。重点是，我们要关注如何矫正与数学能力相联系的心理认知过程的缺陷和对加工过程的缺陷进行有效地补偿，使数学障碍的儿童有更好的表现。

5 对于 WISC-IV 在学习障碍诊断和干预中的建议

5.1 不能将统计差异等同于临床差异

使用 WISC-IV 量表得分时要注意的一个问题就是统计上的差异不代表临床上的显著，不能因为某一低得分就诊断有学习障碍，它只是代表了有学习障碍的可能性。同样，在工作记忆指数（WMI）和加工速度指数（PSI）上的低分不能诊断受测者为 ADHD，而是应将 ADHD 这个可能性考虑进来。再比如，某儿童在言语量表上的得分显著低于操作量表，缺乏经验的诊断者可能仅凭这一点就判定他左半球功能受损，但他如果知道这个儿童前几天因车祸头部受伤，正好言语相关区域损伤导致得分降低，这就正好证实了之前的假设。因此在作出判断是一定要使用其它临床标准而不仅是测验分数。正如 Allport 所言，理论和临床表现综合描述才能看到一个人的独特个性与人格气质，分数只是帮助理解一个人的工具而已。

5.2 将 WISC-IV 与其它量表综合使用

为了确保全面地评估儿童的优势及不足，获得更为精确的信息，WISC-IV 常与其它量表综合使用。如与《韦氏个体成就测验（第二版）》（Wechsler Individual Achievement Test-Second Edition, WIAT-II；The Psychological Corporation, 2001）共同使用，以获得儿童认知能力和学业成就两方面的信息。而学习者过程评价第二版 – 阅读和写作诊断（Process Assessment of the Learner, Second Edition Diagnostic for Reading and Writing, PAL II RW）中的语音编码、语言的接受与表达、正字法编码等分测验在评价阅读障碍上与 WISC-IV 测验相吻合，有助于做出更全面的评价。此外，WISC-IV 还常与语言基础的临床评价（Clinical Evaluation of Language Fundamentals, ELF 4）等综合使用。在中国，当 WISC-IV 与《中国精神障碍分类与诊断标准（第三版）》（CCMD-3）、《标准化阅读理解和识字量测验》以及《学业成就测验》等工具综合使用时，在诊断 ADHD 和阅读障碍上更符合中国儿童的现状，有利于做出更合理的治疗方案。

5.3 由因子分析走向神经认知分析

学习障碍儿童伴随不同程度的认知功能障碍，阅读障碍儿童在阅读、理解语义上有困难，数学障碍儿童在计算、视空理解上有困难，而 ADHD 儿童在工作记忆、加工速度上都有一定的障碍，因此结合不同学习障碍的特点，应更加重视神经认知的分析。根据认知功能诊断做出个性化评估，有利于评估者做出科学的干预治疗方案。尤其是现在心理学研究中眼动、脑电和一系列标准化诊断方法，如 CPT（Continuous Audio-Visual Performance Test）在 ADHD 中的应用，为研究学习障碍的神经机制方面提供了更加广阔的空间。

参考文献

陈海平 . (2008). 韦氏儿童智力测验第四版的修订及其对智力测验开发的启示 . *宁波大学学报（教育科学版），* 30(6), 37– 42.

孔明 , 黄启兵 . (2007). WISC– IV 的新变化及其在特殊

教育评估中的作用 . *中国特殊教育*, 6(84), 28– 34.

徐建平 . (2006). 基于认知加工的阅读增强方案 : PREP 评述 . *中国特殊教育*, *67*(1), 83–90.

张微 , 刘翔平 , 宋红艳 . (2010). "热" 执行对注意缺陷多动障碍和阅读障碍儿童言语工作记忆的影响 . *心理学报*, *42*(3), 415–422.

中华医学会精神病学分会 . (2001). *中国精神障碍分类与诊断标准 . 第 3 版* . 济南 : 山东科学技术出版社 .

周世杰 , 杨娟 , 张拉艳 . (2006). 工作记忆、执行功能、加工速度与数学障碍儿童推理和心算能力的关系 . *中国临床心理学杂志*, *14*(6), 574–578.

Allport, G. W. (1995). *Becoming: Basic Considerations for a Psychology of Personality*. New Haven: Yale University Press.

Aurelio Prifitera, Donald H. Saklofske. (2008). WISC– IV Clinical Assessment and Intervention– Second Edition.

Badian, N. A. (1997). Dyslexia and the double deficit hypothesis. *Annals of Dyslexia, 47*, 69– 87

Barkley, R. A. (1990). Attention Deficit Hyperactivity Disorder: A Handbook for Diagnosis and Treatment.

Berninger, V., Donnell, L. (2005). Research– supported differential diagnosis of specific learning disabilities. WISC– IV Clinical Use and Interpretation, 189– 233.

Bowers, P. G. (2001). *Dyslexia, Fluency, and the Brain*. MD: YorR. Press, 41– 64.

Breier, J. I., Simos, P. G., Fletcher, J. M., et al. (2003). Abnormal activation of temporoparietal language areas during phonetic analysis in dyslexia. *Neuropsychology, 17,* 610– 621.

Daniel, M. H. (2007). Scatter and the construct validity of FSIQ:Coment on Fiorello et al. Applied *Neuropsychology, 14*(4) , 291– 295.

Emilio Ferrer, Elizabeth D. O., & Hare, Silvia A Bunge. (2009). Fluid reasoning and the developing brain. *Front Neurosci, 3*(1), 46– 51.

Evans, J. J., Floyd, R. G., McCrew, K. S., et al. (2002). The relations between measures of Cattell – Horn ‒ Carroll

(CHC) cognitive abilities and reading achievement during childrenhood and adolescence. *School Psychology Review, 31,* 364– 372.

Fiorello, C. A., & Hale, J. B. (2007). Interpreting intelligence test for children with disabilities: is global intelligence relevent? Applied *Neuropsychology, 14*(1), 2– 1.

Fiorello, C. A., Hale, J. B., & SnyderL, E. (2006). Cognitive hypothesis testing and response to intervention for children with reading disabilities. *Psychology in the Schools, 43*, 835– 854.

Flanagan, D. P., & Kaufman, A. S. (2004). Essentials of WISC– IV Assessment.

Geary, D. C., & Hoard, M. K. (2001). Numerical and arithmetical deficit in learning–disabled chidren: relation to dyscalculia and dyslexia. *Aphasiology, 15,* 635– 647.

Hale, J. B., Naglieri, J. A., Kaufman, A. S., et al. (2004). Specific learning disability classification in the idividuals with Disabilities Education Act: the danger of good ideas. *The School Psychologist, 58*(1), 6–14.

Ida, S. B. (2005). Test Review: Wechsler intelligence scale for children–fourth edition(WISC– IV) .*Children Neuropsychology*, (11) :471– 475

James, B. H., Catherine, A. & Fiorello. (2008). Differential ability scales–second edition (Neuro) psychological predictiors of math performance for typical children and children with math disabilites. *Psychology in the Schools, 45*(9), 839– 858.

Kaufman, A. S. (1975). Factor analysis of theWISR– R at eleven age levels between 6 1? 2 and 16 1? 2years. *Journal of Consulting and Clinical Psychology, 43,* 135–147.

Klingberg, T. (2005). Computerized training of working memory in children with ADHD – a randomized, controlled trial. *Journal of American Academic Child Adolesc Psychiatry, 44*(2) , 177– 186.

Leffard, S. A., & Miller, J. A. (2006). Sustantive validity of working memory measures in cognitive functioning test batteries. *Applied Neuropsychology, 13*, 230– 241.

Lyon, G. R. (2005). Why scientific research must guide educational policy and instructional practices in learning disabilites. *Learning Disability Quarterly, 28*, 140– 143.

Matarazzo, J. D. (1990). Psychological assessment versus psychological testing: validation from binet to the school, clinic and courtroom. *American Psychologist, 45*(9), 999– 1017.

Mather, N., & Woodcock, R. W. (2001). Test of cognitive abilities examiner's manual. Itasca, IL: Riverside Publishing, 2– 3.

Mayes, S. D., & Collhoun, S. L. (2006). WISC–IV and WISC–III profile in children with ADHD. *Journal of Attention Disoders*, (9), 486.

Mazzocco, M. M. M. (2001). Math learning disability and math LD subtypes: evidence from studies of Turner Syndrome, Fragile X Syndrome, and Neurofibromatosis Type 1. *Journal of Learning Disbilies, 34*, 520– 533.

McInnes, A., Humphries, T., Hogg, Johnson, S., et al. (2003). Listening comprehension and working memory are impaired in Attention Deficit Hyperactivity Disorder irrespective of language impairment. *Journal of Abnormal Child Psychology, 31,* 427–443

Proctor, B. E., Floyd, R. G., & Shaver, R. B. (2005). Cattell Horn Carroll broad cognitive ability profile of low math achievers. *Psychology in the Schools, 42,* 1–12.

Semel, E., Wiig, E. H., & Secord, W. (2003). Clinical Evaluation of Language Fundamentals Fourth Edition (CELF – 4).

Susan, D. M., & Susan, L. C. (2007). Wechsler intelligence scale for children third and fourth edition predictor of academic achievement in children with Attention Deficit? Hyperactivity Disorder. *School Psychology Quarterly,* (2), 234–249.

Wechsler, D. (1950). Congnitive and non–interactive intelligence. *American Psychologist*, (4), 5–7.

Wechsler D. (2003). Wechsler Intelligence Scale for Children– Fourth Edition (WISC– IV) : Technical and Interpretive Manual.

The value of WISC-IV in the identification and Evaluation of learning disabilities

ZANG Ling[1,2] , ZHANG Wei[1,2] ， LIU Huashan[1,2]

（ [1]Key Laboratory of Youth Network Psychology and Behavior, Ministry of Education ）

（ [2]School of Psychology, Central China Normal University, Hubei Province Key Lab for Human Development and Mental Health, Wuhan 430079, China ）

Abstract　Children with learning disabilities have difficulty in basic cognitive control processes such as listening, speaking, reading and calculating. Different kinds of learning disabilities have specific manifestation in Wechsler Intelligence Scale, due to its inherent structure is connected with specific cognitive processes. In order to enhance clinical effect, WISC–IV has modulation in framework, content and function; especially strengthen clinical discrimination for special learners and emotional and behavioral disorders. Various subtests for different learning disabilities, and cognitive process indices, provide standard for clinical identification, what's more, they also offer references for children with learning disabilities in individual cognitive behavior training.

Keywords　WISC–IV; learning abilities ; evaluation ; intervention

脑与认知研究所团队成员

周治金，博士，教授，脑与认知研究所所长，主要研究领域为创造性思维的认知神经基础，创新性教学等。

刘思耘，博士，教授，主要研究领域为语言的加工与表征，言语知觉，二语习得过程，具身认知。

赵庆柏，博士，副教授，脑与认知研究所副所长，主要研究领域为认知神经科学，网络语言，基于网络的问题解决。

定险峰，博士，副教授，主要研究领域为从具身认知的角度探讨数量、时间和权力等抽象概念的心理表征和神经机制。

范炤，博士，副教授，主要研究领域为视知觉，注意与工作记忆，网络心理学，虚拟现实在心理学中的应用。

高闯，副教授，研究领域为数理心理学、认知心理学，认知神经科学，实验心理学，网络心理学。试图从学理理论，对心理学经验研究领域进行重构，确立数理心理学理论体系。

程晓荣，博士，副教授，主要研究领域为儿童语言发展，语言与阅读的认知加工机制，视知觉与注意。

成良，博士，副教授，主要研究领域为利用神经电生理方法（TDT、ERP、fMRI 等）研究感觉信息加工及脑高级认知活动的神经机制，网络成瘾对脑高级认知活动的影响机制，以动物为研究对象采用分子生物学和遗传学手段研究学习记忆和创造力的神经机制。

李玉杰，博士，讲师，主要研究领域为注意，老龄化，刻板印象。

汉语成语谜语问题解决中思路竞争的眼动研究

黄福荣 [1,2]，周治金 [2]，赵庆柏 [1]

（[1] 青少年网络心理与行为教育部重点实验室 ; 湖北省人的发展与心理健康重点实验室 ; 华中师范大学心理学院 , 武汉 430079）

（[2] 北京市"学习与认知"重点实验室 ; 首都师范大学心理系 , 北京 100048）

摘　要　在谜语问题解决过程中 , 可能存在着通过简单联想和新异联想寻找答案这两种思路。两个实验中设置了包括寻常答案与新颖答案在内的若干备择答案 (实验 1 为 4 个 , 实验 2 为 6 个) 供被试选择 , 利用眼动技术记录被试在解题过程中的不同时间段内对新颖答案与寻常答案的平均注视时间 , 考察汉语成语谜语问题解决中两种思路之间冲突的过程。实验 1 操纵了任务要求 , 实验 2 操纵了规则线索的有效性。实验结果表明：（1）在成语谜语问题解决中的一段时间内新异联想和简单联想能够同时发生且形成竞争；（2）选择"新颖且合适答案"的任务要求 , 提高了成功形成新颖语义联结的概率 , 但是并没有加快新异联想发生、发展的进程 , 也没有改变两种思路相互竞争的局面；（3）有效的规则线索可以抑制简单联想 , 阻止其发生 , 同时可以加快新异联想发生、发展的进程。

关键词　成语谜题 ; 新异联想 ; 简单联想 ; 眼动

分类号　B842

1　引言

在顿悟问题的解决过程中 , 问题解决者往往会经历一个明显的"僵局" (impasses), 感到思维停滞不前 (Schooler, Ohlsson, & Brooks, 1993)。关于僵局产生的原因 , "心理定势理论"认为思维习惯会使问题解决者不再致力于寻找新的、可能是更为有效的问题解决途径。"心理成规理论"认为 , 僵局的产生是由于反复尝试使用某种错误的搜索路径增强了该路径的激活程度 , 相应地降低了搜索其他解题路径的可能性 , 它强调问题解决过程中的"强迫症"倾向。"功能固着理论"认为物体的心理表征与它的常用功能属性相关 , 如果在特定的问题中 , 需要利用它的不寻常的

功能属性 , 自动提取的常用功能属性会阻碍不寻常功能属性的激活和利用 , 僵局就产生了 (罗劲 , 2004)。

那么 , 僵局又是怎样被打破的 ? "表征转换理论"认为顿悟问题的成功解决取决于问题表征方式的转换 , 它可以通过限制消除、组块破解等多种方式对无效的、错误的初始表征进行重构 (Knoblich, Ohlsson, Haider, & Rhenius, 1999), 也可能只需要额外注意和编码问题中的其他关键信息 , 对初始表征进行修复就可以打破僵局 (Kaplan & Simon, 1990)。"进程监控理论"的基本观点是 , 问题解决者会依据将要达到的问题目标状态 , 确定一些看似有效的内在标准 , 并根据这些标准来监控每一个局部行动的有效性。一旦个体意识到这些行动不会成功时 , 就会产生一种内驱力 , 促使个体去解除过去知识经验的限制 , 寻找其它解题途

径 (Chronicle, MacGregor, & Ormerod, 2004; MacGregor, Ormerod, & Chronicle, 2001)。

上述有关顿悟问题解决的两类理论中，前一类理论着眼于说明采用常规的思维方式，就会形成无效的解题思路，从而使问题解决陷入僵局；后一类理论则着眼于说明新颖（有效）的解题思路是如何形成的。由此可见，在顿悟问题解决过程中，可能存在着无效的常规解题思路与有效的新颖解题思路之间的冲突。有关顿悟神经生理机制的研究为两种思路之间的冲突提供了证据。例如，脑功能成像研究发现，当顿悟发生时，与认知冲突的监控相关的前扣带回被激活了 (Aziz-Zadeh, Kaplan, & Iacoboni, 2009; Kounios et al., 2006; 罗劲，2004)。有关顿悟感产生的 ERP 研究也发现，与没有顿悟感的条件相比，顿悟感出现的条件下，新颖答案诱发出 N320/N380 成分，该成分溯源于前扣带回 (买晓琴，罗劲，吴建辉，罗跃嘉，2005; 邱江，罗跃嘉，吴真真，张庆林，2006)，这说明顿悟问题解决过程中包含着两种解题思路之间的冲突。

在两种思路的冲突中，常规的无效思路对新颖的有效思路必然起着干扰和阻碍作用 (öllinger, Jones, & Knoblich, 2008)。虽然表征转换理论和进程监控理论认为，顿悟的产生需要解除过去知识经验的限制，但是没有阐明过去知识经验（即无效解题思路）与新颖有效的解题思路之间冲突的发生、发展过程。毕竟，新颖有效的思路需要战胜常规无效的思路才能出现顿悟。所以，仅描述新颖有效思路形成过程不足以全面反映顿悟产生过程，而深入探究新颖有效思路的发生、发展过程，以及新颖思路战胜常规无效思路的过程，才能更深入地揭示顿悟问题解决的规律。

从表征转换理论可以推论，新颖思路与常规思路之间转换可能是瞬间完成的。罗劲 (2004) 认为，顿悟是问题解决视角的瞬间"新旧交替"过程。有关顿悟问题解决中预热感判断和对关键信息注视时间的变化为此观点提供了实验证据。例如，Metcalfe 考察了被试在问题解决中接近问题解决的预热感 (feeling of warmth, FOW)，研究发现被试对常规问题解决的预热感是逐渐上升的，但是对顿悟问题解决的预热感在问题解决前一直很低，在问题解决瞬间突然上升 (Metcalfe, 1986; Metcalfe & Wiebe, 1987)。Knoblich, Ohlsson 和 Raney (2001) 分析了被试解决"火柴棒算式问题"时的注视时间的变化趋势，发现关键元素的注视时间在问题解决前陡增。这些研究结果倾向于支持顿悟问题解决中新颖有效的解题思路是突然形成，并且在瞬间替代了常规无效的解题思路。但是，这些研究并未直接考察新旧两种思路之间冲突的发生、发展过程。

从进程监控理论可以推论，新颖思路是一个逐渐发生的过程。一些实验研究也确实发现新颖思路不是突然发生的。例如，Yaniv 和 Meyer (1987) 认为 Metcalfe (1986, 1987) 的结论只适用于语义信息不丰富的顿悟问题，在语义信息丰富的顿悟问题解决过程中，与答案相关的语义信息在酝酿期是逐渐积累的。Ellis, Glaholt 和 Reingold (2011) 通过分析被试在解决英语字谜问题过程中的注视时间发现，被试对关键词素和干扰词素的注视时间比例在问题解决初期是没有显著差异的，直到问题解决前几秒钟，干扰词素的注视时间比例逐渐下降。这说明，在语义信息丰富的顿悟问题解决中，新颖有效的解题思路和常规无效的解题思路有可能同时发生，并相互竞争。

由此可见，深入研究顿悟问题解决过程中，新颖有效思路的发生、发展过程及其影响因素，以及新颖思路战胜常规无效思路的过程，是揭示顿悟问题解决认知机制的一条重要途径。谜语问题属于语义丰富的顿悟问题，在谜语问题解决中存在两种解题思路，一种是根据谜面语义所进行的简单联想，另一种是寻找并激活谜面（或谜底）不寻常的语义信息所进行的新异联想 (买晓琴，罗劲，吴建辉，罗跃嘉，2005; 邱江，罗跃嘉，吴真真，张庆林，2006; 沈汪兵，刘昌，张小将，陈亚林，2011; 朱新秤，李瑞菊，周治金，2009)。所以，本研究拟采用汉语成语谜语问题为实验材料，探讨实验任务要求与规则线索的有效性对谜语问题解决中两种解题思路发生、发展过程的影响。在实验任务的选择上，虽然答案生成任务最能体现问题解决的过程与特点，但是采用此实验任务难以监测两种思路发

生、发展的进程。采用呈现答案并确定顿悟感的判断任务，又难以反映完整的问题解决过程。所以，本研究采用选择答案的任务范式，为每一道谜题设置寻常答案、新颖答案 (谜底)、似是而非答案和无关答案等几类备择答案，采用眼动追踪技术记录被试在解题过程中对新颖答案与寻常答案的注视时间，来推测谜语问题解决中新异联想的发生、发展过程，以及新异联想与简单联想之间冲突的过程。

之所以选择采用眼动追踪技术，是因为传统认知实验心理学研究方法，以及 ERP 或 fMRI 等认知神经科学技术，都难以直接考察顿悟问题解决中两种解题思路发生、发展及其相互竞争的过程；而眼动记录仪可以直接"记录"问题解决的进程。近几年来，越来越多的研究者使用这种技术研究问题解决过程 (Bilalic, McLeod, & Gobet, 2008; Kaller, Rahm, Bolkenius, & Unterrainer, 2009; Patsenkoa & Altmanna, 2010; Thomas & Lleras, 2007) 和顿悟问题解决过程 (Ellis, Glaholt, & Reingold, 2011; Grant & Spivey, 2003; Jones, 2003; Knoblich, Ohlsson, & Raney, 2001; Thomas & Llerasb, 2009)。Knoblich 等人 (2001) 除了统计分析顿悟问题解决过程中的总注视时间以外，还把整个问题解决过程分成三个相等的时间阶段，分析不同阶段中注视时间的变化。Ellis, Glaholt 和 Reingold (2011) 把英语字谜问题解决过程按照 100ms 的标准划分成若干相等的时间阶段，分析注视时间的变化趋势，这种数据分析方法能更好地考察创造性问题解决的进程。所以，本研究拟将谜语问题解决过程中的眼动数据进行分时间段分析，考察成语谜语问题解决中新颖有效思路和常规无效思路的发生、发展进程。

2 实验 1

2.1 实验目的

采用眼动仪记录被试解决成语谜题的眼动轨迹，首先对新颖答案和寻常答案的相对注视时间与选择比例进行相关分析，证实对新颖答案和寻常答案的注视能够敏感反映新异联想和简单联想。然后分析解

题过程中的每一时间段内对新颖答案、寻常答案与无关答案的注视时间差异，考察新异联想和简单联想的发生、发展进程，检验两种联想之间是否会发生竞争。本实验还操纵了指导语，尝试检验任务要求是否会改变两种联想的发生、发展进程和竞争过程。

2.2 实验方法

2.2.1 实验设计　实验 1 采用单因素组间设计，自变量为指导语中的任务要求 (合适组 vs 新颖组)，合适组要求被试选择合适答案，新颖组要求被试选择新颖且合适的答案，并且提示被试不是所有与谜面之间存在语义关联的备择答案都满足新颖性要求。新颖组被试在解题过程中可能会意识到存在两个或两个以上的答案，但是实验中并没有给予如何处理的指示。因变量为选择新颖答案和寻常答案的百分数，以及对各类备择答案兴趣区的注视时间。

2.2.2 被试　44 名来自武汉某大学的本科生参加了本实验，其中男生 20 名，女生 24 名。将被试随机分成两组，每组 22 人。所有被试的视力或矫正视力均正常，实验结束后被试获得一件小礼物。

2.2.3 实验仪器　实验仪器是加拿大 SR 公司生产的 Eyelink 1000 型眼动仪，采样频率设置为 1000 Hz，采用瞳孔 + 角膜模式采集数据。呈现刺激的显示器为 21 inch，刷新频率为 60Hz，分辨率为 1024×768。

2.2.4 实验材料　从汉语成语谜语库中 (朱新秤，李瑞菊，周治金，2009) 筛选了 48 道谜题，为每道谜题设置了 4 个备择答案，包括一个新颖答案、一个寻常答案和两个无关答案。新颖答案是成语谜语词典中所提供的谜底，它与谜面之间存在内隐而非直接的语义关联。在预实验中请一组被试根据谜面进行自由联想，生成与谜面之间存在语义关联的答案，然后从被试根据谜面所生成的答案中筛选出生成概率最高的成语作为该谜题的寻常答案。无关答案是从被

表 1　实验 1 中使用的实验材料举例

谜面	寻常答案	新颖答案	无关答案 I	无关答案 II
破晓	旭日东升	毁于一旦	良莠不齐	寿终正寝
越做越快	熟能生巧	积劳成疾	察言观色	言简意赅
善战而多谋	足智多谋	精打细算	旧地重游	孤苦伶仃
交通业的兴起	四通八达	应运而生	洁身自好	精雕细刻

试根据其它谜面生成的答案中选择的，且与谜面不存在语义关联的成语。表1列举了实验1中使用的4道谜题及其备择答案。

2.2.5 实验程序 实验在隔音的眼动实验室内完成。实验开始前，被试的前额和下颚放在托架上以固定头部，眼睛距离显示器大约75 cm。进行校准和确认后，被试先进行12道题的练习，熟悉实验程序之后进入正式实验。正式实验流程是：首先在屏幕中央呈现一个"十"注视点500 ms，然后呈现谜面4000 ms，之后谜面不消失，并且在谜面下方呈现4个备择答案（4个备择答案的位置进行了平衡），要求被试在6000 ms内按要求选择答案。间隔2000 ms后，呈现下一道谜题。

正式实验中呈现备选答案的同时，眼动仪开始记录被试的眼动轨迹等数据，被试按键选择答案之后，眼动仪停止记录。每一道谜题呈现之前，都会进行一次漂移校正以保证眼动记录的准确性。

2.2.6 数据处理 首先在备择答案图片上划分出5个兴趣区：谜面、寻常答案、新颖答案、无关答案I、无关答案II。通过SR公司提供的Dataviewer数据分析软件导出各个兴趣区的眼动数据，然后使用SPSS 11.5统计软件包进行统计分析。

2.3 结果与分析

2.3.1 选择答案的百分数 表2所示的是两组被试选择寻常答案与选择新颖答案的百分数。

表2 选择寻常答案和新颖答案的百分数

任务要求	选择寻常答案	选择新颖答案
合适组	0.62（0.11）	0.32（0.11）
新颖组	0.38（0.22）	0.48（0.18）

注：括号内数据为标准差。

对寻常答案的选择百分数进行独立样本T检验，结果显示，$t(42)=4.75$, $p < 0.001$，合适组的选择百分数更高。对新颖答案的选择百分数进行独立样本T检验，结果显示，$t(42)=3.63$, $p < 0.005$，新颖组的选择百分数更高。

2.3.2 问题解决过程中的总注视时间

表3所示的是两组被试选择寻常答案与选择新颖答案过程中，分别对新颖答案、寻常答案和无关答

案兴趣区的总注视时间（该兴趣区内所有注视点的总时间）。其中无关答案区的总注视时间是两个无关答案选项兴趣区的总注视时间的平均值。

为了检验对新颖答案区、寻常答案区的注视与新异联想、简单联想的关系，对新颖答案和寻常答案的选择百分数与三类兴趣区注视时间比例之间分别进行相关分析。兴趣区注视时间比例是指每类答案区的总注视时间占三类答案兴趣区的总注视时间的百分数。结果显示，新颖答案的选择百分数与对新颖答案兴趣区注视时间的比例呈显著正相关，$r = 0.86$, $p < 0.01$，与对寻常答案区注视时间的比例呈显著负相关，$r = -0.85$, $p < 0.01$，与对无关答案区注视时间的比例也呈显著负相关，$r = -0.85$, $p < 0.01$。寻常答案的选择百分数与对新颖答案区注视时间的比例呈显著负相关，$r = -0.83$, $p < 0.01$，与对寻常答案区注视时间的比例呈显著正相关，$r = 0.91$, $p < 0.01$，与对无关答案区注视时间的比例呈显著负相关，$r = -0.91$, $p < 0.01$。

2.3.3 选择新颖答案过程中的不同时间阶段内的注视时间 表4所示的是两组被试在选择新颖答案过程中的不同时间阶段内对不同兴趣区的注视时间。

为了进一步考察谜语问题解决的加工进程，本实验参考Ellis等人(2011)和沃建中等人(2006)所采用的分时间阶段考察注视轨迹的数据分析方法。具体做法是：取每500 ms为一个时间样本，计算被试在每一个时间样本内对三类答案区的注视时间。实验1统计分析了0~3.5 s的时间范围注视轨迹的数据，作此选择的理由是：被试选择新颖答案的平均时间是3441(±514) ms，我们对实验数据的预处理发现，在3.5s前后的三个时间阶段内，对各兴趣区注视时间之间的数量关系比较稳定，说明无论完成答案选择的时间是短于平均选择时间，还是长于平均选择时间，答案选择过程中所涉及的思维成分与特点具有一致性。

对于选择新颖答案的解题过程的注视时间进行2(任务要求：合适组；新颖组)×3(兴趣区类型：寻常答案区；新颖答案区；无关答案区)×7(时间段：0~0.5 s; 0.5~1 s; 1~1.5 s; 1.5~2 s; 2~2.5 s; 2.5~3 s; 3~3.5 s)三因素方差分析，重点分析兴趣区与时间段的交互

表 3　问题解决过程中的总注视时间 (ms)

任务要求	选择类型	寻常答案区	新颖答案区	无关答案区
合适组	选择寻常答案	884（0.47）	556（0.30）	442（0.23）
	选择新颖答案	548（0.26）	1098（0.53）	433（0.21）
新颖组	选择寻常答案	1031（0.47）	653（0.29）	524（0.24）
	选择新颖答案	532（0.24）	1226（0.56）	438（0.20）

注：括号内数据为注视时间比例。

表 4　选择新颖答案过程中的不同阶段内的注视时间 (ms)

任务要求	兴趣区	0~0.5 s	0.5~1 s	1~1.5 s	1.5~2 s	2~2.5 s	2.5~3 s	3~3.5 s
合适组	新颖答案区	7	62	130	153	146	146	139
	寻常答案区	4	21	95	83	110	76	50
	无关答案区	4	50	89	73	62	54	36
新颖组	新颖答案区	4	53	138	155	170	170	145
	寻常答案区	2	40	98	89	90	74	56
	无关答案区	6	48	75	67	54	56	46

作用，考察在不同时间段内对三类兴趣区注视时间的差异。若对新颖答案区或寻常答案区的注视时间长于无关答案区，说明新异联想或简单联想已经发生了，若对新颖答案区的注视时间也长于寻常答案区，说明新异联想开始占据主导地位。结果显示，任务要求的主效应不显著，$F(1,42) = 0.60, p = 0.44$；兴趣区类型的主效应显著，$F(2,84) = 295.18, p < 0.001$；时间段的主效应显著，$F(6,252) = 207.10, p < 0.001$；任务要求与兴趣区类型的交互作用不显著，$F(2,84) = 1.03, p = 0.36$；任务要求与时间段的交互作用不显著，$F(6,252) = 0.72, p = 0.63$；兴趣区类型与时间段的交互作用显著，$F(12,504) = 14.25, p < 0.001$，从 1s 开始，对新颖答案区的注视时间长于无关答案区，在 2~2.5s 和 2.5~3s 阶段，对寻常答案区的注视时间长于无关答案区，并且从 0.5s 开始对新颖答案区的注视时间长于寻常答案区；三因素交互作用不显著，$F(12,504) = 1.19,$
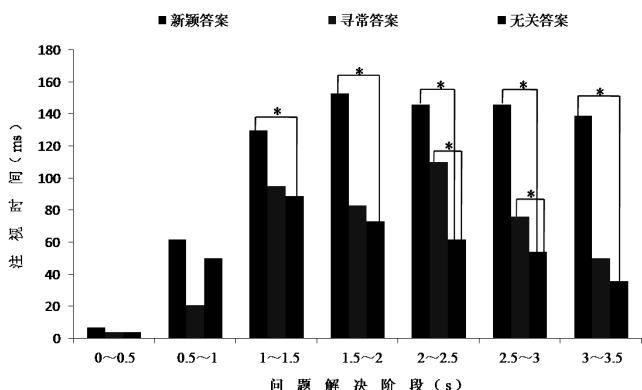
图 1　合适组选择新颖答案的注视时间进程
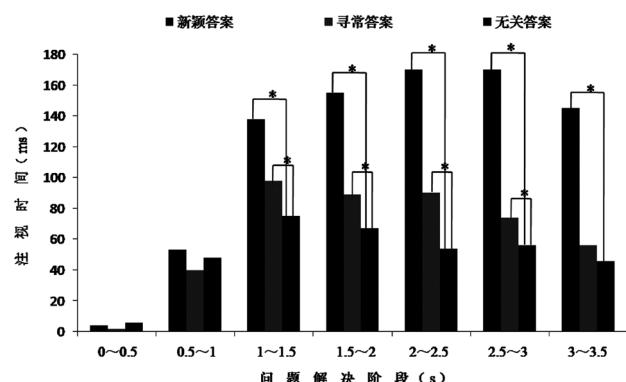
图 2　新颖组选择新颖答案的注视时间进程

$p = 0.28$。合适组和新颖组的注视时间进程分别如图 1、图 2 所示。

在三因素方差分析基础之上，选取新颖答案区和寻常答案区的注视时间都显著长于无关答案区的时间阶段，进一步比较新颖答案区与寻常答案区注视时间之差。结果显示，在合适组解题的 2~2.5s 和 2.5~3s 时间段内对新颖答案区与寻常答案区的注视时间之差没有显著差异，$t(21)=1.44, p =0.17$。在新颖组解题的 1~1.5s、1.5~2s、2~2.5s 和 2.5~3s 时间段内，对新颖答案区与寻常答案区的注视时间之差存在显著差异，$F(3,63)=3.18, p < 0.05$，其中 1~1.5s 阶段的注视时间之差显著小于 2.5~3s 阶段，其余时间段之间的差异均没有达到显著水平。如图 3 所示。

2.4 讨论

实验 1 的结果表明，被试选择新颖答案的百分数
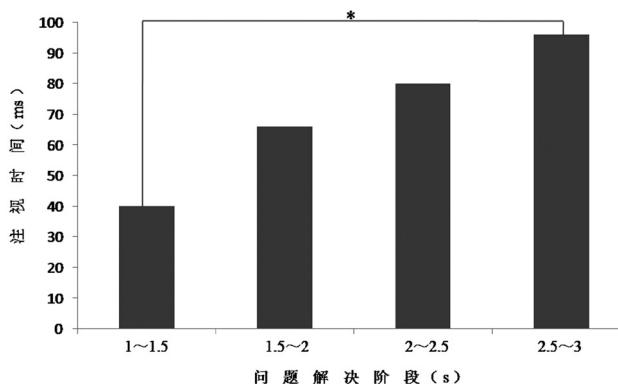
图 3 新颖组的新颖答案与寻常答案区注视时间之差

与其对新颖答案的注视时间呈高度正相关,而与其对寻常答案和无关答案的注视时间呈高度负相关。反过来,被试选择寻常答案的百分数与其对寻常答案的注视时间呈高度正相关,而与其对新颖答案和无关答案的注视时间呈高度负相关。这说明,被试对两类答案的注视(时间)与其对两类答案的选择高度一致。进而言之,被试注视新颖答案时,他所进行的主要是新异联想,被试注视寻常答案时,他所进行的主要是简单联想。

当然,被试注视无关答案时,并不代表他没有进行思考,他可能进行着各种尝试、探索性的思考。当他发现无关答案与谜面之间无法形成语义联结时,就会放弃该答案,转而注视其它答案。这种尝试建立备择答案与谜面之间语义联结的过程,是本实验中较具代表性的思考过程。所以,如果被试对新颖答案或寻常答案的注视时间长于对无关答案的注视时间,就说明新异联想或简单联想发生了,并处于比较稳定的发展过程之中。

在合适组选择新颖答案过程中,从 1s 开始新异联想发生了,在 2~3 s 阶段简单联想也发生了。在新颖组选择新颖答案过程中,从 1s 开始新异联想发生了,在 1~3 s 阶段简单联想也发生了。由此可知,在汉语成语谜语问题解决过程中的某些时间阶段,新异联想和简单联想能够同时发生。而新异联想与简单联想都是针对同一谜题进行的,并且被试往往更习惯进行简单联想,新异联想的顺利进行需要抑制简单联想的思维习惯,所以,两种联想之间可能存在着竞争关系。在新异联想和简单联想都发生的阶段,新颖答案区和寻常答案区的注视时间之差越小表示两种联

想之间的竞争程度越高,随着问题的解决,新异联想与简单联想之间的竞争程度逐渐减弱。在选择新颖答案过程中的思路竞争阶段,对新颖答案的注视时间一直长于寻常答案,表明新异联想占居主导地位。

注视时间的三次交互作用不显著,说明实验任务要求没有改变简单联想和新异联想的发生、发展进程,新颖组和合适组被试的新异联想都从 1 秒后开始占据主导地位,两组被试选择新颖答案的“相对速度”也是一致的,到 3441 ms(平均解题时间)时的累积百分数分别为 50% 和 51%。实验任务要求也没有改变两种解题思路可以同时发生且形成相互竞争的局面,但是影响了两种思路竞争的周期,合适组的思路竞争发生在 2~3 s 阶段,而新颖组的竞争发生在 1~3s 阶段。可能原因在于,新颖组根据指导语中的任务要求知道谜题既有新颖答案,也有寻常答案,能够更加有意识地主动抑制简单联想。

实验 1 中,在选择新颖答案过程中,短于平均选择时间三个标准差的两个时间阶段内 (1~1.5 s 和 1.5~2.0 s),对新颖答案区的注视时间显著长于无关答案区,可能反映了被试对新颖答案的直觉偏好。

为了进一步探索谜语问题解决过程中常规思路与新颖思路发生、发展的特点及其影响因素,实验 2 操纵了规则线索的有效性,重点考察有效和无效规则线索是否影响新颖思路的发生、发展过程,以及常规思路是否同时发生并与新颖思路形成竞争。在实验 1 中,被试有可能采用“排除法”策略来选择答案,为了尽量减少“排除法”策略的使用,提高考察新颖思路发生、发展过程的有效性,实验 2 增加了两个似是而非答案选项,它们与谜面之间有字面上的关联,但不存在严格的语义扣合关系。

3 实验 2

3.1 实验目的

采用眼动仪记录被试解决成语谜题的眼动轨迹,通过分析不同线索条件下的解题过程中的每一时间段内对新颖答案、寻常答案与无关答案的注视时间

表 5　实验 2 中使用的实验材料举例

谜面	总体规则	具体规则	寻常答案	新颖答案	似是而非答案	无关答案
降落伞	正扣	性质特点联想	从天而降	随机应变	落落大方	明察秋毫
五指	正扣	象形联想	长短不一	三长两短	首屈一指	以身作则
泣别	反扣	会意联想	依依不舍	不欢而散	曲终人散	门庭若市
都成眷属	反扣	结果联想	成双成对	无独有偶	独善其身	水泄不通

的差异,考察总体规则线索和具体规则线索如何影响新异联想和简单联想的发生、发展进程,以及两种联想的竞争过程。

3.2　实验方法

3.2.1　实验设计　实验采用 2(总体规则线索有效性:有效 vs 无效)× 2(具体规则线索有效性:有效 vs 无效)的两因素组内设计。为了确保被试信任并利用线索解题,两类规则线索有效的数量都占 2/3,无效的数量占 1/3,实验中为每一道谜题同时提供一条总体规则线索和一条具体规则线索。实验指导语要求被试选择新颖且合适的答案,并且提示被试不是所有与谜面之间存在语义关联的备择答案都符合新颖性要求。因变量为选择新颖答案的百分数,以及选择新颖答案过程中对各类备择答案的注视时间。

3.2.2　被试　22 名来自武汉某大学的本科生参加了本实验,其中男生 10 名,女生 12 名。所有被试的视力或矫正视力均正常,此前未参加过类似实验,实验结束后获赠一件小礼物。

3.2.3　实验仪器　实验仪器同实验一

3.2.4　实验材料　从汉语成语谜语库中抽取 90 道谜题,分别属于正扣和反扣两种总体规则,以及会意、结果、象形和性质特点等四种不同的具体规则。为每道谜题设置 6 个备择答案,其中包括一个新颖答案、一个寻常答案、两个似是而非答案和两个无关答案。新颖答案、寻常答案和无关答案的编制方法同实验 1,似是而非答案是在预实验中从被试根据谜面所生成的答案中选取生成概率非常低,与谜面中的某个字存在关联的成语。表 5 列举了实验 2 中使用的 4 道谜题及其备择答案。

3.2.5　实验程序　实验在隔音的眼动实验室内完成。实验开始前,被试的前额和下颚放在托架上以固定头部,眼睛距离显示器大约 75 cm。进行校准

和确认后,被试先进行 8 道题的练习,熟悉实验程序后进入正式实验。正式实验流程是:首先在屏幕中央呈现一个"十"注视点 500 ms,然后呈现谜面 3000 ms,之后在谜面下方同时呈现两条规则线索,其中具体规则线索在左侧,总体规则线索在右侧,2000 ms 之后再在线索下方呈现 6 个备择答案(6 个备择答案的位置进行了平衡),要求被试在 15s 内选择谜题的新颖答案。间隔 2000 ms 后,呈现下一道谜题。

正式实验中呈现备择答案的同时,眼动仪开始记录被试的眼动轨迹等数据,被试按键选择答案之后,眼动仪停止记录。每一道谜题呈现之前,都会进行一次漂移校正以保证眼动记录的准确性。

3.2.6　数据处理　将整个注视画面划分为 9 个兴趣区:谜面、总体规则线索、具体规则线索、新颖答案、寻常答案、似是而非答案 I、似是而非答案 II、无关答案 I、无关答案 II。通过 SR 公司提供的 Dataviewer 数据分析软件导出各个兴趣区的眼动数据,使用 SPSS 11.5 统计软件包进行统计分析。

3.3　结果与分析

3.3.1　选择新颖答案的百分数　表 6 所示的是被试在不同规则线索条件下选择新颖答案的百分数。

对新颖答案的选择百分数进行 2(总体规则线索有效性)× 2(具体规则线索有效性)的两因素方差分析,结果显示,总体规则线索有效性的主效应显著,$F(1,21) = 28.33$, $p < 0.001$,有效水平的选择比例更高;具体规则线索有效性的主效应不显著,$F(1,21) = 1.33$, $p = 0.26$;交互作用不显著,$F(1,21) = 0.84$, $p = 0.37$。

3.3.2　选择新颖答案过程中的不同时间段内的注视时间　表 7 所示的是被试在选择新颖答案过程中的不同时间段内对不同兴趣区的注视时间。

为了考察谜语问题的正确解决过程中新异联想

表 6　选择新颖答案的百分数

总体规则线索	具体规则线索	选择新颖答案的百分数
有效	有效	0.54（0.17）
	无效	0.50（0.14）
无效	有效	0.37（0.17）
	无效	0.36（0.15）

注：括号内数据为标准差。

和简单联想的发生、发展进程，依然采用分时间段数据分析方法，进行 2（总体规则线索有效性）× 2（具体规则线索有效性）× 3（兴趣区类型：新颖答案区，寻常答案区，无关答案区）× 8（时间段：0~1 s, 1~2 s, 2~3 s, 3~4 s, 4~5 s, 5~6 s, 6~7 s, 7~8 s）四因素组内设计的方差分析。重点分析总体规则线索有效性、具体规则线索有效性、兴趣区类型、时间段的四次交互

$9.16, p < 0.01$；总体规则线索有效性与时间的交互作用显著，$F(7,147) = 3.68, p < 0.005$；兴趣区类型与时间段的交互作用显著，$F(14,249) = 6.73, p < 0.001$；总体规则线索有效性、具体规则线索有效性与时间段的交互作用显著，$F(7,147) = 8.12, p < 0.001$。具体规则线索有效性、兴趣区类型与时间段的交互作用显著，$F(14,294) = 2.05, p < 0.05$。四因素的交互作用显著，$F(14,294) = 3.19, p < 0.001$。其他交互作用均不显著。

对于四因素的交互作用，重点分析总体规则线索有效性与具有规则线索有效性形成的四种组合条件下，不同时间段上，对三类答案兴趣区注视时间的差异。简单效应分析发现：(1) 当总体规则线索和具

表 7　选择新颖答案过程中的不同阶段内的注视时间 (ms)

总体规则线索	具体规则线索	兴趣区	0~1s	1~2s	2~3s	3~4s	4~5s	5~6s	6~7s	7~8s
有效	有效	新颖答案区	105	149	196	190	179	189	157	141
		寻常答案区	103	82	97	78	67	62	50	47
		无关答案区	69	75	61	64	57	50	50	42
	无效	新颖答案区	30	128	147	173	170	160	204	189
		寻常答案区	39	83	77	79	80	87	47	32
		无关答案区	77	66	84	64	57	49	44	37
无效	有效	新颖答案区	53	120	180	168	209	183	199	116
		寻常答案区	48	112	60	70	87	68	77	44
		无关答案区	90	85	55	56	54	55	41	37
	无效	新颖答案区	217	143	124	214	165	232	237	151
		寻常答案区	198	70	71	84	92	58	29	27
		无关答案区	38	46	61	81	80	50	59	30

作用，即考察四种规则线索有效性组合条件下，在不同时间段，对新颖答案区、寻常答案区和无关答案区的注视时间的差异。被试选择新颖答案的平均时间是 8300(± 1636) ms，所以实验 2 分析 8s 以内时间范围注视时间数据，理由同实验 1。

方差分析的结果表明，总体规则线索有效性的主效应边缘显著，$F(1,21) = 4.05, p=0.057$；具体规则线索有效性的主效应不显著，$F(1,21) = 1.32, p = 0.26$；兴趣区类型的主效应显著，$F(2,42) = 129.98, p < 0.001$，对新颖答案区的注视时间长于无关答案区 ($p < 0.001$)，也长于寻常答案区 ($p < 0.05$)；对寻常答案区的注视时间长于无关答案区 ($p < 0.001$)。时间段的主效应显著，$F(7,147) = 5.81, p < 0.001$；总体规则线索有效性与具体规则线索有效性的交互作用显著，$F(1,21) =$

体规则线索都有效时，各时间段上，对新颖答案区的注视时间显著长于无关答案区；在 0~1 s 和 2~3 s 时间段上，对寻常答案区的注视时间显著长于无关答案区；从 1~2 s 时间阶段开始，对新颖答案区的注视时间就显著长于寻常答案区。(2) 当总体规则线索有效但具体规则线索无效时，在 1~2 s 和 2~3 s 时间阶段对新颖答案区的注视时间长于无关答案区的注视时间，差异边缘显著，3 s 以后对新颖答案区的注视时间显著长于无关答案区；对寻常答案区和无关答案区的注视时间在各时间阶段都没有显著差异；从 2~3 s 时间阶段以后，对新颖答案区的注视时间显著长于寻常答案区。(3) 当总体规则线索无效但具体规则线索有效时，从 2~3 s 时间阶段以后，对新颖答案区的注视时间显著长于无关答案区；各时间阶段上，对

寻常答案区和无关答案区的注视时间都没有显著差异；从 2~3 s 时间阶段以后，对新颖答案区的注视时间显著长于寻常答案区。（4）当总体规则线索和具体规则线索都无效时，各时间阶段上，对新颖答案区的注视时间显著长于无关答案区；在 0~1s 阶段，对寻常答案区的注视时间显著长于无关答案区，1s 之后对寻常答案区和无关答案区的注视时间又没有显著差异；从 3~4s 时间阶段以后，对新颖答案区的注视时间显著长于寻常答案区。如图 4、图 5、图 6 和图 7 所示。

3.4 讨论

实验 2 结果表明，有效的总体规则线索提高了新颖答案选择的比例，促进了新颖语义联结的形成；当总体规则线索无效时，它明显阻碍了新颖语义联结的形成。具体规则线索的有效性对新颖答案的选择并没有显著的作用。可能是因为总体规则线索提供的总体联想方向的真伪难以区分，而具体规则线索提供
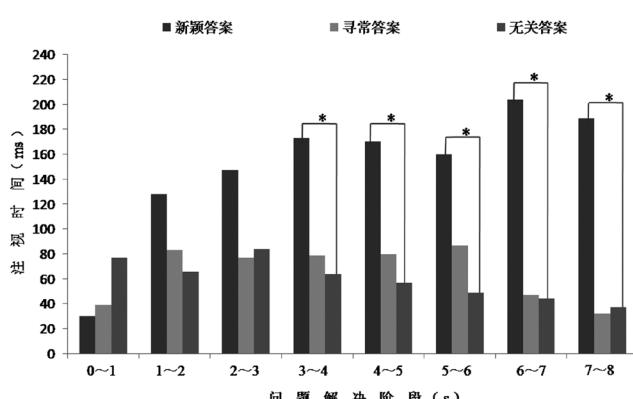
图 6　总体规则无效、具体规则有效

图 7　总体规则无效、具体规则无效

图 4　总规规则有效、具体规则有效
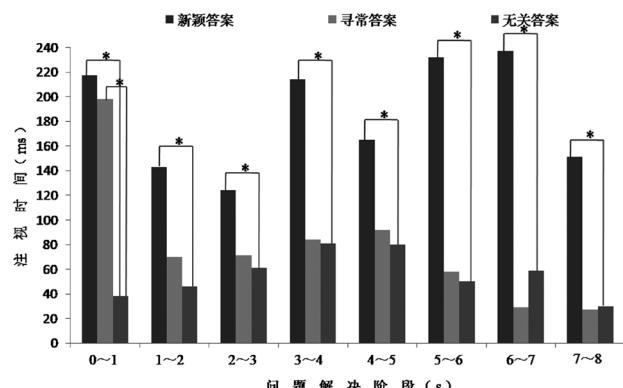
图 5　总体规则有效、具体规则无效

的联想的具体形式的真伪相对容易区分，被试能够批判性地利用具体规则线索，所以只发现了总体规则线索有效性对于新颖答案选择的显著影响。在两条规则线索都无效时，他们选择新颖答案的百分数 (36%) 显著高于随机选择的概率 (16.7%)。这也能够说明，被试并没有完全受线索的约束，他们能够主动调整思路，寻找新颖答案，建立谜面与答案之间的新颖语义联结。

在选择新颖答案过程中，无论两条规则线索是否有效，一般在 3 秒后，被试对新颖答案区的注视时间就显著长于无关答案区，而被试对寻常答案区与无关答案区的注视时间都没有显著差异，这说明新异联想发生了并占据主导地位，而稳定的简单联想基本上没有发生。可能的原因是，在提供了解题线索并要求被试选择新颖且合适答案的条件下，更加激发了被试寻找新颖答案的动机，他们利用两种规则线索努力进行新异联想，同时抑制简单联想。所以，在正确选择新颖答案的解题过程中，没有出现稳定的新异联想与简

单联想同时发生且相互竞争的过程。

规则线索有效性的作用也反映在新异联想的发生、发展进程上。到 8300 ms（平均解题时间）时，在两类规则线索所形成的四种有效性条件下，选择新颖答案的累积百分比分别为 54%、47%、46% 和 48%。在两条规则线索都有效性时，选择新颖答案的"相对速度"快于其它三种实验条件。两类规则线索的有效性对新异联想发生、发展的作用不是独立的，因为正确的新异联想不仅需要确定正向思考或反向思考的总体方向，还需要确定具体的会意、性质特点、结果或象形联想等具体方向，任何一条线索无效，都会对恰当的新异联想产生误导，所以两类规则线索同时有效才能促进新异联想发生、发展的进程。

实验 2 中，在选择新颖答案过程中短于平均选择时间三个标准差的两个时间阶段内 (1~2s 和 2~3s)，对新颖答案区的注视时间显著长于无关答案区，反映了被试对新颖答案的偏好。这种对新颖答案的偏好并不受规则线索有效性的影响，因此可以进一步认为，在这两个时间阶段对于新颖答案更长时间的注视属于直觉加工过程。

4 总讨论

4.1 谜语问题解决中新异联想与简单联想竞争的过程

联结主义认为创造就是把头脑中的观点按照不寻常的、新颖独特且有用的方式加以组合，从而形成一种新颖的联结的过程，或者说通过远距离联想形成信息间新颖联结。但是，由于"思维惰性"，人们往往习惯于进行一些简单的联想。本研究提供了包括寻常答案与新颖答案在内的 4 个（或 6 个）备择答案，在尝试发现并建立谜面与各个答案之间可能的语义联结过程中，通过语义激活扩散就可以自动建立谜面和寻常答案之间的寻常语义联结，这是简单联想；而打破常规思维方式，发现并建立谜面和新颖答案之间的新颖语义联结，这是新异联想。在谜语问题解决中，

被试需要努力尝试解决问题的新颖思路，发现并建立谜面与谜底之间的新颖语义联结。

显然，通过新异联想所形成的新颖且有效的解题思路战胜常规的解题思路是顿悟产生的关键。表征转换理论、进程监控理论以及后续的研究都没有直接考察两种思路冲突的过程，本研究采用眼动技术发现汉语成语谜语问题解决中的一段时间内，新异联想与简单联想可以同时发生并形成竞争（实验 1）。出现这种结果的可能原因是，在选择答案的解题过程中，被试首先需要发现和建立谜面与各个答案之间可能的语义关联，并评估各答案的适切性，在此基础上进一步比较与评估它们的新颖性。在发现并建立谜面与各个答案之间可能的语义联结过程中，虽然简单联想与新异联想都可能满足适切性要求，但是只有新异联想可能满足新颖性要求，而新异联想的发生与发展需要打破简单联想的限制。所以，在选择新颖答案的过程中，简单联想与新异联想可以在一段时间内同时发生，形成了常规思路与新颖思路之间的竞争，直到简单联想得到有效的抑制，两种思路之间的竞争才结束。

新异联想与简单联想之间相互竞争的过程不受实验任务要求的影响。"选择新颖且合适答案"的任务要求虽然提高了被试进行新异联想的认知努力程度，但是并没有改变两种思路相互竞争的过程。可能的原因是，选择"新颖且合适答案"的任务要求只是对答案提出了新颖性限制，并没有指明新异联想的方向，被试在解题过程中依然需要先尝试形成多种可能的语义联结，然后再比较不同联结之间的新颖性。

然而，在提供解题规则线索并要求被试选择"新颖且合适答案"的实验条件下，被试一般会根据总体规则所指示的总体方向进行正向或逆向思考，同时根据具体规则线索和谜面信息，展开某种具体形式的联想。由于两类规则线索提供了进行新异联想的方向，被试只需要沿着新颖的方向寻找具备适切性的答案，简单联想从一开始就受到了抑制，所以没有出现两种思路之间相互竞争的过程。

4.2 谜语问题解决中新异联想发生发展的进程

在提供备择答案的实验条件下，通过语义激活

扩散可以自动建立谜面与寻常答案之间的寻常语义联结，所以，简单联想很快就能发生。然而，发现并建立谜面与新颖答案之间新颖语义联结，则需要抑制谜面或谜底关键字（词）常见意义，而激活其非寻常意义，所以新异联想发生、发展的过程应该较晚。然而，本研究中两个眼动实验的数据都表明，在各种实验条件下选择新颖答案过程中，在问题解决的早期阶段都存在对新颖答案的直觉偏好，反映了对新颖语义关联的无意识加工。国外研究中有类似的结果，例如，Ellis, Glaholt 和 Reingold (2011) 在英语字谜问题解决的眼动研究中发现，答案相关的知识在问题解决早期就已经被无意识地激活了。Bowden 和 Jung-Beeman 也发现，顿悟问题的正确答案在被试有意识地提取之前是处于无意识激活状态的 (Bowden & Jung-Beeman, 1998; Jung-Beeman & Bowden, 2000)。

在被试最终选择新颖答案的情况下，对新颖答案的注视时间（各时间阶段）都显著长于对无关答案的注视时间，这说明，本研究中新颖语义联结的形成是一个逐渐积累的过程。Yaniv 和 Meyer (1986) 也认为，在语义信息丰富的顿悟问题的解决过程中，与答案相关的语义信息是逐渐积累的。本研究的结果比较符合进程监控理论的预期。

本研究还发现，选择"新颖且合适答案"的任务要求增强了问题解决者进行新异联想的认知努力，提高了成功建立新颖语义联结的概率，但是并没有加快建立新颖语义联结的"速度"。可能的原因是，选择"新颖且合适答案"的任务要求虽然对谜面与答案之间的语义联结有了"新颖性"限制，但是，实验任务要求并没有说明如何进行新异联想，也没有说明朝哪个方向进行新异联想，所以实验任务要求没有加快新异联想发生、发展的进程。

在要求选择"新颖且合适答案"的条件下，提供两类规则线索，加快了新颖语义联结形成的"速度"。因为总体规则线索能够引导被试进行正向或逆向思考，具体规则线索能够引导被试根据谜面信息进行象形、会意、因果等具体方式的联想。遵循有效规则线索的引导，被试能够更快地进行恰当的新异联想，所

以有效的规则线索加快了新异联想的发生、发展进程。

4.3 本研究的创新与不足

实验提供新颖答案、寻常答案、似是而非答案和无关答案等几类备择答案，利用眼动技术记录了被试在解题过程中对于新颖答案和寻常答案的注视时间，实验发现对新颖答案的总注视时间与选择新颖答案比例具有高度的正相关，说明对新颖答案注视时间的长短可以作为新异联想的敏感指标。本研究进一步将谜语问题解决时间分为若干阶段，通过比较不同时间阶段内对新颖答案注视时间、对寻常答案的注视与对无关答案的注视时间，直接考察了谜语问题解决过程中新颖有效的解题思路发生、发展的过程，以及常规的解题思路与新颖的解题思路之间冲突过程。本研究所发现的顿悟问题解决过程具有逐渐积累的性质，对于揭示语义信息丰富的顿悟问题解决过程的特性，具有一定的意义。

虽然采用选择答案的实验任务可以较好地操纵（或引导）被试进行新异联想与简单联想，也能有效地控制顿悟问题解决的时间，但是，毕竟产生答案才能最好地体现顿悟问题解决过程。

本研究对被试解决几十道谜题的眼动轨迹，按不同解题时间阶段分别进行平均统计分析，这有利于从整个解题过程上描述新异联想与简单联想发生、发展的特点，但是，采用这种计算方法，难以说明某道谜题的解决过程中是否存在新异联想瞬间替代简单联想的过程。

5 结论

根据两个实验的结果，可以得出如下三个结论：(1) 在谜语问题解决过程中的一段时间内新异联想和简单联想能够同时发生并形成竞争。(2) 选择"新颖且合适答案"的任务要求，提高了成功形成新颖语义联结的概率，但是并没有加快新异联想发生、发展的进程，也没有改变两种思路相互竞争的局面。(3) 有效的规则线索可以抑制简单联想，阻止其发生，同时可以加快新异联想发生、发展的进程。

参考文献

Aziz–Zadeh, L., Kaplan, J. T., & Iacoboni, M. (2009). "Aha!": The neural correlates of verbal insight solutions. *Human Brain Mapping*, 30(3), 908–916.

Bilalic, M., McLeod, P., & Gobet, F. (2008). Why good thoughts block better ones: The mechanism of the pernicious Einstellung (set) effect. *Cognition, 108*(3), 652–661.

Bowden, E. M., & Jung–Beeman, M. (1998). Getting the right idea: Semantic activation in the right hemisphere may help solve insight problems. *Psychological Science, 9*(6), 435–440.

Chronicle, E. P., MacGregor, J. N., & Ormerod, T. C. (2004). What makes an insight problem? The roles of heuristics, goal conception, and solution recoding in knowledge–lean problems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 30(1), 14–27.

Ellis, J. J., Glaholt, M. G., & Reingold, E. M. (2011). Eye movements reveal solution knowledge prior to insight. *Consciousness and Cognition*, 20(3), 768–776.

Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, 9(5), 462–466.

Jones, G. (2003). Testing two cognitive theories of insight. Journal of Experimental Psychology: Learning, *Memory, and Cognition*, 29(5), 1017–1027.

Jung–Beeman, M., & Bowden, E. M. (2000). The right hemisphere maintains solution–related activation for yet–to–be–solved problems. *Memory & Cognition*, 28(7), 1231.

Kaller, C. P., Rahm, B., Bolkenius, K., & Unterrainer, J. M. (2009). Eye movements and visuospatial problem solving: Identifying separable phases of complex cognition. *Psychophysiology*, 46(4), 818–830.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology*, 22, 374–419.

Knoblich, G., Ohlsson, S., & Raney, G. (2001). An eye movement study of insight problem solving. *Memory & Cognition*, 29(7), 1000.

Knoblich, G., Ohlsson, S., Haider, H., & Rhenius, D. (1999). Constraint relaxation and chunk decomposition in insight problem solving. Journal of Experimental Psychology: Learning, *Memory, and Cognition*, 25(6), 1534–1555.

Kounios, J., Frymiare, J. L., Bowden, E. M., Fleck, J. I., Subramaniam, K., Parrish, T. B., & Jung–Beeman M. (2006). The prepared mind: Neural activity prior to problem presentation predicts subsequent solution by sudden insight. *Psychological Science*, 17, 882–890.

Luo, J. (2004). Neural correlates of insight. *Acta Psychologica Sinica*, 36(2), 219–234.

[罗劲 . (2004). 顿悟的大脑机制 . *心理学报* , 36(2), 219–234.]

MacGregor, James N., Ormerod, T. C., & Chronicle, E. P. (2001). Information processing and insight: A process model of performance on the Nine–Dot and related problems. Journal of Experimental Psychology: Learning, *Memory, and Cognition*, 27(1), 176–201.

Mai, X. Q., Luo, J., & Wu, J. H. (2005). "Aha!" effects in a guessing riddle task: An event–related potential study. *Acta Psychologica Sinica*, 37(1), 19–25

[买晓琴 , 罗劲 , 吴建辉 , 罗跃嘉 . (2005). 猜谜作业中顿悟的 ERP 效应 . *心理学报* , 37(1), 19–25.]

Metcalfe, J. (1986). Feeling of knowing in memory and problem solving. Journal of Experimental Psychology: Learning, *Memory, and Cognition*, 12(2), 288–294.

Metcalfe, J., & Wiebe, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition*, 15(3), 238.

öllingera, M., Jones, G., & Knoblich, G. (2008). Investigating the effect of mental set on insight problem

solving. *Experimental Psychology*, *55*(4), 269–282.

Patsenkoa, E. G., & Altmanna, E. M. (2010). How planful is routine behavior? A selective–attention model of performance in the Tower of Hanoi. *Journal of Experimental Psychology: General*, *139*(1), 95–116.

Qiu, J., Luo, Y. J., Wu, Z. Z., & Zhang, Q. L (2006). A further study of the ERP effects of 'insight' in a riddle guessing task. *Acta Psychologica Sinica*, *38*(4), 507–514.

[邱江 , 罗跃嘉 , 吴真真 , 张庆林 . (2006). 再探猜谜作业中"顿悟"的 ERP 效应 . *心理学报* , *38*(4), 507–514.]

Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, *122*(2), 166–183.

Shen, W. B., Liu, C., Zhang, X. J., & Chen, Y. L. (2010). The time course and hemispheric effect of "insight" in three–character Chinese riddles task: An ERP study. *Acta Psychologica Sinica*, *43*(3), 229–240.

[沈汪兵 , 刘昌 , 张小将 , 陈亚林 . (2011). 三字字谜顿悟的时间进程和半球效应:一项 ERP 研究 . *心理学报* , *43*(3), 229–240.]

Thomas, L. E., & Lleras, A. (2007). Moving eyes and moving thought: On the spatial compatibility between eye movements and cognition. *Psychonomic Bulletin and Review*, *14*(4), 663–668.

Thomas, L. E., & Llerasb, A. (2009). Covert shifts of attention function as an implicit aid to insight. *Cognition*, *111*(2), 168–174.

Wo, J. Z., Li, Q., & Tian, H. J. (2006). The eye movements study on children with different ability levels in the figure reasoning problem. *Developmental and Educational Psychology*, *3*, 6–10.

[沃建中 , 李琪 , 田宏杰 . (2006). 不同推理水平儿童在图形推理任务中的眼动研究 . *心理发展与教育* , *3*, 6–10.]

Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. Journal of Experimental Psychology: Learning, *Memory, and Cognition*, *13*(2), 187–205.

Zhu, X. C., Li, R. J., & Zhou, Z. J. (2009). The role of clues in Chinese idiom riddle solving. *Acta Psychologica Sinica*, *41*(5), 397–405.

[朱新秤 , 李瑞菊 , 周治金 . (2009). 谜语问题解决中线索的作用 . *心理学报* , *41*(5), 397–405.]

An eye movement study of associate competition in Chinese idiom riddles solving

HUANG Furong[1,2]; ZHOU Zhijin[1]; ZHAO Qingbai[1]

([1]Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education;

Key Laboratory of Human Development and Mental Health of Hubei Province;

School of Psychology, Central China Normal University, Wuhan 430079, China)

([2] Beijing Key Lab of Learning and Cognition, Department of Psychology; Capital Normal University, Beijing 100048, China)

Abstract   Most Chinese idiom riddles require insightful thoughts to solve. Novel and simple associations can be formed during the process of idiom resolutions. Insightful thought occurs only when novel associations overwhelm simple ones. However, it is unclear how this happens in the mind. According to the Representational Change Theory, the competition happens in a sudden way. But according to the Process Monitoring Theory, it is completed gradually. By using eye–tracking technology, we intended to investigate the time course of insightful problem resolution, using Chinese idiom riddles as experimental materials. In this study, an option selection task was adopted. Chinese idiom riddles were presented, together with four types of options serving as the spare answers to the riddle (novel, ordinary, plausible and absolutely wrong). Participants were asked to make a choice among options. At the same time, the fixation times of the participants spent on different options were recorded. In Experiment 1, the participants were randomly divided into an appropriate group and a novel group. In different groups, the participants were asked to make a choice between an appropriate answer and a novel and appropriate one. In Experiment 2, a 2 (general solution rule: effective vs. ineffective) × 2 (special solution rule: effective vs. ineffective) experimental designs was adopted. Besides, a general solving rule and a specific solving rule were presented at the same time. Results showed that there was a positive correlation between the fixation times of the participants spent on novel or ordinary answers and the percentage of the corresponding selections. The participants were found to have formed novel associations while fixating on novel answers, and simple associations while fixating on ordinary answers. The result further revealed that: (1) Novel associations and simple associations were formed simultaneously and competed to each other for a while before the idiom riddles were solved. (2) The demand of choosing a novel and appropriate answer induced people to make more efforts on novel association formation, and to select more novel answers. However, the task demand did not accelerate the time course of novel association formation, or change the competitive situation. (3) The effective solution rules, which promoted the novel answer selections, not only accelerated the time course of novel association formation, but also inhibited simple association formation, and eliminated the competition between them.

Keywords   Chinese idiom riddle; novel association; simple association; eye movement

语言标签和自我关联对新颖客体类别知觉的影响 *

刘思耘，孟健欣

（青少年网络心理与行为教育部重点实验室；华中师范大学心理学院，武汉 430079）

摘　要　类别知觉 (Categorical perception) 是人类最基本的认知活动之一，探讨语言对类别知觉的影响是心理语言学领域的热门话题之一。在这个研究的 3 个实验中，分别在高、低不同的语言标签表征强度下、高、低不同的客体自我关联程度下观察新颖客体类别知觉的过程。研究发现，语言标签表征程度的增强可促进新颖客体类别知觉的右视野优势效应；客体与自我关联程度的提高会促进左、右视野的类别知觉效应；语言标签的表征程度和客体与自我关联程度同时增强时，语言标签的作用依旧表现出来，但与自我关联的影响产生权衡分配，且其影响力并不足以产生右视野优势效应。

关键词　类别知觉；语言标签；自我关联；右视野优势效应

分类号　B842

1　前言

迄今为止，研究者对语言与思维关系的争论逐步形成两派，一派持语言普遍论 (linguistic universalism) 观点，认为尽管不同语言在表层结构上具有多样性，但其深层内容是一样的，与人类的认知加工并无直接联系，不会造成不同语言不同文化下人们不同的认知行为 (Malt & Wolff, 2010); 另一派则持语言相对论 (linguistic relativism) 观点 (Sapir, 1921; Whorf, 1956)，认为不同语言代表了不同的世界观，语言间的巨大差异会导致人们体验及思维上的差异。持这两种观点的研究者用不同的方法在不同范围内分别找到了大量支持性证据，其中对类别知觉的探讨是最主要的一部分。

类别知觉是指人们对于连续变化的客体刺激感知到的是非连续的、属于不同类别客体的一种现象，如连续变化的可见光波在人类感知的方式下形成了界限分明的七种色彩 (McCullough & Emmorey, 2009)。在类别知觉过程中，类别间客体的差异得到强化，而类别内客体的差异则被弱化，由此人们对类别间客体的感知要优于对类别内客体的感知，即类别知觉效应 (Goldstone & Hendrickson, 2010)。人类对很多客体的感知过程都存在类别知觉效应，如颜色（如 Gilbert, Regier, Kay, & Ivry, 2006)、线条 (Franklin, Catherwood, Alvarez, & Axelsson, 2010)、音符和弦 (Klein & Zatorre, 2011)、语音 (Peng et al., 2010)、情绪面孔（如 Fugate, 2013) 等。

语言作为人类认知系统中特有的结构复杂的心理表征系统，不仅是人际交流的重要媒介和工具，更是整个认知系统操纵各种心理表征并协调各个层次

认知加工的工具 (Gentner & Goldin-Meadow, 2003)。探讨语言对类别知觉过程的影响有助于人们更好地了解语言在整个认知加工过程中，特别是在初级认知加工过程中，所起到的重要作用。

1.1 语言与类别知觉的关系

前人在探讨语言与类别知觉关系时，发现了截然不同的两种结果：一种是语言对类别知觉没有影响；而另一种则相反，认为语言对类别知觉有显著的影响。这两种结果分别得到了不同实证研究的支持。

1.1.1 支持语言普遍论，认为语言对类别知觉无影响的研究

早期研究者在幼儿及前语言阶段婴儿群体中就发现了类别知觉效应，认为如果未习得语言的婴幼儿都可以表现出类别知觉效应，那么该效应很有可能是不受语言影响的 (Franklin & Davies, 2004; Franklin, Pilling, & Davies, 2005; Franklin, Clifford, Williamson, & Davies, 2005a, Franklin et al., 2005; Franklin, Drivonikou, Bevis, Davies, Kay, & Regier, 2008; Ozturk, Shayan, Liszkowski, Majid, 2013)。在成人被试中也有研究者发现，语言标签并未对类别知觉起到促进作用 (Holmes & Wolff, 2012)。如 Sauter 等人比较了德语母语者和玛雅人母语者对愤怒和恶心情绪面孔的类别知觉反应，其中玛雅语中并没有两个不同的词汇分别对应恶心和愤怒两种情绪，结果发现这两种被试对上述两种情绪及其它情绪均可做到类别知觉 (Sauter, LeGuen, & Haun, 2011)。

分析这类研究的一个共性，发现大部分由于被试 (为婴幼儿) 的特殊性，研究中的数据采集也不同于很多其它相同主题的研究。如为了针对婴儿被试而采用习惯化范式 (Franklin & Davies, 2004)，或者采用眼动技术 (Franklin et al., 2005; Ozturk et al., 2013)，分析婴儿对同类或异类刺激注视时长是否有显著差异。进一步分析这两种研究范式所涉及的被试任务及所采集的数据类型，可以发现，无论是对新颖刺激做出新奇的反应，还是注视点停留在差异性更大的刺激上多一些时间，这些反应类型更多体现的是一种初期自动化知觉加工的结果。而早期知觉加工不受到语言因素的影响并不表示输入刺激在后期认知加工中不受语言因素的影响。当被试在测试中被要求完成的任务类型涉及更为复杂的后期认知加工时，那么很有可能会出现不一样的结果。

1.1.2 支持语言相对论，认为语言对类别知觉有影响的研究

尽管上述研究发现语言与类别知觉是相对独立的，但更多的研究发现语言和类别知觉有着密不可分的关系。具体来说，是语言中的语义类别扩大了客体类别间的差异而缩小了类别内的差距，导致在一定程度上影响了人们对客体的知觉 (Gumperz & Levinson, 1996)。研究者认为在类别知觉过程中，语言的标签作用使得类别知觉效应更加显著 (如 Gilbert et al., 2006; Roberson, Damjanovic, & Pilling, 2007; Roberson, Pak, & Hanley, 2008)。

首先，前人的研究发现，如果某种语言用不同的符号标记出两个客体属于不同的类别，则这种语言标签的作用似乎能够促进个体更好地对类别间和类别内客体进行再认和知觉判断 (如：Roberson, Davies, & Davidoff, 2000; Roberson, Davidoff, Davies, & Shapiro, 2005; Lupyan, 2012; Lupyan, Rakison, & McClelland, 2007; Peng et al., 2010; Zhou et al., 2010), 甚至在记忆上造成差异 (Roberson & Davidoff, 2000)。这些研究通过对比不同母语者在类别知觉过程中的不同表现，推断出语言标签在类别知觉中起到的重要作用，相关证据来自于多个语言的母语者对相同材料进行类别知觉的不同表现 (Roberson, Davies, & Davidoff, 2000; Roberson et al., 2005; Goldstein, Davidoff, & Roberson, 2009; Winawer et al., 2007; Roberson et al., 2008; 张积家，刘丽红，陈曦，和秀梅，2008; 谢书书，张积家，和秀梅，林娜，肖二平，2008; McCullough & Emmorey, 2009)。

研究者还利用视觉搜索范式发现了客体类别知觉的右视野优势效应，该效应是语言对类别知觉影响的最直接体现。研究者认为由于语言加工的主要区域在左半球，因此如果语言对类别知觉过程产生影响，那么在左 - 右视野靶子刺激分别呈现的视觉搜索

过程中，右视野 – 左脑这个通道会由于得到语言的促进作用，使得被试能更快地对类别间或类别内的客体做出判断，即产生右视野优势效应 (Gilbert et al., 2006; Gilbert, Regier, Kay, & Ivry, 2008; Paluy, Gilbert, Baldo, Dronkers, & Ivry, 2011)。而右视野优势的产生随着语言在人类婴幼儿时期的发展也有一个逐步变化的过程，即婴幼儿在习得语言之前对类别知觉呈现出右脑优势效应 (Franklin, Drivonikou, et al., 2008)，而随着语言的逐步习得，其类别知觉加工的优势才逐步从右脑向语言加工所在区的左脑转移 (Franklin, Drivonikou, et al., 2008)。

　　语言标签对类别知觉的影响不仅得到行为学研究结果的支持，还得到大量来自认知神经机制研究的支持。研究者们通过采集脑电 (EEG 或 ERP) 和脑成像 (MRI 或 fMRI) 的数据进一步验证了语言对类别知觉过程的影响。如 ERP 脑电研究发现，负责语言加工的左半脑在刺激左 / 右视野呈现时能够产生更大波幅的 N2pc 成分，表明左半脑给予了刺激更多的注意分配，刺激得到更强劲的加工 (Liu, Li, et al., 2009; 姚树霞，杨东，齐森青，雷燕，Ding, 2012)。还有研究发现，在语义上有差别 (即类别间) 的颜色会诱发视觉失匹配负波 (vMMN)，而无差异的则不会 (Thierry, Athanasopoulos, Wiggett, Dering, & Kuipers, 2009; Mo, Xu, Kay, & Tan, 2011)。

1.2　自我关联对客体类别知觉的影响

　　作为高度社会化的人类来说，其思维的方式不仅仅只受到语言的影响，还受到其它诸多来自社会和人格特质方面因素的影响，如客体与自我关联的程度被发现有可能是影响类别知觉的重要影响因素。如 Lupyan 等人 (2007) 让被试在与客体建立高自我关联的情况下，分别在有或无语言标签的条件下对新颖客体进行分类学习，发现学习了语言标签的被试表现出了更好的分类能力。而 Holmes 和 Wolff (2012) 在研究中仅仅是要求被试对毫无意义的新颖图片进行分类，新颖的客体与被试自我之间毫无关联，被试在学习或未学习新颖客体语言标签的条件下完成视觉搜索任务，结果并未发现两组被试的类别知觉效应有显著差异。由此可推断出，这两个研究结果的差异很可能是由于一个研究操纵了客体与自我的关联，而另一个则没有。自我在类别知觉过程中有可能起到重要影响作用。但 Lupyan 等人 (2007) 的研究并没有系统地操纵自我关联与新颖客体之间的关系，原因在于作者在该研究中的主要目的是探讨冗余的语言标签是否会促进新颖客体分类的学习，而且作者让被试与新颖客体建立高自我关联是促使被试进行类别学习的一种手段，自我关联并不是作者探讨的重点。因此，自我关联与新颖客体类别之间的关系本质还不是很清晰。

　　早期研究 (Klein & Loftus, 1988; Klein, Rozendal, & Cosmides, 2002; Rogers, Kuiper, & Kirker, 1977) 发现，自我的意识一旦参与到个体的信息加工过程，其加工的模式和效率都会有所不同，即出现自我参照效应 (self–reference effect)，如记忆效果会明显得到提升。后期多个研究为自我参照效应提供了支持性证据 (Kesebir & Oishi, 2010; Kim & Johnson, 2012; Sui, He, & Humphreys, 2012; Turk et al., 2011)。研究者认为这是因为与自我关联程度高的加工伴随有独特的动机和情感意义 (袁翠平，卢光明，2010)，或伴随出现自我意识 (Ma & Han, 2010)，从而促进个体对这些信息的反应。

　　研究者们还从神经机制的角度去探讨自我的加工脑区，发现了相关的主要大脑结构及其分布状况。其中有研究发现，关于自我的神经机制是广泛分布于左、右脑区的 (Northoff et al., 2006; 程蕾，陈熙海，黄希庭，2011; 杨帅，黄希庭，傅于玲，2012)，个体所表征的关于自我的大部分信息都是全脑表达的 (Powell, Macrae, Cloutier, Metcalfe, & Mitchell, 2010)，这一点与语言加工的脑区分布是不同的。

1.3　问题提出及研究假设

　　在前人的相关研究中，越来越多的证据表明语言标签对客体类别知觉的影响是显著的、稳定的，但同时也有研究对这个结论不断提出了质疑，并提供了确切的数据进行反驳。如果说人类在习得语言之前其类别知觉加工系统已经形成，当习得语言后，语言逐步对这个过程产生了影响，而且这个影响最终发生在早期无意识阶段 (Mo et al., 2011)，即达到高

度自动化程度，那么在语言标签形成的初期，语言对类别知觉的影响应该有一个从控制性加工逐步转变成自动加工的过程，其影响力也是由小变大的，但对于这个过程似乎前人的研究中还未涉及到。因此，现在还不清楚语言标签的表征在形成的初期是否会对类别知觉产生影响或产生何种程度的影响，这个影响过程是否是一个随着表征程度逐步稳定而趋于显著的过程。尽管前人采用不同的技术手段和实验范式，从多个角度为语言相对论提供了大量支持性证据，但毕竟还是有些证据显示语言在类别知觉过程中的影响力是缺失的，由此推断语言在整个认知加工的过程中应该呈现一种动态的影响模式，而不是固定不变的。

在前人研究的基础之上，当前研究从动态的角度探讨语言标签的不同表征程度是否会对类别知觉有不同的影响；同时，根据前文的分析，自我很可能也是影响类别知觉的因素之一，因此将语言与自我这两个因素放在一起进行观察，可以进一步揭示语言与其它因素综合影响类别知觉过程的机制。这样做的目的是，首先，前人在探讨类别知觉过程中很少将语言与其它影响因素综合起来进行观察，而在现实情况下，任何的认知加工都掺杂了复杂的多个影响因素；其次，如果类别知觉会受到来自高级认知层面的语言的影响，那么很有可能也会受到其它抽象概念的影响，那么在与其它因素产生交互的过程中，语言的影响力又是如何表现的呢？该结果将有助于人们更好地理解语言与类别知觉的关系本质。

鉴于此，当前研究将围绕个体在新颖客体类别学习过程中与客体、语言标签产生的不同关系提出以下3个假设：

第一，语言标签与客体之间不同的表征程度对类别知觉过程有影响。

第二，语言标签表征程度较低情况下，客体与自我关联程度高的对类别知觉的影响更大。

第三，语言标签表征程度较高情况下，语言标签和自我关联两个因素对客体类别知觉均有影响，但语言标签的影响作用更大。由于前人大量研究显示语言

标签对类别知觉是有影响的，而且采用的语言标签都是表征程度高且相对稳定的，如熟悉的颜色的名字，熟悉客体的名字等；而仅仅只有 Lupyan 等人 (2007) 及 Holmes 和 Wolff (2012) 两个研究的结果发现了自我对类别知觉的影响。因此，从前人研究的范围和深度来预测，语言的影响力有可能更大、更稳定，这也是在假设中预测语言标签的影响力更大的主要原因。

2 实验前测：研究材料的测评及筛选

2.1 前测目的

由于整个研究均采用连续变化的新颖客体，但前人在探讨连续变化的客体类别知觉时仅采用了颜色，因此为了确保实验所选的材料刺激在类别学习之后会产生典型的类别知觉效应，所以首先对所制作的材料进行了前测。同时，在这个前测过程中，由于被试所学习的客体未被赋予任何语言标签或其它社会意义，因此对于被试来说，其分类判断的标准应该主要是客体本身所展示的不同的特征，即形状上的差异。也就是说，被试必须要能够根据一系列客体在形状上的差异就可以把它们区分成两种类别，产生明显的类别知觉效应。而在后面正式实验中，这一系列客体被赋予了某种语言标签，具备了某种与人有关的社会意义。在这种情况下，被试进行类别判断的标准除了形状上的差异之外，还有语言标签所带来的差异。

2.2 前测方法

2.2.1 被试

招募大学生被试 10 名，均为右利手，母语为汉语，且视力或矫正视力正常。每位被试完成实验后均可获得精美礼品一份。

2.2.2 实验材料

原始图片下载自卡内基梅隆大学相关网址 (www.tarrlab.org) 中的新异物体刺激 Greeble Generator。原始图片材料原本是红色，经过灰白处理之后，再使用专门软件 (Abrosoft FantaMorph, www.fantamorph.com) 制作成两个新颖客体之间连续等距渐变的 12 张图片 (Fugate, Gouzoules, & Barrett, 2010)，相邻图片无论是类
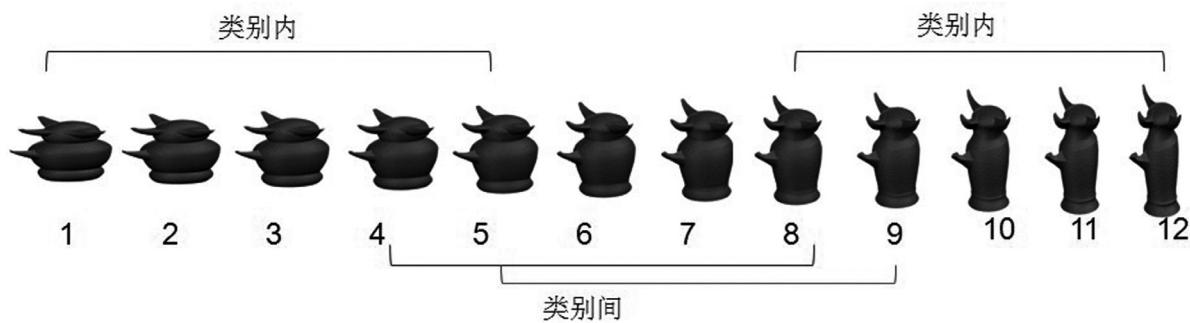
图 1　新颖连续渐变剪影图形

别内还是类别间的差异都是相等的 (如图 1)。这 12 张新颖的图片是使用软件 (AbrosoftFantaMorph, www.fantamorph.com) 在两个原始图片的基础上制作成连续等距渐变的 12 张图片。其中 1 至 6 的新颖图片更多是基于同一个客体，因此相互间为类别内关系；而 7 至 12 则更多基于另一个客体，因此相互间也为类别内关系。但 1~6 中的任意一张图片和 7~12 中的任意一张图片则为类别间的关系。这 12 张图片两两间类别内、类别间关系是由它们的原始图片决定的。

2.2.3　实验设计

采用 (客体类别：类别间 vs. 类别内) 单因素设计，因变量为被试判断刺激所在视野的反应时。

2.2.4　实验程序

采用 E-Prime 2.0 软件对实验设计进行编程，整个实验流程在电脑上进行。被试端坐于实验室中，

片。熟悉过后实验开始。12 张图片将单个随机出现，被试快而准确的按键判断所出现的图片属于哪一类，并按左、右键进行反应。每张图片呈现前有 1 s 的 "+" 注视点呈现时间，随后图片呈现 3 s。被试在 3 s 内做出反应的话，则程序自动跳转到下一个试次。每次按键后屏幕上将会出现反馈 (300 ms)。如果被试反应错误或未及时反应，则随后原图会再呈现一次 (700 ms)，被试熟悉过刚才看到的图片后才进入下一个试次。每两个试次间的空屏时间为 500 ms。每个实验单元包含 12 个试次，连续渐变的每个客体图片均只出现一次。直到某个实验单元中的正确率达到 100%，屏幕上才显示类别学习阶段结束，被试停止学习进入测试阶段。类别学习阶段的实验流程图见图 2。

在测试阶段，被试完成类别判断的经典视觉搜
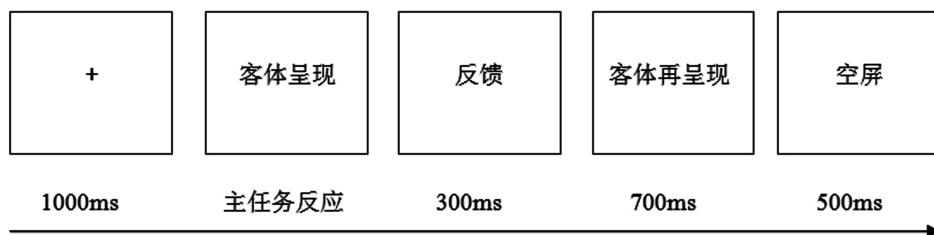
图 2　类别学习阶段实验流程图

与电脑屏幕保持合适距离，屏幕中央位置与被试视线平齐。实验分为两个阶段：类别学习阶段和测试阶段。类别学习阶段的主要目的是为了让被试学习并掌握新颖客体的不同类别。在类别学习阶段之前先给被试同时呈现两个类别客体的典型图片，让被试熟悉图片，并告知被试每类客体都有多张不同图

索任务 (见 Gilbert et al., 2006) （如图 3 ）。为确保目标刺激投射在大脑的半侧，搜索圈的视角设定为 8.5 至 10 度，偏心视角为 4.25 至 5 度 (蔡厚德，1999)。被试的眼睛与屏幕距离得到控制，以确保靶子刺激分别投射到左右视野。被试需要在快速呈现的搜索圈中判断不同的那一张图片是在左侧（"1" 或 "2" 的位置）
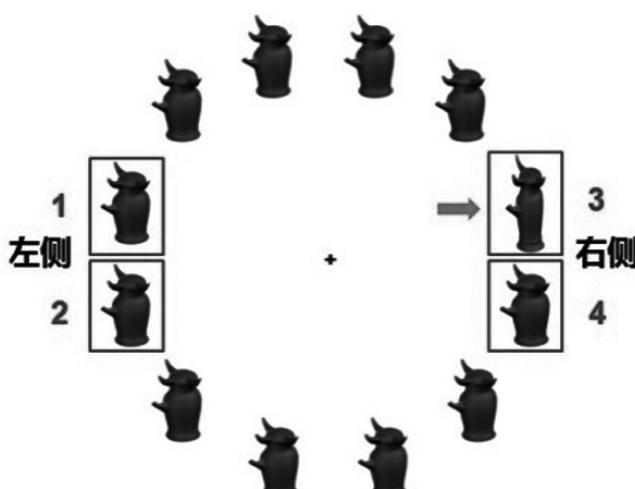
图 3 视觉搜索圈

还是右侧("3"或"4"的位置),并按键盘上相应的"左"、"右"键反应。

实验流程如下:首先屏幕中央呈现一个红色的"+",紧接着快速呈现一个图片圈,要求被试在实验中始终保持对图片圈中央"+"的注视,不随意转移目光。搜索圈消失后尽量快而准确的做出判断,按键反应之后"+"消失,进入下一个判断任务。测试阶段的实验流程图见图 4。每个实验单元的试次设置具体如下:选取 4 对图片作为制作图片圈的原始图片,类别内和类别间各 2 种。每个图片均要充当一次目标和分心刺激,因此产生 8 种搭配情况。此外,由于靶子刺激会出现在 4 个不同的位置上,因此在原来 8 种搭配的基础上共产生 32 种搭配。每种情况再重复 6 次,那么每个被试总共需要完成 192 个试次。整个实验被均衡切分成三部分,每完成一部分被试休息一次。

2.3 结果分析

被招募的 10 名参与前测的被试中有一名因测试阶段正确率低于 50% 而被剔除,最终有 9 份有效数
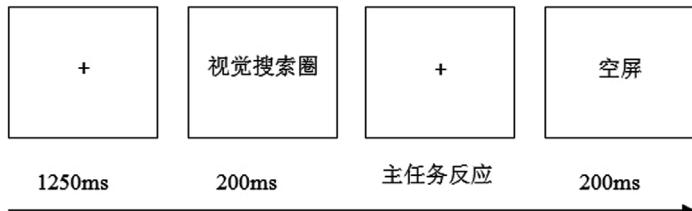
图 4 测试阶段实验流程图

据进入统计分析。运用 SPSS 17.0 对被试的类别变量(类别间 vs. 类别内)进行 t 检验,发现类别的主效应显著 ($t(8) = 2.63$, $p = 0.03$),被试对图片类别间的反应显著快于类别内图片刺激的反应,表明所制作的新颖客体在类别学习后可以产生经典类别知觉效应。

3 实验 1:语言标签表征的不同程度对新颖客体类别知觉的影响

3.1 实验目的

探讨语言标签的不同表征程度对新颖客体的类别知觉效应的影响。通过操纵两组不同被试学习新颖客体语言标签的次数,使新颖客体的语言标签在两组被试中的表征程度不一样,从而将被试分成语言标签表征程度高和表征程度低的两个组。实验 1 的假设是,如果语言标签的表征程度对类别知觉效应有影响,那么两组被试在类别知觉过程中的表现应该是不一样的:语言标签表征程度高的被试会表现出更为显著的类别知觉效应。

3.2 实验方法

3.2.1 被试

招募大学生被试 34 名(男性 9 名,女性 25 名,平均年龄 23.4),均为右利手,母语为汉语,且视力或矫正视力正常。每位被试完成实验后均可获得精美礼品一份。

3.2.2 实验材料

图片材料为在前测中测试好的图片。而语言标签则通过问卷测评在 10 名被试中对 12 个超低频字和 12 个普通低频字进行了识别测试。最终根据笔画数、结构和音节等多方面考虑,选取了识别率均为 0 的两个字"尢"和"玌",并告诉被试这两个字分别念 l í 和 j ǐ。

3.2.3 实验设计

采用 2(语言标签学习程度:高 vs. 低)× 2(客体类别:类别间 vs. 类别内)× 2(视野:左 vs. 右)三因素混合设计。其中语言标签学习程度为被试间变量,客体类别和视野为被试内变量。因变量为被试判断

靶子刺激所在视野的反应时。

本实验采用与预实验中相同的视觉搜索研究范式。与预实验不同的是，在类别学习之后、视觉搜索任务之前插入语言标签学习任务。语言标签学习分高、低两种表征程度在被试间进行，其中语言标签低表征程度学习组的被试对每个客体的名称学习 2 遍，而语言标签高表征程度学习组的被试对每个客体的名称学习 7 遍。在实验 1 中，类别学习阶段和测试阶段的视觉搜索任务要求与预实验相同。

3.2.4 实验程序

采用 E-Prime 2.0 软件对实验设计进行编程，整个实验流程在电脑上进行。实验分 3 个阶段：类别学习阶段，标签学习阶段和测试阶段。类别学习和测试阶段过程和前测中一样。

在语言标签学习阶段，首先告知被试两类图片各自的名字（"先"和"孔"），要求被试认真学习并记住它们相应的名称。实验开始时，首先屏幕中央出现一个红色的注视点"+"（1000 ms），随后出现一张图片（1000 ms），同时下方呈现其相应的语言标签名称。要求被试尽量快而准确地判断名称是否与图片相匹配，并按"是"或"否"键进行反应。如被试未在图片出现后 3 s 内做出反应，则程序自动跳转到下一个试次。每次按键后无论对错，屏幕上将立即出现反馈（300 ms）。每两个试次间的空屏时间为 200 ms。语言标签学习阶段的实验流程图见图 5。每个实验单元 12 个试次，每张图片出现 1 次。被试每学习晚所有图片后休息一次，休息时屏幕中呈现正确匹配的客体和语言标签。语言标签学习程度高的被试组对每个客体的标签学习 7 遍，语言标签学习程度低的被试组对每个客体标签学习 2 遍。

主任务反应
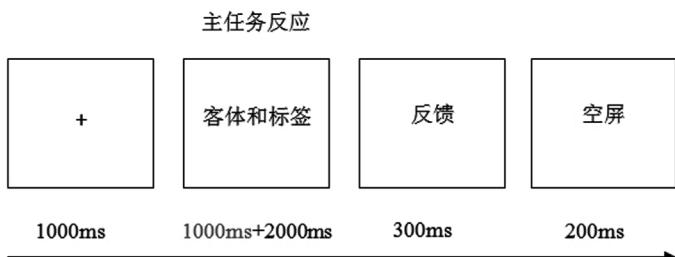
图 5　语言标签学习阶段实验流程图

3.3　结果分析

本实验中类别学习与标签学习阶段均未剔除数据，而测试阶段视觉搜索任务中的正确率和反应时的极端数据处理标准如下：首先，测试阶段正确率方面：为了排除被试猜测判断，根据二项分布定律，删除正确个数为 108 以下的被试数据。猜对与猜错的概率 p 和 q 均为 0.5，n=192，根据公式：$N=u+1.645\alpha$（$u=np$=96，α 的平方 $=npq$）得出完全凭猜测猜对 108 次以下的概率为 95%，因此将剔除正确率低于 108/192=0.56 的被试数据。同时测试阶段反应时方面：也将反应时低于 20ms 和超过 3000ms 的极端反应时试次进行了剔除。

数据经整理后共有 34 份有效数据进入统计分析，高、低标签学习条件下各 17 名被试。运用 SPSS 17.0 软件首先进行 2（标签学习程度：高 vs. 低）× 2（客体类别：类别间 vs. 类别内）× 2（视野：左 vs. 右）的三因素 ANOVA 分析。重复测量的方差分析发现，标签学习程度、客体类别和左、右视野三者之间交互作用
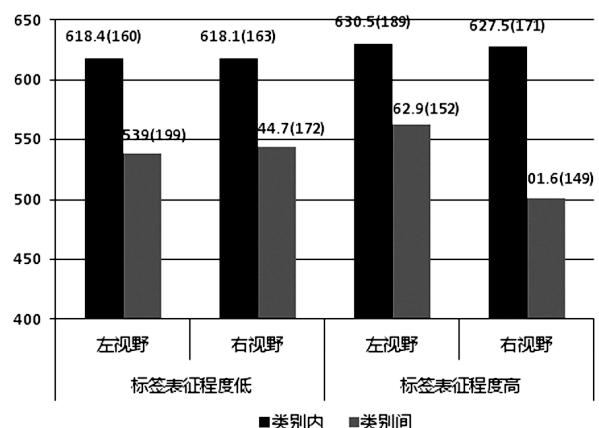
图 6　高、低语言标签学习程度下类别知觉平均反应时（ms）

显著，$F(1,32)$ =4.79，p = 0.036，η_p^2=0.13。进一步进行简单交互效应分析发现，低标签学习组的客体类别（类别间 vs. 类别内）和视野（左 vs. 右）的交互作用不显著，$F(1,32)$=0.08，p=0.77，η_p^2= 0.01；而高标签学习组的客体类别（类别间 vs. 类别内）和视野（左 vs. 右）的交互作用则显著，$F(1,16)$=7.43，p=0.015，η_p^2=0.32。在高标签学习条件下，客体类别：（类别间 vs. 类别内）与视野：（左 vs. 右）的反应时的交互作用显著，$F(1,32)$=7.87，p

=0.008, η_p^2=0.20。进一步进行简单简单效应分析得出：在左视野中，被试对类别间目标刺激的平均反应时显著快于类别内，$F(1,33)$ =5.05, p=0.031, η^2=0.13; 右视野中，被试对类别间目标刺激的平均反应时要比类别内的快，$F(1,33)$ =17.68, p<0.001, η^2=0.35。表明标签学习程度较高条件下产生了类别知觉右视野优势效应（见图 6）。

此外，在标签学习阶段，高标签学习组被试的第七遍正确率 (71.1%) 显著高于低标签学习组被试的第二遍正确率 (57.8%), t=2.47, p=0.019。客体类别知觉的主效应非常显著，$F(1,30)$=23.50, p<0.001, η^2=0.42。在低标签学习条件下 [$F(1,16)$ =5.83, p=0.028, η_p^2=0.27] 高标签学习条件下均分别显著 [$F(1,16)$=34.21, p< 0.001, η_p^2=0.68]，即无论语言标签学习情况，类别间的判断总是优于类别内的判断。

3.4 讨论

实验 1 的结果证实了实验的假设，即高、低不同语言标签表征程度的被试在类别知觉过程中有不同的表现，只有语言标签表征程度较高的被试出现了右视野优势效应。但同时也发现，增强客体与语言标签的表征强度只能促进右视野类别间客体的辨别能力，对类别内客体辨认和左视野的辨别能力均不产生影响，进一步表明语言通过左半大脑对客体类别知觉产生作用，这在一定程度上支持了 Sapir–Whorf 的语言相对论假设。

实验 1 单独考察了语言对类别知觉的影响，在接下来的实验中，将观察语言与其它因素（自我关联）交互影响类别知觉的过程。但为了更好地观察这两者的交互作用，首先单独观察自我关联程度对类别知觉的影响。

4 实验 2：客体与自我不同关联程度对类别知觉的影响：语言标签表征较低情况下的探讨

4.1 实验目的

探讨在新颖客体标签表征程度低的条件下，被分

类客体与自我的关联程度与类别知觉效应及右视野优势效应的关系。前人在探讨语言标签对类别知觉影响的研究中得到不同结论，如当客体涉及自我关联时，其类别知觉会得到促进 (Lupyan et al., 2007); 而当客体与自我无关时，有无语言标签的学习结果均无显著差异 (Holmes & Wolff, 2012)。导致结论不一致的原因有可能是由于客体与自我关联程度在两个实验中并未得到统一处理的缘故所致。因此，在实验 2 中，语言标签的表征程度被控制在一个相对较低的水平，自我与客体知觉的关联程度得到相对独立地系统地操纵，以期观察到类别知觉效应如何受到自我关联的影响。

4.2 实验方法

4.2.1 被试

招募大学生被试 30 名（男性 10 名，女性 20 名，平均年龄 22.6）。被试均为右利手，且母语为汉语。所有被试视力或矫正视力正常。每位被试完成实验后均可获得精美礼品一份。

4.2.2 实验材料

通过谷歌搜索引擎找到代表被试自我及外星人研究专家的人形剪影图片，而语言标签材料（"先"和"弚"）和外星人图片材料则与实验 1 相同。

4.2.3 实验设计

采用 2(客体与自我的关联程度：高 vs. 低) × 2(客体类别：类别间 vs. 类别内) × 2(视野：左 vs. 右) 三因素混合设计。其中客体与自我关联程度为被试间变量，客体类别和视野为被试内变量。因变量为被试判断刺激所在视野的反应时。

4.2.4 实验程序

实验 2 分为 3 个部分：类别学习阶段、标签学习阶段和测试阶段。其中标签学习阶段和实验 1 中的低表征程度语言标签学习阶段相同，测试阶段与前测中的测试阶段相同。而在类别学习阶段则在前测中的类别学习基础上融合了不同程度的自我意识信息。

在高自我关联程度诱导条件中，要求被试根据指导语想象这样的一个情景："地球即将毁灭，科学家及时寻找到另一个适合人类居住的星球，人类开始

图 7　自我关联学习阶段使用的材料
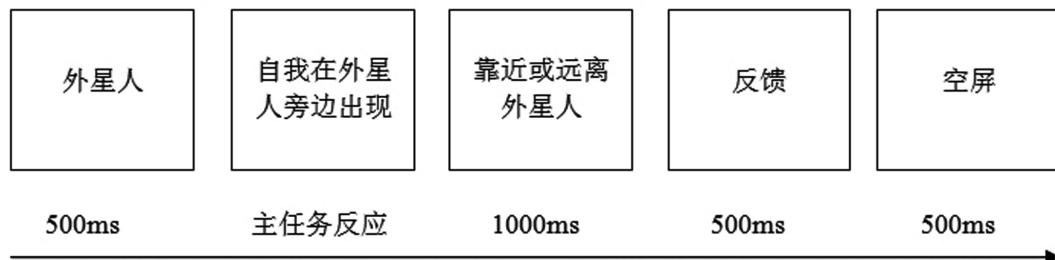
图 8　高、低语言标签学习程度下类别知觉平均反应时 (ms)

逐渐往该星球迁移。而你刚刚抵达星球，得知星球上有两种外星人，其中一种是安全型较友好的，人类可以靠近；另一种是危险型，遇到就会对人类会发起致命攻击，所以见到了必须远离。你只知道两种外星人的样貌，而不清楚哪种是安全的，哪种是危险的。为了在星球上生存下去，被试必须学会辨认出哪种外星人是安全友好可以靠近的，哪一种又是危险需要远离的"（见图 7）。在实验过程中，首先出现一个外星人（即前测中选取的新颖图片），紧接着一个代表人的人形剪影随机出现在外星人左边或右边，该人形剪影代表被试的虚拟自我。要求被试通过在键盘上按"←"键或"→"键来控制自己是远离还是亲近所呈现的外星人。主试事先并不告诉被试哪种外星人是友好的、哪种是致命的，需要被试亲自通过尝试及得到的反馈逐步学习到。

在低自我关联程度诱导条件中，被试读到的指导语基本与高自我关联条件下相同，唯一不同之处在于，不是被试自己而是一位外星人研究专家来到了外星球，这位专家必须对外星人进行安全型和危险型的分类工作。被试在屏幕上控制的人形剪影为他人（外星人研究科学家）而并非自我。被试的任务和具体实验流程与高自我关联条件下的类别学习阶段完全相同（见图 8）。

每个实验单元中的 12 张客体图片均将单个随机出现一次，每一个都判断正确后类别学习阶段才能结束，否则被试将再学习一遍直到某一个实验单元正确率达到 100% 才进入语言标签学习阶段。

4.3　结果分析

在本实验中，类别学习与标签学习阶段产生的数据均有效，没有数据被剔除。而在测试阶段视觉搜索任务中，正确率和反应时的极端数据处理标准如实验 1。数据经初步整理后发现，参与实验的 30 名被试中有 2 名被试因测试阶段正确率过低，其相应数据被剔除，最后总共有 28 份有效数据进入统计分析。采用 SPSS 17.0 软件对数据进行重复测量方差分析发现，客体类别和客体与自我的关联程度的交互作用显著，$F(1,26)=5.77$，$p=0.024$，$\eta_p^2=0.18$；客体类别的主效应显著，$F(1,26)=21.94$，$p<0.001$，$\eta_p^2=0.46$。对类别与自我关联程度的交互作用做简单效应分析发现，在低自我关联程度下，被试并未产生类别知觉效应，$F(1,26) = 2.60$，$p=0.12$，$\eta_p^2=0.09$；而在高自我关联程度下才产生了显著的类别知觉效应，$F(1,26)=25.11$，$p < 0.001$，$\eta_p^2=0.49$。但在两种关联程度条件下均未产生右视野优势效应（见图 9）。

4.4　讨论

实验 2 结果表明，在控制语言标签表征程度较低的情况下，当客体与自我关联程度较高时，被试表现出了非常显著的类别知觉效应，即客体类别间的辨认优于类别内的辨认；而当客体与自我关联程度较低时却未出现类别知觉效应。研究还发现，相对于低自我关联条件下的被试，高自我关联条件下的被试对类别
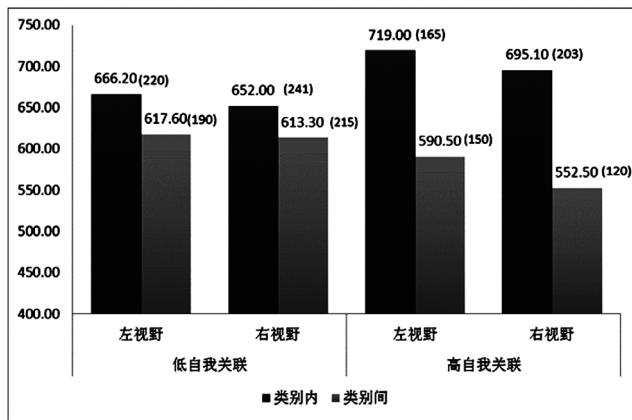
图 9　语言标签表征程度低时，高、低自我关联不同程度条件下类别知觉平均反应时 (ms)

间目标刺激的辨别更快了，且对类别内的目标刺激辨别变慢了。这似乎表明自我关联不仅增加了被试对类别间刺激的心理差异且进一步模糊了被试对类别内刺激的心理差异。这一点与前人对自我意识的探讨结果较为一致，表明由于自我关联导致了自我意识的提升，使得个体的内向注意力增大，从而进一步使得类别间的心理差异放大，类别内的心理差异缩小。同时这一点与语言标签对类别知觉的影响是不同的，语言标签在表征程度高的情况下并没有进一步模糊类别内客体间的差异，也没有进一步放大类别间客体间的差异，表明语言标签和自我关联对类别知觉的影响本质是不同的。

同时前人多个研究表明，大脑关于自我觉知的神经机制是同时分布于大脑的左右半球多个区域的 (Northoff et al., 2006; 程蕾等，2011; 杨帅，黄希庭，傅于玲，2012)，并未出现大脑偏侧化表征的现象，因此实验 2 没有发现左或右视野优势效应与自我的脑机制研究结果是一致的。

5　实验 3：客体与自我不同关联程度对类别知觉的影响：语言标签表征较高情况下的探讨

5.1　实验目的

实验 2 结果表明，当语言标签作用不显著时，高自我关联程度的类别知觉效应是显著的，自我意识对新颖客体的类别知觉有显著影响效果。但实验 2 并未探讨当语言标签也对客体的类别知觉有影响时，自我与客体的关联程度和语言标签的表征程度如何起到一个交互的作用。实验 3 假设在语言标签表征程度较高情况下，当语言标签和自我关联程度同时参与作用时，语言标签和自我关联两个因素对客体类别知觉均有影响，但语言标签的影响作用会更大。

5.2　实验方法

5.2.1　被试

37 名大学生被试 (男性 20 名，女性 17 名，平均年龄 21.9)，均为右利手，母语为汉语，且视力或矫正视力正常。每位被试完成实验后获得精美礼品一份。

5.2.2　实验材料

与实验 2 相同。

5.2.3　实验设计

采用 2(标签自我关联程度：高 vs. 低)×2(客体类别：类别间 vs. 类别内)×2(视野：左 vs. 右) 三因素混合设计。其中客体与自我的关联程度为被试间变量，客体类别和视野为被试内变量。因变量为被试判断刺激所在视野的反应时。

5.2.4　实验程序

实验 3 的 3 个部分与实验 2 基本相同，所不同的是语言的标签学习阶段是学习 7 遍的高程度语言标签表征，而不是学习 2 遍的低程度语言标签表征。

5.3　结果分析

在本实验中，类别学习与标签学习阶段所产生的数据均有效，而在测试阶段视觉搜索任务中，正确率和反应时的极端数据处理标准如实验 1。数据经初步整理后发现，在所招募的 37 名被试中有 1 名被试在测试阶段的正确率过低，其相应数据被剔除，最后共有 36 份有效数据 (高、低标签学习条件下各 18 名被试) 进入最终的统计分析。采用 SPSS 17.0 软件对有效数据进行重复测量的方差分析，发现客体类别和视野的交互作用显著，$F(1,34)=4.42, p=0.043, \eta_p^2=0.12$。对类别与视野的交互作用做进一步简单效应分析，发现左、右视野分别产生了显著的类别知觉效应 [F 左视野 $(1,34)=28.60, p<0.001, \eta_p^2=0.46; F$ 右视野 $(1,34)=35.62,$

$p<0.001$, $\eta_p^2=0.51$]。其中左视野中类别间的反应时比类别内的快65ms，右视野中类别间的反应时比类别内的快104ms，从某种程度上来说右视野的判断更具优势，将左右视野类别内与类别间的平均反应时差值进行t检验，结果发现两者差异显著，$t(35) = -2.11$, $p=0.04$。客体类别的主效应显著 [$F(1,34)=50.03$, $p<0.001$, $\eta_p^2=0.59$]，即类别间的辨别比类别内的要快，而左、右视野的主效应并不显著 [$F(1,34)=1.07$, $p=0.31$, $\eta_p^2=0.03$]，这可能是由于在低自我关联中左视野的类别内和类别间的差异不够大所致。另外，自我关联程度这个变量与客体类别 [$F(1,34)=1.84$, $p=0.18$, $\eta_p^2=0.05$] 和左、右视野 [$F(1,34)=0.05$, $p=0.82$, $\eta_p^2=0.002$] 这两个变量均未产生交互作用，其主效应也不显著，$F(1,34)=0.16$, $p=0.69$, $\eta_p^2=0.01$（见图10）。
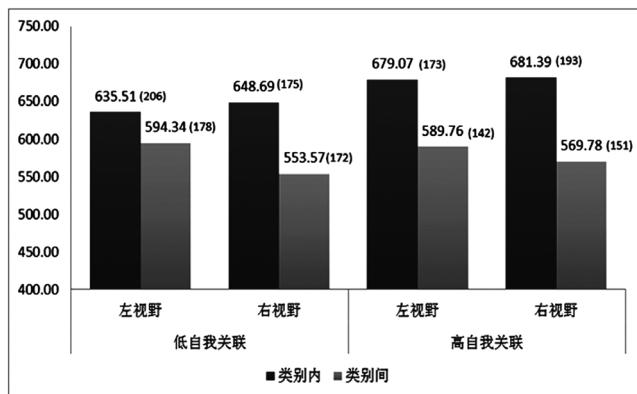
图10　语言标签表征程度高时，高、低自我关联不同程度条件下类别知觉平均反应时 (ms)

5.4 讨论

在实验2中（语言标签表征程度低的情况下），在低自我关联条件下被试并没有表现出类别知觉效应，即类别间的辨认并不占优势；而在高自我关联条件下，客体类别知觉效应才显示出来，即类别间的辨别明显优于类别内的辨认。但在实验3中（语言标签表征程度高的情况下），由于自我关联主效应并不显著，且只发现客体类别与左、右视野两因素交互作用显著，表明自我关联无论高低，右视野的类别间辨认要优于左视野的类别内辨认。但由于左、右视野并未表现出主效应，而客体类别的主效应是显著的，因此右视野的类别知觉优势效应并没有表现得很明显。这有

可能是因为语言标签的高表征程度在低自我关联条件下的作用促进了右视野的类别知觉效应，而在高自我关联条件下，客体与自我之间的高关联又促进了左视野的类别知觉效应，缩小了与右视野类别知觉之间的差异。同时，对比实验2在低自我关联条件下类别知觉效应不显著，高表征程度的语言标签在实验3的低自我关联条件下的作用使得类别间的辨认比类别内的辨认更具优势。因此，综合来看，语言标签和自我关联对类别知觉的影响并不是两者相互简单叠加，而是相互间制约权衡的结果。

但自我关联和语言标签影响力未能叠加的原因并不能通过这个研究来得到确定，这可能有两种原因：一个是两个影响因素在类别知觉过程中有可能会损耗认知资源，如两种力量的权衡过程，从而导致这两个影响力大打折扣；另一个是当两个强有力的因素存在时，只有一个因素起到主要作用。对于这个问题的本质，还有待将来进一步的研究和探讨

6　总讨论

当前研究就类别知觉过程中语言标签表征程度的影响进行了较为深入的探讨。同时，为了进一步揭示语言标签对客体类别知觉影响的本质，引入客体的自我关联变量，并将语言标签和自我关联放在一起观察两者的交互作用，以更好地探讨语言标签和其它变量如何共同影响客体的类别知觉过程。在观察语言标签的影响力时，选择渐变等距的新颖客体作为类别知觉的对象，以最大的限度接近前人对颜色的研究过程。

当前研究设计了3个实验。实验1首先探讨了语言标签的不同表征程度对新颖客体类别知觉效应的影响，结果发现仅当语言标签表征程度高时，被试才表现出显著的右视野优势效应。该结果虽然总体上支持了语言对类别知觉影响的假设，但也指出一个客体在和一个语言标签建立关联的初期，语言标签对该客体的类别知觉的影响是微乎其微的。只有当该客体的语言标签表征相对稳定之后，才会开始对其类

别知觉产生影响。

实验 2 在语言标签表征程度较低情况下探讨客体与自我之间的不同关联程度如何影响客体类别知觉。该实验结果发现当个体与客体建立自我关联时，其对客体的类别知觉显著受到自我意识的影响，即出现了类别知觉效应，且类别间刺激的心理差异进一步扩大，类别内刺激的心理差异被进一步缩小。而当自我关联较低时，客体则没有表现出类别知觉效应。但实验 2 并未发现右视野优势效应，这一点应该与自我并未出现大脑偏侧化现象、且在大脑左 / 右均有分布有关。

实验 3 探讨在语言标签表征较高条件下，客体与自我的不同关联程度对类别知觉产生的影响。结果显示当语言标签和自我同时作用于客体类别知觉时，两个变量同时对客体类别知觉产生影响。与实验 1 结果不同的是，尽管语言标签表征程度与实验 1 中的一样，但右视野优势效应并没有很明显地表现出来，只发现右视野的类别间辨别力比左视野的类别内辨别力有显著优势，表明语言标签和自我关联在共同对类别知觉起作用。与实验 2 结果不同的是，该实验中自我关联的高低在类别知觉上没有表现出差异，进一步表明语言标签和自我关联对类别知觉起到一种联合作用。

6.1 语言标签对类别知觉过程的影响

对于语言标签与类别知觉的关系，持语言普遍论者认为既然类别知觉可以脱离语言而存在，那么这两者之间应该就没有必然的联系。对这一论点的大部分证据来自于语言仍处于发展期的婴幼儿被试，少量来自成人情绪类别知觉的研究。但是，由于人的高度社会化属性，语言的发展是必然，且语言的加工在后期人的社会化过程中不断得到强化而变得越来越自动化。语言加工的这种逐步自动化属性使得它最终在整个认知加工过程中得到优先处理并进一步影响其它加工过程 (如经典的 Stroop 效应)，这一点从前人探讨语言与类别知觉关系时分化成两派不同的观点得到了印证。

对于认为语言标签对类别知觉没有影响的研究者来说，他们所招募的被试大多数是语言系统仍处于发展期的婴幼儿，其语言系统的建构并不成熟、完整，其加工还远未达到自动化程度，因此相比其它信息的认知加工并不占据优势，还无法构成对其它认知加工的影响。而对于语言系统发展完善的成人来说，语言符号加工的高度自动化使得语言不可避免参与到其它认知加工的过程中，毕竟语言符号或者概念的建构是基于初级的感知觉加工 (Barsalou, 1999, 2008)，且两者在神经加工机制上享有共同的通道和激活脑区 (Niedenthal, Barsalou, Winkielman, Krauth-Gruber, & Ric, 2005; Niedenthal, Winkielman, Mondillon, & Vermeulen, 2009)，因此认知加工的这种自上而下的加工属性决定了语言符号影响的必然性，这与大量支持语言影响类别知觉过程的研究结果是一致的。因此可以说，前人发现的所谓的相互矛盾的证据实际上是统一的，符合人的语言发展和影响的规律的。

当前的这个研究则从动态的角度在线观察到了语言对类别知觉影响的变化过程，即当语言标签的表征程度较低，其加工速度还不够快时，其对类别知觉的影响是非常有限的；而当语言标签的表征程度较高，其加工速度足够自动化时，其对类别知觉的影响则会显现出来。

至于有研究者发现成人在情绪的类别知觉过程中并未发现语言标签的影响作用，这很有可能是因为涉及情绪信息加工的脑区主要位于边缘系统的杏仁核，而涉及语言符号加工的脑区则在大脑皮层，这与其他研究者所探讨的客体类别知觉过程 (如颜色，线条，音符等) 主要也位于大脑皮层是有区别的。情绪信息的加工既然在脑机制分布上独立于语言符号加工的脑区，那么其与在大脑皮层进行的认知加工过程也很有可能是相对独立的。从进化的角度来看，人类对情绪面孔的识别应该早于语言的加工和表征，且前者对人类生存更具重要意义，其在生理结构和加工上的独立性可以使其可以得到相对完善的保护。因此人类的情绪加工系统相对独立于其它信息的加工 (Sauter et al., 2011) 是符合生物进化要求和规律的，但这个证据不能用来否定语言系统对其它在大脑皮层完成的认知加工的影响。

6.2 自我关联对类别知觉过程的影响

前人研究发现，人类自我意识的能力和程度会影响到人的行为、判断及心理健康 (Wheeler, Morrison, DeMarree, & Petty, 2008)，而探讨自我意识问题的研究者一直对自我意识在认知加工过程中的影响的本质感兴趣 (Klein & Loftus, 1988; Ma & Han, 2010; Symons & Johnson, 1997; Wang, Zhang, & Sui, 2011)。当前研究发现，自我关联对类别知觉有影响，能够产生类别知觉效应，且相比于非自我关联条件进一步扩大了类别间的心理差异、缩小了类别内的心理差异，表明自我关联所导致的自我意识的增强，很有可能是通过注意力内向化导向的方式使得类别知觉过程在心理上的认知消耗减小，从而使类别知觉过程变得更容易。

同时当前研究发现，自我关联与语言标签共同对类别知觉产生影响，且两者的影响作用呈现权衡较量结果，并非简单作用力的叠加。该结果表明：(1) 在探讨语言对其它认知加工影响的时候，有必要考虑除语言以外的其它影响因素的存在。由于语言并不是现实生活中唯一对认知加工过程有显著影响的一个因素，综合考虑其它因素与语言的交互作用是语言认知研究更具生态效应的一个标志。(2) 根据前人研究发现，自我意识对自动化的认知加工是有影响的。那么自我意识在激活过程中，很有可能同时激活了与自我相关的概念表征，而自我概念表征的激活意味着相关语言符号的激活，从而产生类似语言标签的作用。这或许就是造成语言标签和自我关联对客体类别知觉的作用不是简单叠加，而是有重叠的原因。

6.3 对外界因素影响类别知觉过程本质的思考

在人类所生存的客观世界中存在各种各样的客体，这些客体之间的差别是多样化的，有些在外形上有区别，有些在属性或功能上有所不同；有些差别很明显，而有些差别则不易为人所察觉。人类在与这些不同客体的交互作用过程中，为了更高效地处理外界的信息，必然会动用各种资源来帮助自己更有效地对这些客体进行分类。也就是说，在高级认知层面，凡是与分类有关的过程都有可能对客体的类别知觉产生促进作用，而语言标签的建立和自我意识产生的过程应该恰好都具备了这样的属性，所以才导致它们在对客体类别知觉的过程中产生相似的促进作用。

首先，根据前人对语言标签功能的研究，语言标签可以从心理层面对任意两个被划分到不同类别的客体间的差异起到放大的作用，这个作用从本质上来说是一种主观上的分类过程，在两个差异不明显的客体知觉过程中显得尤为重要。人类在前语言阶段的婴儿时期就可以做到类别知觉，表明客体的客观差异在足够明显时，是不必依赖主观上的帮助来完成分类的。但在人类适应外部世界的过程中，有很多时候需要对客观差异不够明显或客观差异过大的一些客体进行类别的划分。针对前者，人类要将看上去相似的两个客体分成两类；而针对后者，则是要将看上去不怎么相似的两个客体分成一类。这时，语言标签的作用在于放大前者的差异、缩小后者的差异，使得前者的微小差异得到注意，而后者的巨大差异得到忽略。语言的这种调节作用，表明类别知觉过程融合了自上而下和自下而上的认知加工过程。

其次，自我意识的过程实际上是将自我及与自我密切相关的因素和他人及任何其它外界因素进行分类的一个过程，这一点从个体对自我和他人具有不同的表征结构和脑区激活模式就可以推断出来 (Burris & Rempel, 2008; Dijksterhuis & Bargh, 2001; Macrae, & Roseveare, 2002; Mitchell, Macrae, & Banaji, 2006; Ruby & Decety, 2001; 杨红升，朱滢，2004; Han et al., 2008; Heatherton et al., 2006; Jenkins & Mitchell, 2011; Kelley et al., 2002; Lombardo et al., 2010; Moran, Heatherton, & Kelley, 2009; Wu, Wang, He, Mao, & Zhang, 2010; Zhu, Zhang, Fan, & Han, 2007)。因此可以说，自我意识的过程在某种程度上也是一种分类的过程，那么这个过程对外界客体的分类过程产生影响就不足为奇了。

而自我意识的产生对客体类别知觉产生促进作用而不是抑制作用，其原因很可能在于当个体对所加工的信息进行自我关联之后，会在随后的信息通达和记忆中会表现出显著优势 (Klein & Loftus, 1988; Ma & Han, 2010; Symons & Johnson, 1997; Wang et al., 2011)。前人 (Vogeley & Fink, 2003) 认为这种优势是由

个体看待事物的视角所致。个体在自我关联之后，他们看待事物的角度就变成了第一人称视角 (the first-person perspective)，而这个视角位于个体基于自身为参照所建构的多模态体验空间的中心，是自我意识形成的重要指标。而自我意识又与个体注意力内向化导向紧密相关，自我意识越强，则表明个体的内向注意力越强，个体越不容易受到外界事物的干扰或启动 (Dijksterhuis & van Knippenberg, 2000)，即自我意识的提升或降低会对行为的启动效应 (prime-to-behavior effects) 产生影响，而且对自动化行为的影响也是多方面、多样化的 (Wheeler et al., 2008)。综合这些研究结果，可以推测出，一旦将某个客体或事件自我关联之后，自我意识会随之提升，这种自我意识的提升对同时进行的其它认知行为，特别是已自动化的认知行为，产生促进作用。

7 结论

当前研究通过采用经典的视觉搜索范式，从语言标签和客体的自我关联两个角度探讨了高级概念层面的信息对新颖客体类别知觉的影响。研究结果显示：(1) 语言标签表征程度的增强可促进新颖客体类别知觉的右视野优势效应，且这种优势只体现在对右视野类别间客体的辨别上。(2) 客体与自我关联程度的提高会促进左、右视野的类别知觉效应；且不仅增大了类别间客体的心理差异，还缩小了类别内客体的心理差异。(3) 语言标签的表征程度和客体与自我关联程度同时增强时，语言标签的作用依旧表现出来，但与自我关联的影响产生权衡分配，且其影响力并不足以产生右视野优势效应。

参考文献

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, 22(4), 577–609.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, 59, 617–645.

Burris, C. T., & Rempel, J. K. (2008). Me, myself, and us: Salient self-threats and relational connections. *Journal of Personality and Social Psychology*, 95(4), 944–961.

Cai, H. D. (1999). Some methodological issues of tachistoscopic visual half-field technique. *Psychological Science, 22*(3), 265–272.

[蔡厚德 . (1999). 半视野速示技术的若干方法学问题 . *心理科学* , 22(3), 265–272.]

Cheng, L., Chen, X. H., & Huang, X. T. (2011). Role of the right or the left hemisphere? Dispute on neural mechanism of Self-awareness. *Advances in Psychological Science*, 19(9), 1319–1327.

[程蕾 , 陈煦海 , 黄希庭 . (2011). 左脑还是右脑——自我觉知神经机制的述评 . *心理科学进展* , 19(9), 1319–1327.]

Dijksterhuis, A., & Bargh, J. A. (2001). The perception-behavior expressway: Automatic effects of social perception on social behavior. *Advances in Experimental Social Psychology*, 33, 1–40.

Dijksterhuis, A., & Van Knippenberg, A. (2000). Behavioral indecision: Effects of self-focus on automatic behavior. *Social Cognition*, 18(1), 55–74.

Franklin, A., & Davies, I. R. L. (2004). New evidence for infant colour categories. British *Journal of Developmental Psychology*, 22(3), 349–377.

Franklin, A., Pilling, M., & Davies, I. R. L. (2005). The nature of infant color categorization: Evidence from eye movements on a target detection task. *Journal of Experimental Child Psychology*, 91(3), 227–248.

Franklin, A., Clifford, A., Williamson, E., & Davies, I. R. L. (2005). Color term knowledge does not affect categorical perception of color in toddlers. *Journal of Experimental Child Psychology*, 90, 114–141.

Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R. L., Kay, P., & Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants,

but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences of the United States of America, 105*(9), 3221–3225.

Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. L. (2008). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(47), 18221–18225.

Franklin, A., Catherwood, D., Alvarez, J., & Axelsson, E. (2010). Hemispheric asymmetries in categorical perception of orientation in infants and adults. *Neuropsychologia*, *48*, 2648–2657.

Fugate, J. M. B. (2013). Categorical perception for emotional faces. *Emotion Review*, *5*(1), 84–89.

Gentner, D., & Goldin–Meadow, S. (Eds.). (2003). Language in mind: *Advances in the study of language and thought*. Cambridge, MA: MIT Press.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(2), 489–494.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2008). Support for lateralization of the whorf effect beyond the realm of color discrimination. *Brain and Language*, *105*(2), 91–98.

Goldstein, J., Davidoff, J., & Roberson, D. (2009). Knowing color terms enhances recognition: Further evidence from English and Himba. *Journal of Experimental Child Psychology*, *102*(2), 219–238.

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. Wiley *Interdisciplinary Reviews: Cognitive Science, 1,* 69–78.

Gumperz, J. J. & Levinson, S. C. (1996). Introduction to part I. In J. J. Gumperz, & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 21–35).

Cambridge: Cambridge University Press.

Han, S., Mao, L., Gu, X., Zhu, Y., Ge, J., & Ma, Y. (2008). Neural consequences of religious belief on self–referential processing. *Social Neuroscience*, *3*, 1–15.

Heatherton, T. F., Wyland, C. L., Macrae, C. N., Demos, K. E., Denny, B. T., & Kelley, W. M. (2006). Medial prefrontal activity differentiates self from close others. *Social Cognitive and Affective Neuroscience*, *1*, 18–25.

Holmes, K., & Wolff, P. (2012). Does categorical perception in the left hemisphere depend on language? *Journal of Experimental Psychology: General, 141*(3), 439–443.

Jenkins, A. C., & Mitchell, J. P. (2011). Medial prefrontal cortex subserves diverse forms of self–reflection. *Social Neuroscience*, *6*, 211–218.

Kelley, W. M., Macrae, C. N., Wyland, C. L., Caglar, S., Inati, S., & Heatherton, T. F. (2002). Finding the self? An event–related fMRI study. *Journal of Cognitive Neuroscience*, *14*, 785–794.

Kesebir, S., & Oishi, S. (2010). A spontaneous self–reference effect in memory. *Psychological Science*, *21*, 1525–1531.

Kim, K., & Johnson, M. K. (2012). Extended self: Medial prefrontal activity during transient association of self and objects. *Social Cognitive and Affective Neuroscience*, *7*, 199–207.

Klein, S., & Loftus, J. (1988). The nature of self–referent encoding: The contributions of elaborative and organizational processes. J*ournal of Personality and Social Psychology*, *55*, 5–11.

Klein, S., Rozendal, K., & Cosmides, L. (2002). A social–cognitive neuroscience analysis of the self. S*ocial Cognition*, *20*, 105–135.

Klein, M. E., & Zatorre, R. J. (2011). A role for the right superior temporal sulcus in categorical perception of musical chords. *Neuropsychologia*, *49*(5), 878–887.

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Sadek, S. A., Pasco, G., Wheelwright, S. J., ⋯ Baron–Cohen, S. (2010). Atypical neural self–representation in autism. *Brain, 133*, 611–624.

Lupyan, G. (2012). Linguistically modulated perception and cognition: The label–feedback hypothesis. *Frontiers in Psychology*, 3, 54.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, 18(12), 1077–1083.

Ma, Y., & Han, S. (2010). Why respond faster to the self than others? An implicit positive association theory of self advantage during implicit face recognition. Journal of Experimental Psychology: *Human Perception and Performance*, 36(3), 619–633.

Macrae, C. N., & Roseveare, T. A. (2002). I was always on my mind: The self and temporary forgetting. *Psychonomic Bulletin & Review*, 9(3), 611–614.

Malt, B., & Wolff, P. (2010). Words and the mind: *How words capture human experience.* Oxford University Press.

Mitchell, J. P., Macrae, C. N., & Banaji, M. R. (2006). Dissociable medial prefrontal contributions to judgments of similar and dissimilar others. *Neuron, 50*(4), 655–663.

McCullough, S., & Emmorey, K. (2009). Categorical perception of affective and linguistic facial expressions. *Cognition*, 110(2), 208–221.

Mo, L., Xu, G. P., Kay, P., & Tan, L.H. (2011). Electrophysiological evidence for the left–lateralized effect of language on preattentive categorical perception of color. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 14026–14030.

Moran, J. M., Heatherton, T. F., & Kelley, W. M. (2009). Modulation of cortical mid–line structures by implicit and explicit self–relevance evaluation. *Social Neuroscience*, 4, 197–211.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth–Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9*(3), 184–211.

Niedenthal, P. M., Winkielman, P., Mondillon, L., & Vermeulen, N. (2009). Embodiment of emotion concepts. *Journal of Personality and Social Psychology*, 96(6), 1120–1136.

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self–referential processing in our brain—A meta–analysis of imaging studies on the self. *NeuroImage*, 31(1), 440–457.

Ozturk, O., Shayan, S., Liszkowski, U., & Majid, A. (2013). Language is not necessary for color categories. *Developmental Science, 16*(1), 111–115.

Paluy, Y., Gilbert, A., Baldo, J., Dronkers, N., & Ivry, R. (2011). Aphasic patients reveal a reversal of hemispheric asymmetries in categorical color perception. *Brain and Language*, 116, 151–156.

Peng, G., Zheng, H. Y., Gong, T., Yang, R–X., Kong, J. P., & Wang, W. S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics, 38*(4), 616–624.

Powell, L. J., Macrae, C. N., Cloutier, J., Metcalfe, J., & Mitchell, J. P. (2010). Dissociable neural substrates for agentic versus conceptual representations of self. *Journal of Cognitive Neuroscience*, 22(10), 2186–2197.

Roberson, D., Damjanovic, L., & Pilling, M. (2007). Categorical perception of facial expressions: Evidence for a "category adjustment" model. *Memory & Cognition*, 35(7), 1814–1829.

Roberson, D., & Davidoff, J. (2000). The categorical perception of colors and facial expressions: the effect of verbal interference. *Memory and Cognition*, 28(6),

977–986.

Roberson, D., Davies, I., & Davidoff, J. (2000). Color categories are not universal: Replications and new evidence from a stone age culture. J*ournal of Experimental Psychology: General*, *129*(3), 369–398.

Roberson, D., Davidoff, J., Davies, I. R. L., & Shapiro, L. R. (2005). Color categories: Evidence for the cultural relativity hypothesis. *Cognitive Psychology*, *50*(4), 378–411.

Roberson, D., Pak, H., & Hanley, J. R. (2008). Categorical perception of colour in the left and right visual field is verbally mediated: Evidence from Korean. *Cognition*, *107*(2), 752–762.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self–reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*(9), 677–688.

Ruby, P., & Decety, J. (2001). Effect of subjective perspective taking during simulation of action: A PET investigation of agency. *Nature Neuroscience*, *4*(5), 546–550.

Sapir .E(1921) Language : An Introduction to Study of Speech. NewYork :Harhcount,Brance & Company

Sauter, D. A., LeGuen, O., & Haun, D. B. (2011). Categorical perception of emotional facial expressions does not require lexical categories. *Emotion*, *11*(6), 1479–1483.

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: Evidence from self–prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 1105–1117.

Symons, C. S., & Johnson. B.T. (1997).The self–reference effect in memory: A meta–analysis .*Psychological Bulletin*, *121*(3), 371–394

Thierry, G., Athanasopoulos, P., Wiggett, A., Dering, B., & Kuipers, J. R. (2009). Unconscious effects of language–

specific terminology on preattentive color perception. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(11), 4567–4570.

Turk, D. J., Van Bussel, K., Brebner, J. L., Toma, A. S., Krigolson, O., & Handy, T. C. (2011). When "it" becomes "mine" : Attentional biases triggered by object ownership. *Journal of Cognitive Neuroscience*, *23*, 3725–3733.

Vogeley, K., & Fink, G. R. (2003). Neural correlates of the first–person–perspective. *Trends in Cognitive Sciences*, *7*, 38–42.

Wang, L. Y., Zhang, M., & Sui, J. (2011). Self–face advantage benefits from a visual self–reference frame. Acta *Psychologica Sinica*, *43*, 494–499

Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(19), 7780–7785.

Wheeler, S. C., Morrison, K. R., DeMarree, K. G., & Petty, R. E. (2008). Does self–consciousness increase or decrease priming effects? It depends. *Journal of Experimental Social Psychology*, *44*, 882–889.

Whorf, B. L. (1956). Language, thought, and reality: *Selected writings of Benjamin Lee Whorf*. New York: MIT Press.

Wu, Y., Wang, C., He, X., Mao, L., & Zhang, L. (2010). Religious beliefs influence neural substrates of self–reflection in Tibetans. *Social Cognitive and Affective Neuroscience*, *5*, 324–331.

Xie, S. S., Zhang, J. J., He, X. M., & Xiao, E. P. (2008). Culture's effects on 'black' and 'white' Color cognition of undergraduates from Yi Nation, Bai Nation, Naxi Nation and Han Nation. *Acta Psychologica Sinica*, *40*(8), 890–901.

[谢书书 , 张积家 , 和秀梅 , 林娜 , 肖二平 (2008). 文化差异影响彝、白、纳西和汉族大学生对黑白的认知 .

心理学报, *40*(8), 890–901.]

Yang, H. S., & Zhu, Y. (2004). The self and retrieval—induced forgetting. *Acta Psychologica Sinica, 36*(2), 154–159.

[杨红升 , 朱滢 . (2004). 自我与提取诱发遗忘现象 . *心理学报* , *36*(2), 154–159.]

Yang, S., Huang, X.T., & Fu, Y. L.(2012).Medial Prefrontal Cortex: Neural Basis of the Self. Advances in Psychological Science,*20*(6),853–862

[杨帅 , 黄希庭 , 傅于玲 .（2012）内侧前额叶皮质——"自我"的神经基础, *心理科学进展*, *20*(6), 853–862.]

Yao, S. X., Yang, D., Qi, S. Q., Lei, Y., & Ding, C. (2012). Studies on the N2pc component in visual spatial attention. *Advances in Psychological Science, 20*(3), 365–375.

[姚树霞 , 杨东 , 齐森青 , 雷燕 , Ding, C. (2012). 视觉空间注意研究中的 N2pc 成分述评 . *心理科学进展* , *20*(3), 365–375.]

Yuan, C. P., & Lu, G., M. (2010). The advance of Self–reference processing of fMRI. *Chinese Journal of Medical Imaging Technology*, *26*(12), 2382–2384.

[袁翠平 , 卢光明 . (2010). 自我参照加工的功能磁共振成像研究进展 . *中国医学影像技术* , *26*(12), 2382–2384.]

Zhang, J. J., Liu, L. H., Chen, X., & He, X. M. (2008). The relationship of color cognition and color terms in Naxi people. *Minority Languages of China*, (2), 49–55.

[张积家 , 刘丽红 , 陈曦 , 和秀梅 (2008). 纳西语颜色认知关系研究 . *民族语文* , (2), 49–55.]

Zhou, K., Mo, L., Kay, P., Kwok, V. P. Y., Tiffany, N. M. I., & Tan, L. H. (2010). Newly trained lexical categories produce lateralized categorical perception of color. *Proceedings of the National Academy of Science of the United States of America, 107*(22), 9974–9978.

Zhu, Y., Zhang, L., Fan, J., & Han, S. (2007). Neural basis of cultural influence on self–representation. *NeuroImage*, *34*, 1310–1316.

The influence of language labels and self-reference on new object categorical perception

LIU Siyun; MENG Jianxin

(Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education;

School of Psychology, Central China Normal University, Wuhan 430079 China)

Abstract　Categorical perception is one of the most basic cognitive processes of human beings. When humans process incoming information, categorization help them further clarify and simplify the sophisticated inputs. The categorical perception effect refers to the phenomenon that individuals will respond faster or more accurately when discriminating two stimuli that cross a category boundary than when discriminating two stimuli from the same category, despite between– and within–category stimuli being equated in distance (Bornstein & Korda, 1984). However, inconsistent evidence has been obtained from previous studies on how language would influence categorical perception. The language label theory suggested that the language label is a cue that helps individuals to categorize information unconsciously and automatically; whereas perceptual feature theory suggested that categorical perception is based on a pure perceptual process, which arises from life experiences that eventually changes the mappings of perceptual neurons. The current study systematically investigates how language labels would interact with self–reference factor to play a mutual role on new object categorical perception. The hypothesis of this study is that language labels are not the only important factors that would influence categorical perception, other social or personality factors may also play a role, but it is not sure whether language labels would play a more important role. In this study, three experiments were conducted. In Experiment 1, a 2 (language label learning times：more vs. less) × 2 (object category：between vs. within) × 2 (visual field：left vs. right) mixed experiment was designed, and it is found that only the participants who have learned the labels for more times showed a significant right visual field advantage effect. In Experiment 2, a 2 (object self–reference connection: tight vs. loose) × 2 (object category：between vs. within) × 2 (visual field：left vs. right) mixed experiment was designed. This time, all the participants learned the language labels for only twice. As a result, it is found that only the participants who had built a tight self–reference connection with the new object revealed a significant categorical perception effect, however, on both left– and right–visual fields. Furthermore, higher level of self–reference made the participants show a better discriminating ability for between–category objects, but not for within–category objects. In Experiment 3, a 2 (object self–reference connection：tight vs. loose) × 2 (object category：between vs. within) × 2 (visual field: left vs. right) mixed experiment was designed. This time all the participants learned the language labels for seven times. As a result, it is found that under higher level representation of language labels, both language label and the self–reference play a role on categorical perception. In summary, this study revealed an important and complicated role of language labels on categorical perception, and a nonetheless very important influence of self–reference as well.

Keywords　categorical perception; language label; self–reference; right visual field advantage effect

网络使用经验对动作动词加工的影响 *

刘思耘，周宗奎，李 娜

（青少年网络心理与行为教育部重点实验室；华中师范大学心理学院 , 武汉 430079）

摘 要 具身认知理论认为高级概念认知和低级感知觉认知紧密关联，且两者共享相同的神经系统。本研究设计了 3 个实验：实验 1 观察面部表情动词加工是否受面部情绪表达的影响，结果发现网络使用经验多的被试在促进和抑制面部积极表情条件下对动词的反应没有表现出显著差异。实验 2 探讨网络使用经验对肢体动作词汇加工的影响，结果发现网络使用经验多的被试在动词 – 名词转换时没有产生显著认知损耗。实验 3 探讨两类被试对正常序列和随机序列图式动词加工是否有差异，结果发现网络使用经验多的个体对随机序列和正常序列的回忆成绩并无显著差异。本研究从动词认知加工角度验证了高级认知与低级感知觉加工间的紧密联系，揭示了网络行为与高级认知加工间的相关关系。

关键词 具身认知；转换消耗效应；图式；网络使用经验

分类号 B842

1 前言

根据 2012 年中国互联网络信息中心（CNNIC）发布的中国互联网网络发展状况，中国 30 岁以下的青少年网民占整体比例的 56.8%，而网瘾者在 2007 年就达到 2.4%，可见上网人数及成瘾人数之多。自从有互联网以来，人们花在网络上的时间越来越多，并越来越依赖于网络以处理各种日常事宜。那么网络科技给人们生活模式带来的改变是否会影响人的高级认知加工呢？如果有影响，会在哪方面造成影响？影响程度如何？本研究将比较不同程度网络使用经验者的语言加工过程，探讨网络使用经验与高级认知加工间关系，揭示网络行为对人们生活影响的实质。

1.1 网络成瘾行为对高级认知加工的影响

当网络成瘾行为极大程度地影响到人们的生活时，研究者开始研究其形成机制及对认知功能的影响，并提出相应的理论模型（牛更枫，孙晓军，周宗奎，魏华，2013）。对网络成瘾行为的研究主要从成瘾者的人格特质的角度及影响网络成瘾环境等因素展开讨论（Weinstein & Lejoyeux, 2010; 贺金波，郭永玉，向远明，2008），对网络成瘾者认知功能变化的探讨则主要采用脑成像和脑电信号采集的方法，但网瘾形成和维持机制还不太明确（Weinstein & Lejoyeux, 2010; 戴珅懿，马庆国，王小毅，2011; 黄敏等，2010）。

研究者对网络成瘾的认知神经机制探讨的结果发现，成瘾者在认知的多个方面具有功能受损的表

现，具体如下：1）成瘾者的大脑信息加工能力降低（郁洪强等，2009；贺金波，郭永玉，柯善玉，赵仑，2008）。2）成瘾者面孔和表情识别能力降低（高文斌，陈祖妍，2006；赵仑，高文彬，2007）。3）成瘾者的记忆和注意能力受损，但该结论是基于对芬兰 15–64 岁的 2000 名被试的问卷调查（N?si & Koivusilta, 2013）。4）成瘾者的语言加工能力受损（金璞，傅先明，钱若兵，牛朝诗，韩晓鹏，2009）。5）成瘾者的思维认知能力受损（郁洪强等，2009）。6）成瘾者的决策能力受损（Pawlikowski & Brand, 2011；梁三才，游旭群，2010；Sun, Chen, Ma, Zhang, Fu, & Zhang, 2009）。

尽管多个研究发现网络成瘾者的高级认知加工能力受到严重损坏，但研究者大多采用认知神经的方法和技术，从脑神经机制的角度来推断网络成瘾者的认知功能变化，而较少从行为学的角度系统探讨网络成瘾者认知加工能力的状况。究其原因，可能是网络成瘾者在行为上已经与正常个体的行为产生了明显分化，而如果在实验室环境下仍只探讨成瘾者群体的行为特异性，其启示意义远不如探究其脑神经机制的改变。但是，对于大多数对网络有一定程度依赖但还远未达到病理性成瘾状态的网络使用者来说，首先从行为层面去界定他们的行为本质及与正常行为的偏差程度是具有重要意义的。本研究将关注这类群体，他们不仅使用网络时间较长，且在一定程度上对网络已形成心理依赖，但其心理健康受损程度并没有达到病理性的严重程度。我们称这类群体为网络使用经验多的群体。

1.2 具身认知为网络行为研究提供的新视角

我们真实的社交空间是具身的，而网络虚拟空间是非具身的（Hanlon, 2001）。当人们在虚拟空间中进行各种交际活动时，很多社交场合所必需的肢体动作、面部表情等都大大削弱和减少，这使得人们不得不更多地依赖于抽象的语言符号或者人工表情符号去弥补交流时线索的缺乏（Gackenbach, 2011）。网络的使用还使得人们的身体活动范围和频率也受

到限制，使一些必要的肢体的动作和运动逐步减少，最终导致具身体验机会的减少（Alessi, 2001; Hanlon, 2001; Kang, 2007; Kim, et al., 2012）。本研究将采用具身认知的视角探讨网络使用经验与高级认知加工间的关系。

1.2.1 具身认知的主要理论及其相关支持性证据

高级认知加工的理论中影响最大的有两种观点：一个是传统的符号加工理论，它认为高级的认知加工只涉及抽象的符号运算，并不涉及低级感知 – 运动系统的参与（Fodor, 1975）；另一种观点认为在概念加工过程中，低级感知觉系统也会得到激活（Damasio, 1989; Snodgrass, 1984），并逐渐形成具身认知理论（embodied cognition theory）。具身认知理论认为认知是基于身体的各种体验形成，个体高级的认知加工过程会受到身体状态的直接影响，大脑对先前的感知 – 运动经验的再激活（re-activation）在认知加工过程中扮演着至关重要的作用。

具身认知理论包括身体观和模拟观，这两个观点并不对立，而是相辅相成，从不同侧面印证身体体验和高级认知加工之间的紧密关系。身体观强调自下而上的加工路径，而模拟观则强调自上而下的加工路径。身体观是指实际的身体状态（如身体的各种感知觉体验、姿势、动作等）会影响高级认知加工（Larsen, Kasimatis, & Frey, 1992; Strack, Martin, & Stepper, 1988）；而模拟观则指高级的认知加工会激活先前经验所涉及的感知觉、运动、内省系统（Barsalou, 1999, 2008; Barsalou, Kyle Simmons, Barbey, & Wilson, 2003; 谢久书，张常青，王瑞明，陆直，2011）。

前人从模拟观的角度提供了大量行为学层面的支持性证据（Marques, 2006; Pecher, Zeelenberg, & Barsalou, 2003; Vermeulen, Corneille, & Niedenthal, 2008）以及神经机制方面的证据（Goldberg, Perfetti, & Schneider, 2006; Kellenbach, Brett & Patterson, 2001; Simmons et al., 2007）。在行为学实验中，研究者主要采用转换消耗效应（switching costs effect）和知觉负荷效应（sensory load effect）（Pecher et al., 2003; Marques, 2006）范式。如有研究者（Pecher et al.,

2003）发现当要求被试对所呈现的词进行属性判断，且先后呈现的特征词均属相同感觉通道时（如头发 – 黑色的 vs. 芒果 – 黄色的）被试的反应比不同通道时（如头发 – 黑色的 vs. 草莓 – 酸酸的）要快。在采用知觉负荷效应范式时，研究者（Vermeulen et al., 2008）发现，当第一次呈现的感觉刺激占用了概念特征判断任务所需的资源时，被试对随即呈现的相同感觉通道的刺激加工会变缓，表明概念加工对低级感觉通道有再次激活的功能（Vermeulen et al., 2008）。在神经机制方面，研究者用 PET 和 fMRI 技术均验证了高级概念加工对与该概念相关的感知觉脑区的再激活（Goldberg et al., 2006; Kellenbach et al., 2001; Simmons et al., 2007）。

不仅客体的概念加工会再激活低级感知觉通道，复杂的社会认知加工也会出现类似现象。许多有关情绪信息的加工研究（Halberstadt, Winkielman, Niedenthal, & Dalle, 2009; Oosterwijk, Rotteveel, Fischer, & Hess, 2009; Price, Peterson, & Harmon–Jones, 2012; Tan, Walter, Scheck, Hrabal, Hoffmann, Kessler, & Traue, 2012; Vermeulen, Niedenthal, & Luminet, 2007; Wiswede, Münte, Kr?mer, & Rüsseler, 2009）和社会知觉加工研究（Andersen, Reznik, & Manzella, 1996; Vanman, Paul, Ito, & Miller, 1997）均提供了支持性证据。如当由感觉通道属性判断转向情绪信息判断时，反应时会延长（Vermeulen et al., 2007）；情绪概念的加工会激活面部肌肉运动系统（Halberstadt et al., 2009; Price et al., 2012; Tan et al., 2012; Wiswede et al., 2009），个体的身体体验也会受到高级社会认知的影响（Andersen et al., 1996; Vanman et al., 1997）。

研究者从身体观的角度也提供了诸多支持性证据。研究者发现个体的身体体验会反过来影响社会认知加工，其中面部肌肉运动、身体姿态和各种感知觉体验对情绪、社会知觉、态度都会产生显著影响。如研究者通过对比面部肌肉受损和正常被试两大群体发现，面部肌肉运动会影响个体对情绪面孔的识别（Ponari, Conson, D'Amico, Grossi, & Trojano, 2012; Oberman, Winkielman, & Ramachandran,

2007; Niedenthal, 2007; Neal & Chartrand, 2011; Havas, Glenberg, Gutowski, Lucarelli, & Davidson, 2010; Williams, Whiten, & Singh, 2004; Oberman et al., 2007; Stel & van Knippenberg, 2008）；个体身体姿态的具身体验对其情绪的表达（Duclos et al., 1989; Van den Stock, Righart, & de Gelder, 2007; Wallbott, 1998）和社会态度的形成（Briñol & Petty, 2008; Cacioppo, Priester, & Berntson, 1993; Förster & Strack, 1998; Taylor, Lord, & Bond, 2009; Wells & Petty, 1980）也有重要影响；个体的各种感知觉体验如温度感觉、触觉、物理距离知觉等也能影响某些社会知觉（Ackerman, Nocera, & Bargh, 2010; IJzerman & Semin, 2009; Jostmann, Lakens, & Schubert, 2009; Williams & Bargh, 2008）。

1.2.2 与动作动词加工相关的具身研究

语言中动词表征的是人们在实际生活中逐步形成的典型动作，是联接不同客体间关系的重要逻辑枢纽（Tanenhaus, Carlson, & Trueswell, 1989）。研究者从行为学和脑神经机制两方面为动词加工提供了大量实证证据。在行为学方面，研究者发现动词短语的理解和动作系统存在着相互作用，即当被试执行一个动作时，若同时阅读一个和该动作方向一致的动词短语，则被试执行该动作的速度会得到促进（Glenberg & Kaschak, 2002; Lindemann, Stenneken, Van Schie, & Bekkering, 2006）。研究者认为被试在短语理解过程中对所描述的动作进行了模拟，语言理解过程再激活了相应的运动系统，从而对共享相同运动系统的动作产生了促进作用。这个效应表明动词理解是以低级运动系统的再激活为基础的，需要动作经验的参与。

研究者还采用 fMRI（Hauk, Johnsrude, & Pulvermüller, 2004; Filimon, Nelson, Hagler, & Sereno, 2007）、TMS（Buccino, Riggio, Melli, Binkofski, Gallese, & Rizzolatti, 2005; Pulvermüller, Hauk, Nikulin, & Ilmoniemi, 2005）、MEG（Hauk & Pulvermüller, 2004; Hauk, Shtyrov, & Pulvermüller, 2008）等技术手段对动词加工过程进行了深入的探讨，从脑机制层面进一步验证了动词加工过程中个体对相应动作的模拟（De Zubicaray, Postle, McMahon, Meredith, & Ashton,

2010）。

1.2.3 长期具身经验对高级认知加工的影响

研究发现不仅短暂的身体体验变化对高级认知有影响，长期具身经验也具类似影响。其中躯体特异性理论（body-specificity hypothesis）（Casasanto, 2009）就指出主体与外界环境互动经验的不同会形成不同的概念加工方式。如右利手被试倾向于将喜欢的、好的物体放在右侧，将讨厌的、不好的物体放在左侧；而左利手被试则正好相反。这种不同的身体经验造成了他们对"好"、"坏"概念的不同空间表征（Casasanto, 2009; Casasanto & Henetz, 2012）。

概念隐喻理论（Lakoff & Johnson, 1999）也认为身体经验是概念的投射源，概念系统是以低级的感知觉经验为基础形成的。如果作为投射源的身体经验不足，那么身体经验和概念之间的隐喻联接强度就会减弱，最终会影响对相应概念的理解。如早期缺乏与母亲身体接触的个体会表现出较弱的物理温度 – 人际间情感（physical warmth – interpersonal warmth）隐喻联接强度（Fay & Maner, 2012; Ijzerman, Karremans, Thomsen, & Schubert, 2013）。也有研究发现由于女性比男性具有更丰富的面部表情体验，因此女性对情绪的表达与识别都强于男性（LaFrance & Hecht, 2000）。

1.3 问题提出与研究假设

基于以上文献的回顾和分析，我们提出如下实验假设：一、面部动作具身体验的减少将会影响个体对面部动作词汇的加工；二、肢体具身体验的减少将会影响人们对肢体动作词汇的加工；三、肢体具身体验的减少将会影响人们对动作图式中动词系列的加工。

上述的三个假设将从不同的角度和层面探测网络使用经验不同的群体在动词加工过程中的差异。其中前两个假设的提出是基于人们在真实场景的社交过程中会涉及大量面部表情动作和肢体动作，而网络使用经验多的群体会由于将更多的时间用于网络，从而缺失足够的真实社交体验。假设三是在前两个假设的基础上进一步深入探测网络使用经验多的群体是否在一系列具有逻辑关联的动词加工上会有所不同表现。3 个实验从一定的广度和深度探讨了网络使用经验与高级认知加工之间的关系。

2　实验 1：网络使用经验对面部动作词汇加工的影响

2.1 实验目的

探讨网络使用经验的多少与面部表情动词加工间关系。实验一假设，网络使用经验过多的个体不易受具身状态的影响，即当这类群体的面部表情动作受到抑制或促进时，其相应概念的加工不会发生改变；而网络使用经验少的群体则会有明显的变化。

2.2 实验方法

2.2.1 被试

本研究中被试备选库的准备：抽取出白羽和樊富珉（2005）的《大学生网络依赖测量量表》中的人际健康维度、周治金和杨文娇（2007）的《网络成瘾类型》中网络游戏成瘾维度和网络信息成瘾维度编制成一份新问卷，作为筛选出本研究所需的网络使用经验多和少的被试的测量工具。此问卷共 19 个项目，每个项目采用 5 点计分方式，从"极不符合"到"非常符合"分值依次增加。从某高校中随机选取 149 名大学生进行施测，问卷回收率 100%。最终分别选取总分在两端的 60 名被试作为备用被试，其中低分组 30 名（男生 11 名，女生 19 名）作为网络使用经验少的被试群体；高分组 30 名（男生 15 名，女生 15 名）作为网络使用经验多的被试群体。被试的年龄范围在 19 岁 ~25 岁之间（M=22.42，SD=2.16）。经独立样本 t 检验，网络使用经验多的备用被试群体在总分、人际健康维度、网络信息成瘾维度、网络游戏成瘾维度及日平均上网时间上的平均得分均显著高于网络使用经验少的备用被试群体（$t_{总分}$（58）=18.48, p<0.001; $t_{健康维度}$（58）=10.98, p<0.001; $t_{信息成瘾维度}$（58）=9.10, p<0.001; $t_{游戏成瘾维度}$（58）=8.36, p<0.001; $t_{上网时间}$（58）=7.11, p<0.001）。

实验一被试：从 60 名备用被试中随机选取低分

组18名（男生6名，女生12名）作为网络使用经验少的被试群体，高分组20名（男生11名，女生9名）作为网络使用经验多的被试群体。被试的年龄范围在19岁~25岁之间（$M=22.84$，$SD=2.01$）。经独立样本t检验，网络使用经验多的被试群体在总分、人际健康维度、网络信息成瘾维度、网络游戏成瘾维度和日平均上网时间上的平均得分均显著高于网络使用经验少的被试群体（$t_{总分}(36)=17.90$，$p<0.001$；$t_{健康维度}(36)=9.50$，$p<0.001$；$t_{信息成瘾维度}(36)=7.32$，$p<0.001$；$t_{游戏成瘾维度}(36)=6.89$，$p<0.001$；$t_{上网时间}(36)=5.97$，$p<0.001$）。

2.2.2 实验材料

从《现代汉语词典》（2005）中选取140个四字成语或三字惯用语（见附录1）。在15名随机招募的被试中进行情绪效价的7点量表测评，最后确定符合情绪效价要求的材料共132个（消极词71个，积极词61个）。然后再随机选取15名被试对这132个词是否和面部表情动作有关进行7点量表评定，最终选取相关程度均分在5.80以上的80个成语和惯用语（消极和积极词各40个）作为实验材料，且在与面部表情动作的相关程度上没有显著差异（$F(1,78)=0.03$，$p=0.87$），而在情绪效价上有显著差异（$F(1,78)=1585.10$，$p<0.001$）。

2.2.3 实验设计

采用2（被试类别：网络使用经验少 vs. 网络使用经验多）× 2（面部操纵状态：促进积极情绪 vs. 抑制积极情绪）× 2（词的情绪效价：积极 vs. 消极）三因素混合设计。其中被试类别为被试间变量，词的情绪效价和面部操纵状态为被试内变量。因变量为被试判断刺激情绪效价的反应时和正确率。

2.2.4 实验程序

采用E-Prime2.0软件对实验设计进行编程，整个实验流程在电脑上进行。正式实验中，按照对被试面部肌肉的不同操纵方法分为两个反应时段，一个为促进积极情绪时段，另一个为抑制积极情绪时段。促进积极情绪的实现手段是，要求被试用门牙紧紧咬住一支筷子，且在这个过程中不能让嘴唇碰触到

筷子（见图1左图）；抑制积极情绪的实现手段是，要求被试用嘴唇紧紧固定住一支筷子，且在这个过程中不能让牙齿碰触到筷子（见图1右图）（Niedenthal, 2007; Strack et al., 1988）。要求被试在实验进行中时刻保持正确的咬筷子方式，并明确告知被试有录像设备对他们是否正确操作进行监控。这两种控制条件的先后顺序在被试间做了平衡处理。被试的任务

图1 抑制积极情绪和促进积极情绪动作模拟的示意图（摘选自 Niedenthal, 2007)

是对所呈现的词汇进行积极和消极情绪效价的判断。一半的被试用左手按"积极"键，右手按"消极"键，另一半则相反。

每个反应时段由练习和正式实验组成。正式实验中每个试次的具体流程如下：首先在电脑屏幕正中呈现一个红色的"+"注视点，500ms后注视点消失，随即在相同位置呈现一个词语，时长为800ms。被试需在准确的基础上尽量快的做出情绪效价的判断。反应后空屏时长为1000ms，随即下一个试次开始。每个反应时段有80个试次，且两个反应时段中实验材料相同。刺激在每个反应时段内随机呈现，且每个反应时段内被试休息两次，每次5min。被试在休息时，安排被试完成一份无关问卷，以削弱被试对上一个时段刺激的记忆。整个实验一共160个试次，共耗时15min左右。

2.3 实验结果

38名被试中有2名被试的数据因正确率低于80%而被剔除，最终有36份有效数据进入统计分析。其中网络使用经验多的有19名，网络使用经

验少的有 17 名。运用 SPSS17.0 对有效数据进行三因素重复测量的方差分析，发现被试类别、面部操纵状态和词的情绪效价三因素交互作用显著（F（1,34）=4.53, p=0.04, η_p^2=0.12）。进一步进行简单交互作用分析发现，被试类别和面部操纵状态的交互作用在积极词上显著（F（1,34）=4.63, p=0.039, η_p^2=0.12），而在消极词上不显著（F（1,34）=0.03, p=0.88, η_p^2=0.001）。进一步进行简单简单效应分析发现，网络使用经验少的被试在促进和抑制积极情绪状态下对积极词的反应显著不同（F（1,34）=7.59, p=0.009, η_p^2=0.18）。而网络使用经验多的被试则无显著差异（F（1,34）=0.05, p=0.83, η_p^2=0.001）。同时，被试类别的主效应显著（F（1,34）=5.16, p=0.03, η_p^2=0.13），即网络使用经验少的被试对词
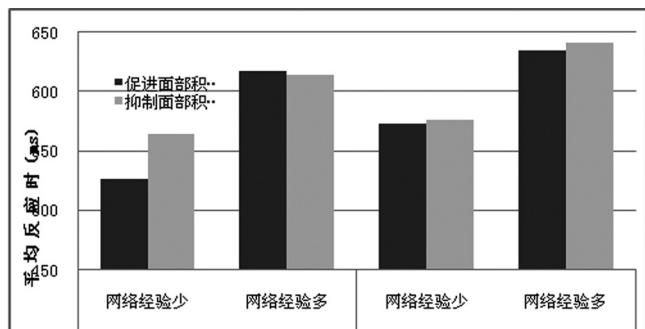
图 2　两类被试群体在不同面部状态下的平均反应时

汇的反应显著快于网络使用经验多的被试。词的情绪效价主效应显著（F（1,32）=19.04, p<0.001, η_p^2=0.36），即所有被试对积极词的反应要快于对消极词的反应。

2.4　讨论

实验 1 的结果表明网络使用经验多的被试对面部动作词汇的反应与网络使用经验少的被试有显著不同。前者在积极情绪面部表达受促进或受抑制情况下，对词汇的判断无差别；而后者则有显著不同，实验 1 的假设得到了证实。但两组被试只在积极情绪词汇的判断上表现出差异，在消极情绪的词汇判断中则未表现出不同，表明积极情绪概念的加工似乎比消极情绪概念的加工更易受具身经验减少的影

响，但也有可能是在本实验中被试受到操纵的只是积极情绪的表达，而不是消极情绪的表达。与此同时，网络使用经验多的被试对所有词汇的总体反应都要比网络使用经验少的被试要慢，表明前者整体的概念加工已不如后者。

3　实验 2：肢体具身体验的减少对肢体动作概念加工的影响

3.1　实验目的

网络使用经验有可能不仅影响面部表情词汇的加工，还对肢体动作词汇加工有影响。实验 2 比较网络使用经验不同被试在加工肢体动作动词时是否有所不同。实验 2 假设，网络使用经验多的被试由于在感知觉层面的动作表征减弱，因此会比网络使用经验少的被试在词类转换时产生更少的认知转换消耗。

3.2　实验方法

3.2.1　被试

从被试备用库中随机选取低分组 20 名（男生 5 名，女生 15 名）作为网络使用经验少的被试，高分组 19 名（男生 8 名，女生 11 名）作为网络使用经验多的被试。被试的年龄范围在 19 岁 ~25 岁之间（M=22.69，SD=2.12）。经独立样本 t 检验，网络使用经验多的被试群体在总分、人际健康维度、网络信息成瘾维度、网络游戏成瘾维度和日均上网时间上的平均得分均显著高于网络使用经验少的被试群体（$t_{总分}$（37）=16.22, p<0.001; $t_{健康维度}$（37）=9.81, p<0.001; $t_{信息成瘾维度}$（37）=6.51, p<0.001; $t_{游戏成瘾维度}$（37）=6.85, p<0.001; t（37）=6.25, p<0.001）。

3.2.2　实验材料

从《现代汉语常用词词频词典》（刘源等，1990）中选取 140 个双字词，并随机招募 15 名被试对所选词汇属于动词还是名词进行 3 点量表评定，最后选取符合要求的词汇 60 个（动词、名词各 30 个）。根据 Pecher 等人（2003）实验中对刺激的配对方法对这 60 个词进行配对，具体方法为：从 60 个词里选择 10 个动词和 10 个名词作为目标词，使其出现在配对

词汇中后面的位置，剩下的 40 个词汇作为启动词汇出现在配对词汇中前面的位置。当目标词为动词（如：抽打）时，从启动词汇中选取动词（如：按摩）与之配对成具有相同加工通道（均为动词）的词对（如：按摩－抽打），或选取名词（如：戒指）与之配对成具有不同通道的词对（如：戒指－抽打）。用同样的方法另外组成目标词为名词的相同通道的词对和不同通道的词对。每个目标词会得到两次配对的机会，最后一共配成 40 个真词－真词对作为实验二的靶刺激。所有词对的词频两两差异不显著（$F_{(2,57)}$=0.04，p=0.96），且靶刺激中所有名词和所有动词的平均词频差异不显著（$F_{(1,58)}$=0.001，p=0.98）。最后另外准备 80 个真词和 160 个假词随机配成真词－假词、假词－假词、假词－真词各 40 对作为填充刺激，由此组成 160 个词对作为实验二的实验材料。

3.2.3 实验设计

采用 2（被试类别：网络使用经验少 vs. 网络使用经验多）× 2（转换方式：相同通道 vs. 不同通道）两因素混合设计，其中转换方式为被试内变量。因变量分为两类：一类是被试对启动词和目标词均正确反应条件下对目标词的反应时和正确率，用于分析两个被试群体的转换消耗效应的大小；另一类是被试对所有动词的反应时和正确率，用于分析两个被试群体对动词加工是否存在差异。

3.2.4 实验程序

采用 E-Prime2.0 软件对实验设计进行编程，整个实验流程在电脑上进行。实验的单个试次流程如下：首先在电脑屏幕中间呈现一个红色的"＋"注视点，500ms 后注视点消失，随即在相同的位置上呈现一个双字词，时长 500ms。要求被试在准确的基础上尽快地按键判断呈现的词是真词还是假词。在指导语中，我们给被试指出"真词是指符合语义、常见的词语，如"挠头、衣服"等；假词则指不符合语义、本来不存在的词语，如"片和、灯度"等。"。一个试次结束后有 1000ms 的空屏时间。整个实验由 160 对词对、320 个反应试次组成。被试对所呈现的刺激进行真词和假词的判断。整个实验分两个时段完成，所有实验材料被均分成两份并以预先配好的词对顺序伪随机逐个呈现。即词对中的两个词分别先后呈现，但词对的呈现顺序则完全随机。两个时段间被试休息 5min。整个实验需要 15min 左右。

3.3 实验结果

首先分析两类被试群体是否均出现了转换消耗效应，如果都出现了，两者是否有显著差异。由于所有刺激均以配对假随机形式呈现，而配对刺激中的第一个刺激被认为是类似启动的一种刺激，在计算转换消耗效应时，被试对其的反应并不被考虑。每个配对中的第二个刺激为靶刺激，是计算转换消耗效应时的关键刺激。比较被试在相同通道条件下对靶刺激的平均反应时与不同通道条件下对靶刺激的平均反应时是否有显著差异，从而确定是否产生了转换消耗效应。

本实验 39 名被试中有 3 名被试的数据因正确率低于 80% 而被剔除，最终有 36 份有效数据进入统计分析，其中两类被试各 18 名。运用 SPSS17.0 对有效数据进行两因素重复测量方差分析，发现被试类别和转换方式两因素交互作用显著，$F_{(1,34)}$=4.31，p=0.05，η_p^2=0.11。进一步进行简单效应分析，发现网络使用经验少的被试在相同通道条件下对目标词的反应比不同通道条件下显著快些（$F_{(1,34)}$=18.38，p<0.001，η_p^2=0.35），表明出现了转换消耗效应；而
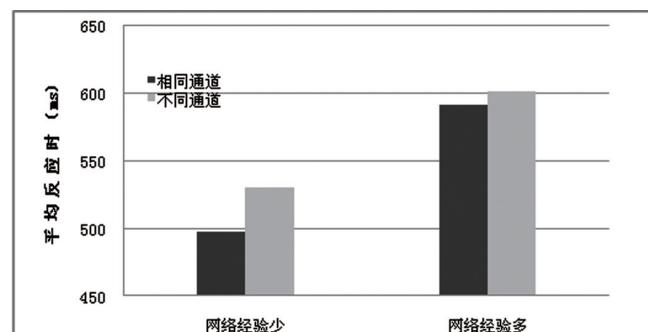
图 3 两类被试群体在两种转换方式下对目标词的平均反应时

网络使用经验多的被试在两种不同通道条件下对目标词的反应没有显著差异（$F_{(1,34)}$=1.83，p=0.19，η_p^2=0.05），表明没有出现转换消耗效应。

然后分析两类被试群体对动词加工是否存在整

体上的差异。所有 39 名被试的数据均为有效数据，且正确率均高于 80%。采用 SPSS17.0 对两类被试群体加工动词的反应时进行独立样本 t 检验，发现被试类别的主效应显著（$t_{(37)}=3.16$, $p=0.003$），即网络使用经验少的被试对动词的反应要快于网络使用经验多的被试。

3.4　讨论

实验二的结果表明两类被试群体对肢体动作动词的反应有显著不同，网络使用经验多的被试没有表现出应有的转换消耗效应，实验二的假设得到验证，即网络使用经验多的被试在加工动词和名词时，表现出更多的相似性，导致其认知转换消耗减少。根据具身认知理论，由于在肢体动作动词的加工过程中，个体需要激活先前的运动经验，而网络使用经验多的被试其肢体运动具身体验的感知觉表征有可能减弱了，从而导致他们在肢体动作词汇加工过程中对肢体运动系统的再激活变得困难，使得其加工更趋同于名词。但这一结论还有待更多的研究加以证实。

实验二还发现网络使用经验多的被试从总体上来说对动词加工要比网络使用经验少的被试要慢，这表明尽管网络使用经验多的群体还未达到病理性的网络成瘾状态，但由于长期的网络使用经验，其高级动作概念的加工已经产生了异化。

4　实验 3：肢体具身体验的减少对动作图式加工的影响

4.1　实验目的

前面两个实验探讨了网络使用不同经验者在加工单个动作概念时的差异。实验三将比较这两类群体在加工一系列有内在逻辑关联的动词时的不同表现。实验三采用图式理论中动作图式概念（action schema）。图式（schema）是从过去经验中抽取出来的重要且稳定的元素，是一种高水平、有组织的概念表征方式（Barsalou, 2000）。动作图式是图式的一种，是个体在观察和不断练习的基础上，

在大脑中形成的一种概括化的动作结构（Galotti, 2009）。动作图式中的动作结构的核心成分是由动词组成的、逻辑上归属于同一个事件的若干动词（短语）。实验三假设，网络使用经验多的被试对以正常顺序呈现和以随机顺序呈现的系列动词的系列回忆成绩没有显著差异，而网络使用经验少的被试则会受打乱顺序的影响。

4.2　实验方法

4.2.1　被试

由于被试备用库中有部分被试流失，因此实验三依据实验一筛选被试的方法补充了一些被试，形成低分组 24 名（男生 6 名，女生 18 名）作为网络使用经验少的被试，高分组 23 名（男生 12 名，女生 11 名）作为网络使用经验多的被试。被试的年龄范围在 19 岁 ~25 岁之间（$M=22.62$，$SD=2.10$）。经独立样本 t 检验，网络使用经验多的被试群体在总分、人际健康维度、网络信息成瘾维度、网络游戏成瘾维度和日平均上网时间上的平均得分均显著高于网络使用经验少的被试群体（$t_{总分(45)}=16.12$, $p<0.001$; $t_{健康维度(45)}=10.38$, $p<0.001$; $t_{信息成瘾维度(45)}=7.89$, $p<0.001$; $t_{游戏成瘾维度(37)}=7.25$, $p<0.001$; $t_{上网时间(45)}=5.44$, $p<0.001$）。

4.2.2　实验材料

正式实验前准备 22 个情景事件（见附件 3），让随机招募到的 15 名被试根据这些情景写出最能表示这一情景的 6-8 个符合逻辑的连续动词（即动作图式），每个词不要超过 4 个字。如在描述"去餐馆吃饭"情景中，与之相匹配的动作图式可以是：开门、看菜单、点餐、吃东西、付账、离开。问卷有效回收率为 100%。对被试所列的词语进行频次统计，把每个事件中的动词系列按频次从高到低进行排列，选出事件中排在前五位的动词作为动作图式中的代表性动词，作为本实验的材料。

符合本实验要求的动作图式共准备了 16 个，每个图式中的系列动词有 5 个。16 个图式中的系列动词分为正常序列和随机序列两种，共组成 32 个动词序列。每个被试所接受的随机序列动词系列无

论在项目上还是在序列上都是一样的。将这 32 个动词序列按照拉丁方的方法分配给被试，使每个被试只接受一个动作图式中的一种序列。每个被试共接受 16 组动作序列，8 组正常序列，8 组随机序列。这 16 组正常与随机序列在每个被试内的呈现顺序完全随机。

4.2.3 实验设计

采用 2（被试类别：网络使用经验少 vs. 网络使用经验多）× 2（动词序列类型：正常序列 vs. 随机序列）两因素混合设计，其中网络使用经验为被试间变量。因变量为被试对每个词汇回忆的项目正确率和项目位置正确率。

4.2.4 实验程序

采用 E-Prime2.0 软件对实验设计进行编程，整个实验流程在电脑上进行。正式实验中每个试次的具体流程如下：首先在电脑屏幕中间呈现一个红色的"+"注视点，500ms 后，屏幕中央依次呈现 5 个词汇，每个词汇呈现 1000ms，ISI 为 500ms。词汇呈现完毕后，屏幕上呈现"请按词语呈现的先后顺序依次作答"字样，提醒被试马上在答题纸上由左至右按词汇的呈现顺序尽量准确和快地写出刚才呈现的词语。作答完毕后按空格键进入下一个试次。实验共有 16 个试次，做完 8 个试次后被试休息几分钟。整个实验持续 15min 左右。

4.3 实验结果

本实验的 47 名被试中有 1 名被试由于操作不当数据被剔除，最终有 46 份有效数据进入统计分析，其中两类被试各有 23 名。采用 SPSS17.0 对项目正确率（只要项目被正确回忆出来即视为正确）、项目和位置正确率（项目和位置均被正确回忆出来才视为正确）进行两因素重复测量的方差分析发现，被试类别和动词序列类型的两因素交互作用显著（$F_{项目正确}(1,44)=5.276$, $p_{项目正确}=0.026$, $\eta^2_{p 项目正确}=0.107$; $F_{项目和位置正确}(1,44)=5.57$, $p_{项目和位置正确}=0.02$, $\eta^2_{p 项目和位置正确}=0.11$）。进一步进行简单效应分析得出，网络使用经验少的被试对随机序列动词的项目正确率显著低于正常序列动词的项目正确率（$F_{项目正确}(1,44)$

$=12.98$, $p_{项目正确}=0.001$, $\eta^2_{p 项目正确}=0.22$; $F_{项目和位置正}$
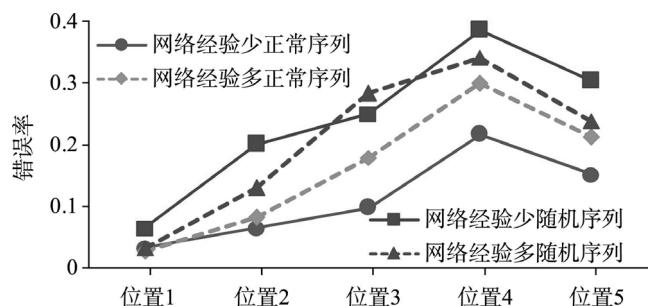
图 4　两类被试群体的项目正确回忆的系列位置曲线
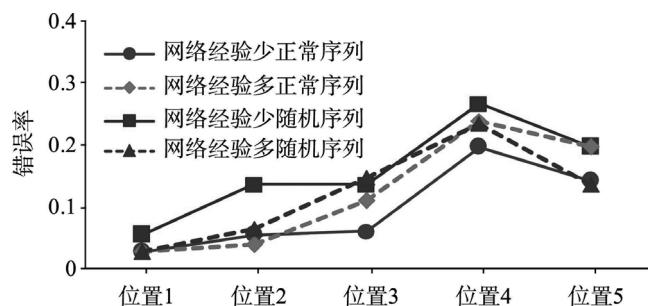
图 5　两类被试群体的项目和位置正确回忆的系列位置曲线

$确(1,44)=26.60$, $p_{项目和位置正确}<0.001$, $\eta^2_{p 项目和位置正确}=0.37$），而网络使用经验多的被试则没有表现出不同（$F_{项目正确}(1,44)=0.08$, $p_{项目正确}=0.78$, $\eta^2_{p 项目正确}=0.002$; $F_{项目和位置正确}(1,44)=3.31$, $p_{项目和位置正确}=0.08$, $\eta^2_{p 项目和位置正确}=0.06$）。

4.4 讨论

实验三发现网络使用经验多的被试在对动作图式中的系列动词作系列回忆时，无论是对动词本身还是对系列动词中的序列信息，都表现出与网络使用经验少的被试有显著不同。网络使用经验多的被试对以正常顺序和以随机序列呈现的系列动词没有表现出回忆成绩的不同，而网络使用经验少的被试则在这两种条件下表现出了明显差异。根据具身认知理论，网络使用经验多的被试很可能由于长期缺乏足够的具身动作体验，对动作图式中的动作及其顺序表征已经不够敏感，从而在复杂概念的表征与加工上产生异化，但具体原因还有待进一步深入的研究。

5　总讨论

有研究表明网络环境和使用的特殊性及人们对网络本身的迷恋程度对人类尤其是青少年的身心健康造成了严重损伤，甚至改变了相应的大脑神经加工机制。如网络经验越多，被试体验到的孤独感和抑郁感越强，体验到的社会支持度也越低（Kang, 2007）；网络成瘾者在体验离身刺激时，所激活的脑区与精神分裂、抑郁症患者特有的脑激活区一致，且网瘾时间越长，这些变异性的脑区激活程度越高（Kim et al., 2012）。本研究正是在这样一个大的社会环境背景之下，从具身认知理论出发，探讨网络使用经验的不同程度与个体加工高级认知概念之间的关系。

5.1　网络使用经验对单个动作词汇加工的影响

根据具身认知理论，个体在加工动作词汇时，会再激活大脑中的感知运动系统，对动词描述的动作进行模拟。本研究中的实验一和实验二采用表示面部表情或肢体动作的动词作为刺激，来测试具有不同网络使用经验的被试群体是否有不同的行为反应。这两个实验的结果均验证了我们的假设，即网络使用经验多的人无论是对面部表情动作词汇还是对肢体动作词汇，其加工方式都与网络使用经验少的有所不同。分析其原因，我们认为这很可能是由于网络使用经验多的个体由于将更多的时间投入到网络上，且在一定程度上对网络产生了心理依赖，使得他们对面部表情的体验和肢体运动的体验严重减弱，导致低级的感知觉系统的再激活有可能受到阻碍，从而在相应的概念加工过程中产生异化。本研究的结果进一步证实了具身感知觉经验的不同与高级认知活动之间的密切关系，支持了具身认知理论。但网络使用经验多的群体与高级认知加工之间关系的实质还有待进一步研究。

实验二的结果进一步验证了具身理论的观点。前人对失语症病人的研究发现不同部位的大脑皮层区受损会影响名词和动词的加工，暗示名词和动词的神经表征系统是相对独立的 (Shapiro & Caramazza,

2003; 刘涛 等 , 2008)。因此当先后加工动词和名词时，由于不同的脑区得到激活，在认知上会产生一定损耗；而如果脑区激活模式趋同的话，则损耗会减少。实验二中发现的网络使用经验多的被试并未产生显著的转换消耗效应，表明他们对动词加工的模式已逐步趋同于名词的模式。而最值得引起注意的是，由于动词加工在语义知识理解、存储及提取的过程中都承载着重要作用（Kemmerer & Gonzalez- Castillo, 2010），因此如果这种差异的减小真的是由于对动词加工的减弱引起的，那么长期沉迷于网络的的行为有可能最终在语义知识体系建构与表征上产生本质性改变。

5.2　网络使用经验对系列动词加工的影响

本研究的实验三发现，网络使用经验多的被试对一系列具有逻辑顺序的动词的加工已产生异化。网络使用经验多的被试在刺激系列的顺序被打乱后，其系列回忆成绩并无显著变化，这表明被试对动作动词逻辑顺序的表征变弱，导致序列信息对其系列回忆成绩产生不了影响。根据身体图式（body schema）（Assaiante, Barlaam, Cignetti, &Vaugoyeau, 2014; Dijkerman & de Haan, 2007）理论，人类自出生以来就通过本体感受和视觉获取的方式，为将来与外界的交流开始建构身体各部件之间及与周围环境间的内部表征（Head & Holmes, 1911），这种表征包含了方位、速度等维度。由于身体图式主要基于身体的移动及与他人的互动所产生的感知觉信息建构起来的，因此实际的身体体验对于身体的图式表征就显得非常重要。由此，我们推断网络使用经验过多者对身体动作图式中系列动词加工产生的异化现象，很可能是由于他们对动作的具身体验减弱所导致，这一结果进一步验证了网络使用经验与高级认知加工之间的关系。

5.3　未来研究展望

虽然本研究针对不同程度网络使用经验的被试设计了三个实证实验，验证了网络使用经验的多少与动作动词加工之间的关系。但由于我们在实验过程中并未操纵被试的网络使用行为，所以从本质上来

说本研究是一个相关研究，我们还不清楚最终导致两组被试在动作动词加工上显著差异的是网络使用经验本身，还是网络的长期使用引起了其它认知能力的变化所导致。因此，未来的研究除了进一步操纵和控制个体网络使用的某些变量，以进一步探讨网络使用中的哪些行为是引起高级认知功能变差的原因，还应该找到相应的预防措施，以防止人类在无法避免的网络接触过程中将面临的认知能力的逐步衰退。

6 结论

本研究从具身认知理论的角度，探讨了网络使用经验与汉语动词加工间的关系。研究的结果表明当前的网络行为与高级认知加工有着紧密关联，至少在动词加工方面表现在：1）网络使用经验多的被试其面部肌肉动作的内部表征可能已经弱化；2）网络使用经验多的被试其肢体动作的内部表征可能已经弱化；3）网络使用经验多的被试其身体图式的内部表征可能已经弱化。

参考文献

Ackerman, J. M., Nocera, C. C., & Bargh, J. A. (2010). Incidental haptic sensations influence social judgments and decisions. *Science*, *328*(5986), 1712–1715.

Alessi, N. (2001). Disembodiment in cyberspace is not a myth. *CyberPsychology & Behavior*, *4*(4), 537–538.

Andersen, S. M., Reznik, I., & Manzella, L. M. (1996). Eliciting facial affect, motivation, and expectancies in transference: Significant–other representations in social relations. *Journal of Personality and Social Psychology*, *71*(6), 1108–1129.

Assaiante, C., Barlaam, F., Cignetti, F., & Vaugoyeau, M. (2014). Body schema building during childhood and adolescence: A neurosensory approach. *Clinical Neurophysiology*, *44*, 3–12.

Bai, Y., & Fan, F. M. (2005). A study on the internet dependence of college students: The revising and applying of a measurement. *Psychological Development and Education*, *21*(4), 99–104.

[白羽, 樊富珉. (2005). 大学生网络依赖测量工具的修订和应用. *心理发展与教育*, *21*(4), 99–104.]

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*(4), 577–660.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645.

Barsalou, L. W., Kyle Simmons, W., Barbey, A. K., & Wilson, C. D. (2003). Grounding conceptual knowledge in modality–specific systems. *Trends in Cognitive Sciences*, *7*(2), 84–91.

Briñol, P. & Petty, R. E. (2008). Embodied persuasion: Fundamental processes by which bodily responses can impact attitudes. *Embodiment Grounding: Social, Cognitive, Affective, and Neuroscientific Approaches* (pp. 184–207). Cambridge: Cambridge University Press.

Buccino, G., Riggio, L., Melli, G., Binkofski, F., Gallese, V., & Rizzolatti, G. (2005). Listening to action–related sentences modulates the activity of the motor system: A combined TMS and behavioral study. *Cognitive Brain Research*, *24*(3), 355–363.

Cacioppo, J. T., Priester, J. R., & Berntson, G. G. (1993). Rudimentary determinants of attitudes: II. Arm flexion and extension have differential effects on attitudes. *Journal of Personality and Social Psychology*, *65*(1), 5–17.

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right–and left–handers. *Journal of Experimental Psychology – General, 138*(3), 351–367.

Casasanto, D. & Henetz, T. (2012). Handedness shapes children's abstract concepts. *Cognitive Science*, *36*(2), 359–372.

Dai, S. Y., Ma, Q. G., & Wang, X. Y. (2011). Attentional

bias to addiction-related stimuli in internet addiction patients: An ERP study. *Journal of Psychological science*,*34* (6), 1302–1307.

[戴砷懿 , 马庆国 , 王小毅 . (2011). 网络游戏成瘾者对成瘾相关线索的注意偏向 : 一项 ERP 研究 . *心理科学* , *34*(6), 1302–1307.]

Damasio, A. R. (1989). Time-locked multiregional retroactivation: A systems-level proposal for the neural substrates of recall and recognition. *Cognition*, *33*(1–2), 25–62.

De Zubicaray, G., Postle, N., McMahon, K., Meredith, M., & Ashton, R. (2010). Mirror neurons, the representation of word meaning, and the foot of the third left frontal convolution. *Brain and Language*, *112*(1), 77–84.

Dictionary Department of the Institute of Linguistics of the Chinese Academy of Social Sciences (Ed.). (2005). *Modern Chinese Words Dictionary* (5th Ed.). Beijing, China: Commercial Press.

[中国社会科学院语言研究所辞典编辑室(编). (2005). *现代汉语词典* (第五版). 北京：商务印书馆 .]

Dijkerman, H. C. & de Haan, E. H. (2007). Somatosensory processes subserving perception and action. *Behavioral Brain Science*, *30*(2): 189–201.

Duclos, S. E., Laird, J. D., Schneider, E., Sexter, M., Stern, L., & Van Lighten, O. (1989). Emotion-specific effects of facial expressions and postures on emotional experience. *Journal of Personality and Social Psychology*, *57*(1), 100–108.

Fay, A. J. & Maner, J. K. (2012). Warmth, spatial proximity, and social attachment: The embodied perception of a social metaphor. *Journal of Experimental Social Psychology*, *48*(6), 1369–1372.

Filimon, F., Nelson, J. D., Hagler, D. J., & Sereno, M. I. (2007). Human cortical representations for reaching: mirror neurons for execution, observation, and imagery. *Neuroimage*, *37*(4), 1315–1328.

Fodor, J. A. (1975). *The Language of Thought.* Cambridge, MA: Harvard University Press.

Förster, J., & Strack, F. (1998). Motor actions in retrieval of valenced information: II. Boundary conditions for motor congruence effects. *Perceptual and Motor Skills*, *86*(3 Pt 2), 1423–1426.

Gackenbach, J. (Ed.). (2011). *Psychology and the internet: Intrapersonal, interpresonal, and transpersonal implications* (2nd Ed.). Burlington, MA: Academic Press.

Galotti, K. M. (2009). *Cognitive Psychology*: *In and Out of the Laboratory*. Toronto: Cengage Learning.

Gao, W. B., & Chen, Z. Y. (2006). A Study on Psychopathology and Psychotherapy of Internet *.Advances in Psychological Science, Addiction, 14*(4), 596–603.

[高文斌 , 陈祉妍 . (2006). 网络成瘾病理心理机制及综合心理干预研究 . *心理科学进展* , *14*(4), 596–603.]

Glenberg, A. M. & Kaschak, M. P. (2002). Grounding language in action. *Psychonomic Bulletin & Review*, *9*(3), 558–565.

Goldberg, R. F., Perfetti, C. A., & Schneider, W. (2006). Perceptual knowledge retrieval activates sensory brain regions. *The Journal of Neuroscience*, *26*(18), 4917–4921.

Halberstadt, J., Winkielman, P., Niedenthal, P. M., & Dalle, N. (2009). Emotional conception: How embodied emotion concepts guide perception and facial action. *Psychological Science*, *20*(10), 1254–1261.

Hanlon, J. (2001). Disembodied intimacies: Identity and relationship on the Internet. *Psychoanalytic Psychology*, *18*(3), 566–571.

Hauk, O., Johnsrude, I., & Pulvermüller, F. (2004). Somatotopic representation of action words in human motor and premotor cortex. *Neuron*, *41*(2), 301–307.

Hauk, O. & Pulvermüller, F. (2004). Neurophysiological distinction of action words in the fronto-central cortex.

Human Brain Mapping, 21(3), 191–201.

Hauk, O., Shtyrov, Y., & Pulverm ü ller, F. (2008). The time course of action and action–word comprehension in the human brain as revealed by neurophysiology. *Journal of Physiology–Paris, 102* (1–3), 50–58.

Havas, D. A., Glenberg, A. M., Gutowski, K. A., Lucarelli, M. J., & Davidson, R. J. (2010). Cosmetic use of botulinum toxin–A affects processing of emotional language. *Psychological Science*, *21*(7), 895–900.

He, J. B., Guo, Y. Y., Ke, S. Y., & Zhao, L. (2008). Cognition deficit in internet game addicts: an auditory oddball P300 study. *Journal of Psychological science*, *31*(2), 380–384

[贺金波 , 郭永玉 , 柯善玉 , 赵仑 . (2008). 网络游戏成瘾者认知功能损害的 ERP 研究 . *心理科学* , *31*(2), 380–384.]

He, J. B., Guo, Y, Y., & Xiang, Y. M. (2008) Forming mechanism of Adolescents' Internet– game Addiction. *Chinese Journal of Clinical Psychology*, *16*(1), 46–48.

[贺金波 , 郭永玉 , 向远明 . (2008). 青少年网络游戏成瘾的发生机制 . *中国临床心理学杂志* , *16*(1), 46–48.]

Head, H. & Holmes, G. (1911). Sensory disturbances from cerebral lesion. *Brain, 34*:102–254.

Huang, M., Qian, R. B., Fu, X. M., Wang, C. X., Liu, Y., Han, X. P. ⋯ Wang, Y. H. (2010). Location of related brain activation in internet game addicts: an fMRI study. *Chin J Neurome*, *9*(2), 167–171

[黄敏 , 钱若兵 , 傅先明 , 王昌新 , 刘影 , 韩晓鹏 , ⋯ 汪业汉 . (2010). 网络游戏成瘾者相关脑区功能定位的 fMRI 研究 . *中华神经医学杂志* , *9*(2), 167–171.]

IJzerman, H., Karremans, J.C., Thomsen, L., & Schubert, T. W. (2013). Caring for sharing. *Social Psychology*, *44*(2), 160–166.

IJzerman, H. & Semin, G.R. (2009). The thermometer of social relations: Mapping social proximity on

temperature. *Psychological Science*, *20*(10), 1214–1220.

Jin, P., Fu, X. M., Qian, R. B., Niu, C. S., & Han, X. P. (2009). Event–related potentials N400 in adolescents with internet addiction disorder. *Chin J Stereotact Funct Neurosurg*, *21*(6), 333–335

[金璞 , 傅先明 , 钱若兵 , 牛朝诗 , 韩晓鹏 . (2009). 青少年网络成瘾的事件相关电位 N400 研究 . *立体定向和功能性神经外科杂志* , *21*(6), 333–335.]

Jostmann, N. B., Lakens, D., & Schubert, T. W. (2009). Weight as an embodiment of importance. *Psychological Science*, *20*(9), 1169–1174.

Kang, S. (2007). Disembodiment in online social interaction: Impact of online chat on social support and psychosocial well–being. *CyberPsychology & Behavior*, *10*(3), 475–477.

Kellenbach, M., Brett, M., & Patterson, K. (2001). Large, colorful, or noisy? attribute– and modality–specific activations during retrieval of perceptual attribute knowledge. *Cognitive, Affective, & Behavioral Neuroscience*, *1*(3), 207–221.

Kemmerer, D. & Gonzalez–Castillo, J. (2010). The two-level theory of verb meaning: an approach to integrating the semantics of action with the mirror neuron system. *Brain and Language*, *112*(1), 54–76.

Kim, Y. R., Son, J. W., Lee, S. I., Shin, C. J., Kim, S. K., Ju, G., ⋯ Jo, S. (2012). Abnormal brain activation of adolescent internet addict in a ball–throwing animation task: possible neural correlates of disembodiment revealed by fMRI. *Progress in Neuro–Psychopharmacology and Biological Psychiatry*, *39*(1), 88–95.

LaFrance, M., & Hecht, M. A. (2000). Gender and smiling: a meta–analysis. In *Gender and Emotion*: *Social Psychological Perspectives* (pp. 118–142). New York: Cambridge University Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*:

The Embodied Mind and Its Challenge to Western Thought. New York: Basic Books.

Larsen, R. J., Kasimatis, M., & Frey, K. (1992). Facilitating the furrowed brow: An unobtrusive test of the facial feedback hypothesis applied to unpleasant affect. *Cognition & Emotion*, *6*(5), 321–338.

Liang, S. C. & You, Y. Q. (2010) Affective Decision–making in Patients with Internet Addiction, *Chinese Journal of Clinical Psychology*, *18*(5), 597–599.

[梁三才 , 游旭群 . (2010). 网络成瘾者情感决策能力的对照研究 . *中国临床心理学杂志* , *18*(5), 597–599.]

Lindemann, O., Stenneken, P., Van Schie, H. T., & Bekkering, H. (2006). Semantic activation in action planning. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(3), 633–643.

Liu, T., Yang, Y. M., Zhang, H., Zhang, S. S., Liang, D. D., Gu, J. X., & Hu, W. (2001). Neural distinction between Chinese nouns and verbs in the grammatical context: an ERP study. *Acta Psychologica Sinica*, *40*(6), 671–680.

[刘涛 , 杨亦鸣 , 张辉 , 张珊珊 , 梁丹丹 , 顾介鑫 , 胡伟 . (2008). 语法语境下汉语名动分离的 ERP 研究 . *心理学报* , *40*(6), 671–680.]

Liu, Y. (Ed.). (1990). *Modern Chinese Frequency Dictionary of common words*. Beijing, China: Aerospace Press.

[刘源 (编). (1990). *现代汉语常用词词频词典* . 北京 : 宇航出版社 .]

Marques, J. F. (2006). Specialization and semantic organization: evidence for multiple semantics linked to sensory modalities. *Memory & Cognition*, *34*(1), 60–67.

Näsi, M. & Koivusilta, L. (2013). Internet and everyday life: the perceived implications of internet use on memory and ability to concentrate. *Cyberpsychology, Behavior, and Social Netwroking*, *16*(2), 88–93.

Neal, D. T. & Chartrand, T. L. (2011). Embodied emotion perception amplifying and dampening facial feedback modulates emotion perception accuracy. *Social Psychological and Personality Science*, *2*(6), 673–678.

Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*(5827), 1002–1005.

Niu, G., Sun, X., Zhou, Z., & Wei, H. (2013). A review of cognitive neuroscience studies on internet addiction. *Advances in Psychological Science*, *21*(6), 1104–1111.

[牛更枫 , 孙晓军 , 周宗奎 , 魏华 . (2013). 网络成瘾的认知神经科学研究述评 . *心理科学进展* , *21*(6), 1104–1111.]

Oberman, L. M., Winkielman, P., & Ramachandran, V. S. (2007). Face to face: Blocking facial mimicry can selectively impair recognition of emotional expressions. *Social Neuroscience*, *2*(3–4), 167–178.

Oosterwijk, S., Rotteveel, M., Fischer, A.H., & Hess, U. (2009). Embodied emotion concepts: How generating words about pride and disappointment influences posture. *European Journal of Social Psychology*, *39*(3), 457–466.

Pawlikowski, M. & Brand, M. (2011). Excessive internet gaming and decision making: Do excessive World of Warcraft players have problems in decision making under risky conditions? *Psychiatry Research*, *188*(3), 428–433.

Pecher, D., Zeelenberg, R., & Barsalou, L.W. (2003). Verifying different–modality properties for concepts produces switching costs. *Psychological Science*, *14*(2), 119–124.

Ponari, M., Conson, M., D'Amico, N. P., Grossi, D., & Trojano, L. (2012). Mapping correspondence between facial mimicry and emotion recognition in healthy subjects. *Emotion, 12*(6), 1398–1403.

Price, T. F., Peterson, C. K., & Harmon–Jones, E. (2012). The emotive neuroscience of embodiment. *Motivation and Emotion, 36*(1), 27–37.

Pulvermüller, F., Hauk, O., Nikulin, V. V., & Ilmoniemi, R. J. (2005). Functional links between motor and language systems. *European Journal of Neuroscience*, *21*(3), 793–797.

Shapiro, K. & Caramazza, A. (2003). Grammatical processing of nouns and verbs in left frontal cortex? *Neuropsychologia*, *41*(9), 1189–1198.

Simmons, W. K., Ramjee, V., Beauchamp, M. S., McRae, K., Martin, A., & Barsalou, L. W. (2007). A common neural substrate for perceiving and knowing about color. *Neuropsychologia*, *45*(12), 2802–2810.

Snodgrass, J. G. (1984). Concepts and their surface representations. *Journal of Verbal Learning and Verbal Behavior*, *23*(1), 3–24.

Stel, M., & van Knippenberg, A. (2008). The role of facial mimicry in the recognition of affect. *Psychological Science*, *19*(10), 984–985.

Strack, F., Martin, L. L., & Stepper, S. (1988). Inhibiting and facilitating conditions of the human smile: A nonobtrusive test of the facial feedback hypothesis. *Journal of Personality and Social Psychology*, *54*(5), 768–777.

Sun, D. L., Chen, Z. J., Ma, N., Zhang, X. C., Fu, X. M., & Zhang, D. R. (2009). Decision–making and prepotent response inhibition functions in excessive internet users. *CNS Spectr, 14*(2), 75–81.

Tan, J. W., Walter, S., Scheck, A., Hrabal, D., Hoffmann, H., Kessler, H., & Traue, H. C. (2012). Repeatability of facial electromyography (EMG) activity over corrugator supercilii and zygomaticus major on differentiating various emotions. *Journal of Ambient Intelligence and Humanized Computing*, *3*(1), 3–10.

Tanenhaus, M. K., Carlson, G., & Trueswell, J. C. (1989). The role of thematic structures in interpretation and parsing. *Language and Cognitive Processes*, *4*(3–4), SI211–SI234.

Taylor, C. A., Lord, C. G., & Bond, C. F. (2009).

Embodiment, agency, and attitude change. *Journal of Personality and Social Psychology*, *97*(6), 946–962.

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion, 7*(3), 487–494.

Vanman, E. J., Paul, B., Ito, T., & Miller, N. (1997). The modem face of prejudice and structural features that moderate the effect of cooperation on affect. *Journal of Personality and Social Psychology*, *73*(5), 941–959.

Vermeulen, N., Corneille, O., & Niedenthal, P. M. (2008). Sensory load incurs conceptual processing costs. *Cognition, 109*(2), 287–294.

Vermeulen, N., Niedenthal, P. M., & Luminet, O. (2007). Switching between sensory and affective systems incurs processing costs. *Cognitive Science*, *31*(1), 183–192.

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology*, *28*(6), 879–896.

Weinstein, A., & Lejoyeux, M. (2010). Internet addiction or excessive internet use. *The American Journal of Drug and Alcohol Abuse*, *36*(5), 277–283.

Wells, G. L. & Petty, R. E. (1980). The effects of over head movements on persuasion: Compatibility and incompatibility of responses. *Basic and Applied Social Psychology, 1*(3), 219–230.

Williams, J. H., Whiten, A., & Singh, T. (2004). A systematic review of action imitation in autistic spectrum disorder. *Journal of Autism and Developmental Disorders*, *34*(3), 285–299.

Williams, L. E. & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science*, *322*(5901), 606–607.

Wiswede, D., Münte, T. F., Krämer, U. M., & Rüsseler, J. (2009). Embodied emotion modulates neural signature of performance monitoring. *PLoS One, 4*(6), e5754.

Xie, J. S., Zhang, C. Q., Wang, R. M., & Lu, Z. (2011).

Paradigms in the study of perceptual symbol systems. *Advances in Psychological Science, 19*(9), 1293–1305.

[谢久书 , 张常青 , 王瑞明 , 陆直 . (2011). 知觉符号理论及其研究范式 . *心理科学进展* , 19(9), 1293–1305.]

Yu, H. Q., Zhao, X., Li, N., Wang, M. S., & Zhou, P. (2009). Effect of excessive Internet use on the time–frequency characteristic of EEG. *Progress in Natural Science*, 19(10), 1383–1387.

[郁洪强 , 汪曈 , 赵欣 , 李宁 , 刘海婴 , 王明时 . (2009). 网络成瘾患者的 EEG 小波熵与复杂度特征分析 . *中国生物医学工程学报* , 28(1), 157–160.]

Zhao, L. & Gao, W. B. (2007). Early face processing of internet addiction patients by face specific N170. *Space Medicine & Medical Engineering*. 20(1), 72–74.

[赵仑 , 高文彬 . (2007). 网络成瘾患者早期面孔加工 N170 的研究 . *航天医学与医学工程* , 20(1), 72–74.]

Zhou, Z. J. & Yang, W. J. (2007). Different types of internet addiction scale for undergraduates. *The Handbook of Psychological Assessment Scales pp.* 503–507.

[周治金 , 杨文娇 . (2007). 大学生网络成瘾类型问卷 . *心理评定量表手册*（第一版）, 503–507.]

The impact of cyber-experience on action verb processing

LIU Siyun; ZHOU Zongkui; LI Na

(Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education;

School of Psychology, Central China Normal University, Wuhan 430079 China)

Abstract　Two major different approaches have been held when researchers study higher level cognitive processes. The classic symbolic approach suggested that our higher level cognitive processing belongs to an independent system from that of lower level cognitive processes; whereas embodied cognition theory proposed that our conceptual knowledge is grounded in our sensorimotor systems and shares common neural systems with them. Up till now, a growing number of behavioral and neurological data have provided supporting evidence for embodied cognition theory. In this study, we explored whether different degree of cyber-experience would affect higher level of cognitive processing. Three experiments were designed to investigate the impacts of cyber-experience on the processing of facial expression verbs, body action verbs and action schema verbs. In Experiment 1, two experimental groups of participants were presented with facial expression verbs while their facial positive expressive capability was either facilitated or inhibited. Results showed that participants who had excessive cyber experiences showed no different performance under two different facial muscle controlling conditions, while the participants with less cyber experience recognized positive facial verbs more quickly under facilitated condition than that under inhibited condition. In Experiment 2, the switching costs paradigm was used to explore the impact of cyber-experience on body action verb processing. Results showed that the participants with excessive cyber experience did not show any cost while they switched between verbs and nouns, but the participants with less cyber experience showed significant cognitive cost while switching. In Experiment 3, the serial recall experimental paradigm was applied to explore the impact of cyber-experience on action schema verb processing. Results showed that the excessive cyber-experience participants' recall performance of the action schemas verbs was no different between logic sequence and random sequence conditions, whereas those with less cyber-experience showed significantly worse performance in random sequence condition than that in logic sequence condition. In summary, current findings suggested that excessive cyber behaviors may hurt individuals' higher level of cognitive processing, in that their verb processing may be weakened or delayed as a result of less normal conceptual representations. Our study also provided supportive evidence for the close relationship between the sensorimotor systems and the higher level of conceptual processing.

Keywords　embodied cognition; switching cost effect; schema; cyber-experience

Capacity limit of simultaneous temporal processing: How many concurrent 'clocks' in vision?

Xiaorong Cheng[1,2,3], Qi Yang [1,2,3], Yaqian Han [1,2,3], Xianfeng Ding [1,2,3], Zhao Fan [1,2,3,*]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan, China)

([2] School of Psychology, Central China Normal University, Wuhan, China)

([3] Key Laboratory of Human Development and Mental Health of Hubei Province, Wuhan, China)

Abstract A fundamental ability for human is to monitor and process multiple temporal events that occur at different spatial locations simultaneously. A great amount of studies have demonstrated simultaneous temporal processing (STP) in human and animal participants, i.e., multiple 'clocks' rather than a single 'clock'. However, to date, we still have no knowledge about the exact limitation of the STP in vision. Here we provide the first experimental measurement to this critical parameter in human vision by using two novel and complementary paradigms. The first paradigm combines merits of a temporal oddball–detection task and a capacity measurement widely used in the studies of visual working memory to quantify the capacity of STP (CSTP). The second paradigm uses a two–interval temporal comparison task with various encoded spatial locations involved in the standard temporal intervals to rule out an alternative, object individuation–based, account of CSTP, which is measured by the first paradigm. Our results of both paradigms indicate consistently that the capacity limit of simultaneous temporal processing in vision is around 3 to 4 spatial locations. Moreover, the binding of the 'local clock' and its specific location is undermined by bottom–up competition of spatial attention, indicating that the time–space binding is resource–consuming. Our finding that the capacity of STP is not constrained by the capacity of visual working memory (VWM) supports the idea that the representations of STP are likely stored and operated in units different from those of VWM. A second paradigm confirms further that the limited number of location–bound 'local clocks' are activated and maintained during a time window of several hundreds milliseconds.

Keywords Simultaneous Temporal Processing; Capacity Limit; Spatial Encoding; Vision

1 Introduction

The ability of simultaneous temporal processing can be life−critical for human. For example, a lifeguard with responsibility for a swimming pool full of children has to adopt a strategy of singling out a potential hazard by estimating how long a child has submerged. Since the children spread around the pool, it is necessary to establish individual−based temporal estimation at each child's location and monitor multiple locations simultaneously. Besides, the pool and its vicinity are full of irrelevant distractors, such as moving adults, person's shouting, laughing and water splashes. The mechanisms of how the lifeguard ignores those distractors and accomplishes the simultaneous temporal task to spot out a potential hazard are not yet fully understood.

Visual adaptation studies shed some light on the property of the simultaneous temporal processing in vision by demonstrating that the encoding of visual duration operates in a spatially localized way. For example, temporal estimation to a 10 Hz grating of 600 ms duration (tester) was significantly compressed by a 15s adaptation of an oscillating 20 Hz grating presented on the same part of the retina as the tester. This finding implicated that the multiple, independent time estimators / 'clocks' of the STP might operate in a retinotopic or spatialtopic way, i.e., a time−space binding in visual modality.

Consider the predominant theory of time perception, i.e., the pacemaker−accumulator model for a moment. This model posits that the duration encoding is accomplished by a 'clock'−like structure, consisting of a pacemaker and an accumulator. The elapsed time is psychologically represented by the arithmetic summation of pulses emitted from the pacemaker at a regular pace, which are stored and added in the accumulator. In the context of this model, the most parsimonious way to time multiple durations concurrently is to have independent pulse making and accumulating units for those 'local clocks'. This idea has been proposed, tested and confirmed in several studies of simultaneous temporal processing in both animals and humans (but see for a single clock strategy that might apply in simple tasks requiring minimum cognitive resource). Indeed, parallel timing across the visual and auditory modalities had been observed in a study with human participants by using a stop−reaction−time (stop−RT) paradigm. Similarly, a recent research, using a temporal reproduction task on human, demonstrated that two independent clocks are involved in the timing of two multi−second intervals that are presented simultaneously in the visual modality. However, it is not clear so far how many independent timers / 'clocks' are operated concurrently in vision at most. The majority literatures of simultaneous temporal processing used temporal intervals at multi−second level and did not measure the capacity of STP. In this study, we are interested in the capacity magnitude of STP at a time scale of sub−second level since previous literature had revealed that the mechanisms of visual estimations between sub−second and supra−second levels are different and the former involves more automatic processing.

Another research line indicated the necessity of spatial attention in visual temporal processing. For example, by using a dual−task paradigm, Cicchini and Morrone found a perceived temporal compression by up to 40% to sub−second temporal intervals when attention was divided spatially. This finding implied that the operation of each 'local clock' of STP might require some attentional resource. Thus it is legitimate to ask how many location−bound 'clocks' can be activated and maintained simultaneously in vision when full attention is available, i.e., the spatial capacity of simultaneous temporal processing in vision. To our knowledge, up to now, there has been no study to directly measure this important

capacity in visual modality.

The aim of the present study was to: (a) measure directly the capacity of simultaneous temporal processing (CSTP) in sub-second level and to elucidate the properties of the location-bound multiple 'clocks', including (b) whether the CSTP is constrained by other critical cognitive limitations, such as capacity of an individual's visual working memory (VWM), and (c) whether spatial attention plays a critical role in the binding between a 'local clock' and its spatial location. We addressed these questions by using two novel and complementary paradigms, i.e. a temporal oddball-detection task to measure CSTP directly and a two-interval temporal comparison task to test an alternative account of CSTP measured by the first task.

2 Experiment 1a

2.1 Materials and methods

Twenty Nine paid participants (24 female, 5 male; mean age = 21 years) from Central China Normal University (CCNU) took part in Experiment 1a & 1b. Procedures for all experiments in this study were approved by the institutional review board of CCNU. All participants in this study were right-handed, and naive to the aim of this research. All of them had normal or corrected-to-normal vision, and had no history of neurological disorders or color blindness. All experiments in this study had been approved by the Research Ethics Committee of CCNU, and participants gave their written informed consent to participate.

The displays of all the experiments were programmed at a spatial resolution of 1024*768 pixels and a refresh rate of 100Hz. The stimuli were displayed on an IIYAMA HM903DT color monitor, driven by a NVIDIA GeForce FX 5200 Graphics Adapter. Participants' responses were recorded via a keyboard connected to the PC with their head stabilized by a chin rest and a viewing distance of

57cm. The displays of all the experiments in this study were programmed in MATLAB (MathWorks Inc.) and Psychophysics Toolbox.

The first paradigm used in Experiment 1a, i.e., a temporal oddball-detection paradigm (TODP), was adapted from the well-established change detection paradigm (CDP) by replacing stationary to-be-memorized object features with dynamic to-be-differentiated temporal intervals. Here, participants were asked to detect an oddball temporal interval which might occur in one of a set of spatial locations (Figure 1A). The stimuli and task of the first paradigm are shown in Figure 1A. In a typical trial, participants first viewed a location array to know how many locations were involved in this trial and where they were. Then a dynamic period, called 'flashing array', was followed during which objects (squares) were presented at each of those locations with desynchronized timing. At an unpredictable moment, one object was presented with oddball duration, resulting in an 'oddball array'. A probe array then came out and participants were required to report whether the oddball occurred in a location circled by an outlined grey ring. For the condition of only one spatial location involved, half trials contain an oddball while the other half not and participants were required to report whether there was an oddball presented in that location.

Criteria for trials with excessive eye movements. In Experiment 1a and subsequent experiments involving TODP, eye movement was monitored throughout the experiment by using an eye tracker of Eyelink1000 (SR Research Ltd.) to avoid the usage of saccade-based strategies to scan multi-locations sequentially, e.g. participants switched their gazes rapidly between multiple locations sequentially during the detection of the oddball. The trials of excessive eye movement were operationally defined with following criterion. The average gazes of the whole trial should be
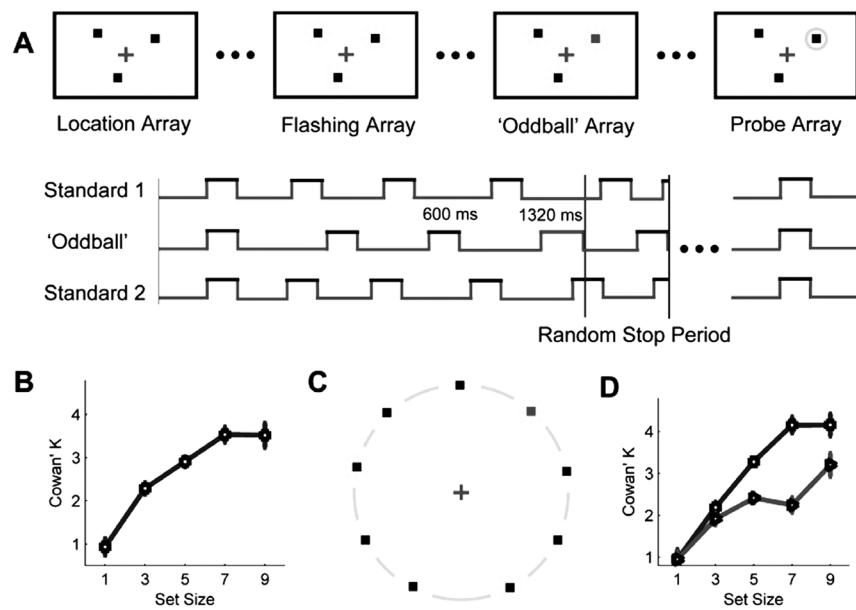
Figure 1. Experimental approach and results of Experiment 1a, 2 and 3. A, Temporal–oddball task (an exemplary trial of set size = 3; upper panel for stimuli setup, lower panel for timing traces). Red square denotes oddball and did not appear in the experimental display, so did for the yellow rectangles (denoting the repeated sequences of onsets and offsets of the squares). In the timing traces, the black/red bars denote the standard/oddball intervals and the two vertical blue lines indicate a period during which the dynamic array stops at a random time. **B**, Results of Experiment 1a (N=29, blue curve and squares). **C**, An exemplary stimulus layout of Experiment 3 of set size = 3 (with 6 stationary distractors). Experiment 2 was the same as Experiment 3 except for no stationary distractors on the display. **D**, Results of Experiment 2 (N=21, squares and blue curve) and 3 (N=20, triangles and red curve). Error bars are within–subjects SEs.
doi:10.1371/journal.poen.0091797.g001.

confined within a square area (no objects would appear inside this area by the predefined stimulus arrangement), centered at the fixation point with its boarder 3.75 deg from the fixation point. By using this criterion, averagely 6.3% trials were labeled as excessive eye movement trials and thus excluded from subsequent data analysis. The eye tracker was recalibrated after each experimental block of 20 trials.

In the basic TODP paradigm of Experiment 1a, the stimuli, 1.2 deg × 1.2 deg black squares (luminance of 5.03 cd/m^2, measured with a Minolta CS–200 Chromameter photometer), were randomly positioned within a square area, subtending 16 deg × 16 deg, against a white background (luminance of 110.02cd/m^2). The minimum distance between centers of any two squares was 2 deg and the squares never overlapped with each other. A fixation cross, 0.24 deg width and 0.07 deg thickness, was presented at the center of the display. Each border of the squares was at least 3.75 deg away from the fixation point. Participants were instructed to

fixate that cross throughout the experiment.

The standard intervals of the objects were kept constant as 600ms and the oddball interval 1320 ms, which was determined by an oddball duration experiment (see below). The blank intervals between two temporally neighbored squares were randomly sampled from 800 to 1600ms. Except for the first square onsets and offsets of all locations, which were synchronized, the timings of all other square onsets and offsets were independently produced. For any trial containing an oddball interval at a spatial location, the oddball square took either the 4th, the 5th or the 6th temporal position with equal possibilities in the square sequence of that spatial location. The oddball never appeared at either the 2nd or the 3rd temporal position, aiming to allow the desynchronizations of the timings of the squares across different spatial locations. For the oddball–present trials, a random stop procedure was used in which each trial stopped at a random time between the offset of the oddball and the offset of

its immediate next standard interval. For those trials containing no oddballs, e.g. half trials of the single location condition, similar random stop procedure was applied but modified in that the trials stopped between the offset of either the 4th, the 5th, the 6th standard interval and the offset of its immediate next standard interval. This resulted in an average trial length of 11.8 seconds and matched trial lengths between the oddball–present and oddball–absent trials. Above procedures were adopted carefully in order to avoid any consistent cuing, which might be based on the timing–relationship of multiple objects on the display and be implicitly used by participants during the oddball detection.

Five options of set size (1, 3, 5, 7 or 9 locations) were used. Each set size was repeated 40 times, including 20 'Yes' trials (an oddball appeared in the circled spatial location) and 20 'No' trials (no oddball appeared in the circled spatial location). A gray circle of 2.88 deg radius was used in the response frame to highlight one spatial location. Participants gave their response by pressing one of two response keys. The mappings between 'Yes' / 'No' responses and the two response keys were counterbalanced across participants. The whole experiment included 200 trials (10 blocks) in total and different conditions were randomly mixed in the test.

Before Experiment 1a, an oddball duration experiment was performed to establish the minimum duration that was required for individual participant to detect the oddball interval from a background of the standard intervals (600 ms). This parameter is important because it was chosen to be long enough for participants to differentiate the oddball interval from the standard interval but short enough to avoid unnecessary cuing. Twenty paid participants (16 female, 4 male; mean age = 22.15 years) from CCNU took part in the oddball duration experiment. The stimulus layout of the oddball duration experiment (Figure 2A) was similar to that of Experiment 1a with following exceptions.

Firstly, there was only one spatial location in this experiment, i.e., set size always equals to one. Secondly, in half trials, there was no oddball interval. In the other half, an oddball interval occurred, which might take one of five possible values, i.e., 650, 750, 950, 1250 or 1650 ms. Thirdly, the random stop procedure in Experiment 1a was not used here since only one spatial location was involved in this experiment and there was no cuing based on the timing–relationship of multiple objects on the display. In order to avoid that participants identify the oddball via counting, the number of squares at the single location varied from 7 to 11 with equal possibilities and the oddball square, if presented, took a temporal position randomly varied from the 2nd to the 11th in the square sequence. The raw data of the oddball duration experiment of each participant were fitted by a sigmoid curve to calculate the 95% point of correctness (Figure 2B). Considering the fact that participants might have low–level response errors by mistakenly pressing a wrong key, it was more appropriate to take the rule of 95% correctness than the rule of 100% correctness. A similar criterion of 95% correctness was also used in previous literature. The maximum oddball duration corresponding to the 95% correctness of 21 participants was 1320 ms. This critical parameter (1320 ms) was then used in Experiment 1a where multiple spatial locations were introduced to measure the CSTP.

2.2 Results and discussion

With an increase of the set size or the number of involved spatial locations, participants' performance deteriorated (Figure 3) and capacity estimation reached saturation (Figure 1B). A formula of calculating Cowan's K was used to estimate the spatial capacity of simultaneous temporal processing in visual modality, similar to the calculation of the capacity of VWM in CDP. Here Cowan's K was defined as: K = (hit rate + correct rejection rate − 1) × N; N equals to set size. This approach was widely used to measure capacity of the visual working memory,
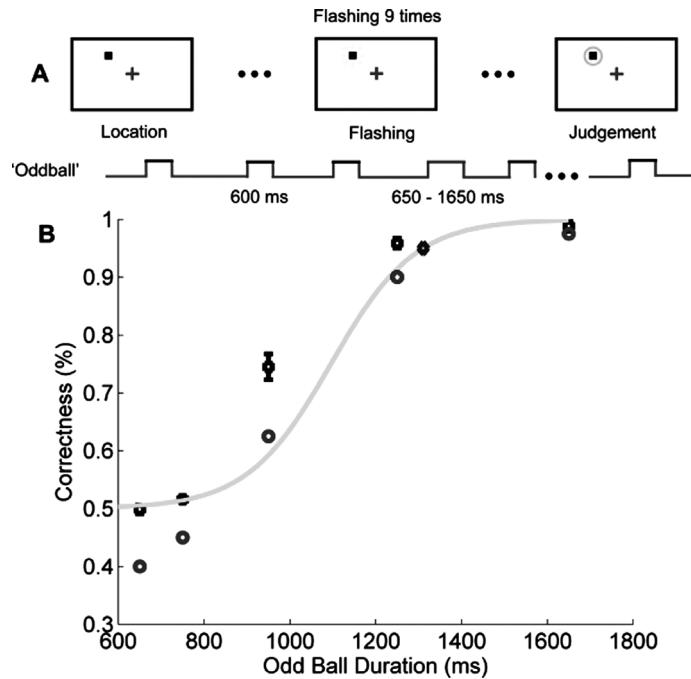
Figure 2. Experimental approach and results of the oddball duration experiment. A, Experimental approach of the oddball duration experiment. **B**, Results of the oddball duration experiment. Black squares are mean correctness of 20 participants. Red circles are correctness data from a single participant who had the maximum odd ball duration corresponding to 95% correctness (denoted by a blue diamond) after fitted by a sigmoid model (green curve). Error bars are within–subjects SEs .
doi:10.1371/journal.poen.0091797.g002.

including subsequent working memory experiment in this study.

The results of Experiment 1a (Figure 1B) showed that the Cowan's K was close to one in the condition of set size 1 (i.e., perfect performance at set size 1), indicating a valid estimation of the required minimum oddball duration. With increase of the set size, the Ks also increase but tend to be saturated.

Determine capacity based on cowan's K. In the experiments involving TODP and working memory (see below Experiment 1b), capacity was operationally defined as the mean of Cowan's Ks for a subset of all set size conditions, where K value of each subset member was significantly larger than K values of those conditions that were outside of the subset and had smaller set sizes. Meanwhile the K value of the largest set size condition inside the subset shouldn't be significantly smaller than those of the rest subset members (This was to correct the calculation of capacity when there is a drop of K value at the largest set size condition). By this definition, the

capacity in Experiment 1a was defined as the averaged K values of set size 5, 7, and 9. This was based on the post–hoc pairwise comparisons (after Bonferroni corrections, see Table 1 for the statistical results). Capacity calculations of both multi–temporal processing and working memory (Experiment 1b, see Table 2 for the statistical results) were based on the same principle throughout this study.

By using above statistical procedure, we operationally defined the capacity of simultaneous temporal processing (CSTP) and found that the average capacity was 3.32 (N=29, SD=1.58). These results demonstrated that the capacity of STP is limited for human, in a range between 3 to 4 locations, with a relatively large individual difference (SD=1.5755).

The result suggested that our visual timing system can co–activate and maintain 3 to 4 independent 'local clocks'. However, this suggestion seems contradicting with the conclusion of an earlier study by Morgan and colleagues which proposed a single "stopwatch" for duration
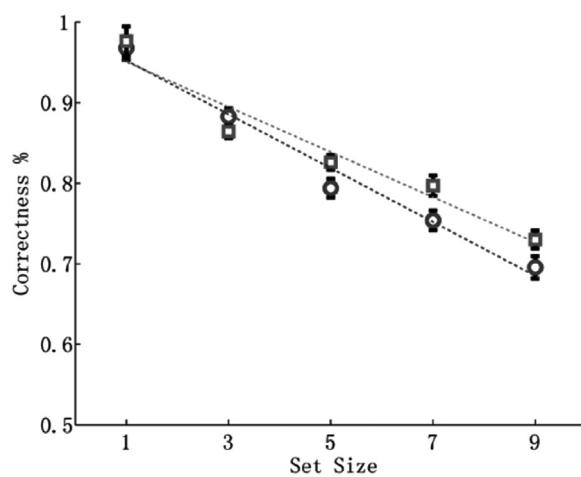
Figure 3. Percentages of correct responses of each set size condition in Experiment 1a (N=29; red circles) and Experiment 2 (N=21; green squares) and their linear regression curves. The red curve is for Experiment 1a (slope=−0.034; R^2=0.979; $F_{(3, 1)}$= 139.227; p< .002) and the green curve for Experiment 2 (slope=−0.028; R^2=0.941; $F_{(3, 1)}$= 47.739; p< .007). Error bars are within−subjects SEs. The slope of regression curve in Experiment 1a was not significant different from the slope of regression curve in Experiment 2 (Independent samples t−test; t(48)= −1.397; p=0.169), suggesting the percentages of correct responses in both experiments obeyed a Weber´s law.
doi:10.1371/journal.poen.0091797.g003

estimation, similar to a single "ruler" for size estimation (but see for a different opinion on size estimation). The TODP paradigm in our study is also similar to the visual search paradigm used in the Morgan et al.'s study. In the 'General Discussion', we gave a thorough comparison and analysis on similarities and differences between these two paradigms and concluded that the claims of these two studies are not necessarily incompatible.

3 Experiment 1b

Experiment 1a provides the very first measurement to the capacity of the location−bound simultaneous timing system in vision. According to the result of Experiment 1a, it is likely that our visual timing system can co−activate and maintain 3 to 4 independent 'local clocks'. This is consistent with recent researches on rats and human participants, demonstrating that subjective time estimation in a simultaneous temporal task from millisecond to second

level is represented by multiple clocks. Those clocks operate independently to make temporal judgment in a context−dependent way. However, our understanding to how those 'local clocks' operate is still scarce. In the next several experiments, we tried to explore the nature of the 'local clocks' from two important aspects. First, due to the requirement of our task, the representation of accumulated pulses of each 'local clock' need to be maintained and updated online in some memory−like unit. This raises the question of whether the maintenance and updating work in a similar way as the manipulation of representations in the visual working memory. In other words, are those representations of clock−pulses transferred into VWM before the online manipulation? If so, it would be reasonable to predict that the capacity of an individual's VWM constrains the capacity of the individual's STP. Alternatively, if the representations of clock pulses are manipulated in units different from those of visual working memory, e.g. some type of pre−working memory registers characterized with easy access, temporary keeping and rapid manipulation, the two capacities should not be constrained by each other.

3.1 Materials and methods

In order to answer the above question, the same group of participants was invited into a working memory experiment. The memory array in this experiment (see Figure 4A) consisted of colored square(s) with one out of five set size options (1, 3, 5, 7 or 9). Each square was sampled at random from a set of seven highly discriminable colors (red, blue, violet, green, yellow, black and cyan), and a given color could appear no more than twice within an array. The memory array was presented for 100ms. Then it was followed by a 1000−ms white blank interval and finally a presentation of the test array which would disappear until the participants made a response. One color of the item in the test array was different from the corresponding item in the memory array on 50% of trials; otherwise the memory

Table 1. Mean k value differences and p values of pairwise comparisons in Experiment 1a.

Set Size	1		3		5		7	
	MD	p	MD	p	MD	p	MD	p
3	1.354***	.000						
5	1.974***	.000	0.620*	.031				
7	2.602***	.000	1.248***	.000	0.628	.098		
9	2.587***	.000	1.233*	.039	0.613	.532	−0.015	1.00

Note: * p < 0.05, *** p < .001

Mean K value differences in Experiment 1a (Bonferroni corrected pairwise comparisons), indicating that K values of set size 5, 7 and 9 were not significantly different from each other. However, these K values were all significantly larger than those of set size 1 and 3.
doi:10.1371/journal.poen.0091797.t001

Table 2. Mean differences of perceived time distortion and p values of pairwise comparisons in experiment 4.

Number of encoded spatial locations	1L		2Ls		3Ls	
	MD	p	MD	p	MD	p
2Ls	57.2	.199				
3Ls	146.8***	.000	89.7***	.000		
4Ls	151.6***	.000	94.5***	.000	4.8	1.00

Note: *** p < .001
Mean differences of perceived time distortion in Experiment 4 (Bonferroni corrected pairwise comparisons) indicating that the perceived time distortions of 1L and 2Ls conditions were significantly different from those of 3Ls and 4Ls conditions.
doi:10.1371/journal.poen.0091797.t002

and test arrays were identical. Participants were required to judge whether the color square in memory and test array were identical or not. Other aspects of this experiment, such as the size and layout of the squares, were identical as Experiment 1a.

3.2 Results

The mean VWM capacity was 2.56 (N=29, SD=0.64) in this experiment (see Figure 4B), a value close to the capacity of 2.8 in Vogel and Machizawa's 2004 study , suggesting a high consistence between these two studies that used similar paradigms. The individual-based correlation between the capacities of VWM and STP was not significant (p=0.28) (see Figure 4C), suggesting that these two types of capacities are likely based on different cognitive resources and the representations of accumulated time pluses in each 'local clock' are not transferred into the VWM units automatically. This is consistent with a recent research, demonstrating a dissociation of visual working memory and the number of encoded spatial locations.

4 Experiment 2 & 3

Other factors that might influence the result of Experiment 1a include stimulus crowding and eccentricity. Experiment 1a used a stimulus layout similar to most change detection paradigms in VWM studies, where the eccentricities of different targets and the distances between them were randomized to counterbalance the potential effects of stimulus crowding and eccentricity. We want to ask whether the capacity obtained in Experiment 1a is stable after controlling stimulus properties related with crowding and eccentricity. Experiment 2 & 3 still used the same paradigm, but the spatial layouts of the square(s) were different from Experiment 1a in that all the locations were now positioned along a virtual circle centered at the fixation point (Figure 1C). This stimulus layout allowed
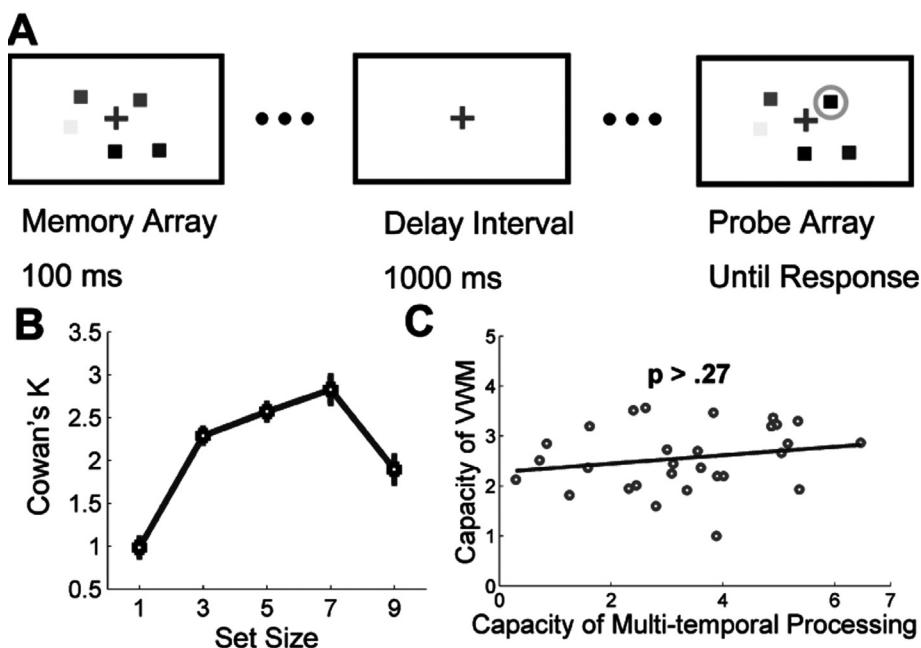
Figure 4. Experimental approach and results of Experiment 1b (the working memory experiment). A, Experimental approach of the working memory experiment (an exemplary trial of set size = 5). **B**, The mean Cowan´s Ks of each set size condition in the working memory experiment. **C**, individual´s working memory capacity was not significantly (p >.27) correlated with his or her spatial capacity of multi–temporal processing. Error bars are within–subjects SEs .
doi:10.1371/journal.poen.0091797.g004

the spatial attention to be distributed along a ring–shape area to obtain a constant eccentricity. Meanwhile only nine locations on this invisible ring were used as candidate locations for targets. Thus, the crowding effect was minimized by relatively large and strictly–controlled distances between the targets.

4.1 Materials and methods

Twenty one paid participants (17 female, 4 male; mean age=22.14 years) from CCNU took part in Experiment 2. An additional 20 paid participants (17 female, 3 male; mean age=21.5 years) from CCNU took part in Experiment 3. All aspects of Experiment 2 were the same as Experiment 1a, except that all objects were now positioned along a virtual circle with a radius of 11.3 deg and centered in the fixation (Figure 1C), to control the potential effect of stimulus crowding and eccentricity. The minimum angular distance between any two objects was 40° due to a maximum set size of 9 and all the potential locations were equally distributed along the circle. In a new trial, the nine positions were given a random jitter (maximum rotation angle as 40°) to avoid expectation from the previous trial. Experiment 3 was identical as Experiment 2, except that those locations without objects in Experiment 2 were now occupied by stationary black squares of same size as other squares, i.e., the distracting objects, throughout the trial.

4.2 Results

With above changes applied, the CSTP in Experiment 2 was 3.86 (Figure 1D, N=21, SD=1.57, same capacity definition as Experiment 1a), not significantly different from that of Experiment 1a (Independent samples t–test, t(48)= −1.198, p= 0.237). This showed that the capacity of STP is reliable after controlling crowding and eccentricity–related spatial properties of the stimuli.

A second aspect in our exploration to the nature of the 'multiple clocks' was to evaluate what role spatial attention plays in a simultaneous temporal task. In Experiment 3, distracting objects, i.e., the black squares which were always stationary and thus perceptually highly–

distinguishable from the dynamic targets, were used (Figure 1C). Independent samples t–tests were used to compare the results of Experiment 2 and 3. We found a significant decrease of Cowan's Ks in Experiment 3 (Figure 1D) at set sizes of 5 (Difference = −0.87, t(39)= −2.251, p<.05, SE= 0.39) and 7 (Difference = −1.91, t(48)= −3.267, p<.003, SE= 0.58), but not at set sizes of 1 (p =0.95), 3 (p=0.23) and 9 (p =0.19) relative to that of Experiment 2. This result demonstrated that the spatial attention was necessary for the 'local clocks' to become location–bound. When the allocated attention on the targets was reduced, such as the set size 5 and 7 relative to the set size 3, the binding between clock's pulse accumulation and its local location was weakened. This can be seen by the mistaken conjunctions between the pulse accumulation of a target and the distracting information from other should–be–ignored locations of the distractors. Those mistaken conjunctions resulted in reduced Cowan's K at set size 5 and 7 but not at set size 3 where adequate attention was allocated on each of the targets.

5 Experiment 4 & 5

Previous research on multiple–object–tracking had demonstrated that humans are able to monitor four or five objects simultaneously. Xu and Chun proposed a two–stage model, including object individuation and object identification, to account for how multiple objects are attended and perceived. This model assumes that capacity limits of various visual phenomenon, such as working memory, enumeration, and multiple object tracking, which are all around four to five objects or regions of interest, are largely due to the constrains of object individuation. The typical explanation to the limitation of object individuation is that our cognitive resource, such as spatial attention, is limited at any given moment and can not cover more than four or five objects simultaneously. Here, we tried to

argue that the simultaneous temporal processing is limited in its capacity even without the explicit involvement of the object individuation (at least in its current definition). For that purpose, we designed a second paradigm called spatial splitting of a temporal interval (SSTI) to investigate this issue. The essence of this paradigm was to measure how perceived duration of a physically–fixed temporal interval (1100ms, standard interval) varied by the number of its encoded spatial locations (1, 2, 3 or 4 location(s)). Here, participants were not required to allocate spatial attention to multi–locations which might trigger the object individuation automatically. There were four critical conditions (denoted as 1L, 2Ls, 3Ls and 4Ls in this paper) in Experiment 4 and three (i.e., 2Ls, 4Ls, and 6Ls) in Experiment 5, corresponding to the number of different spatial locations involved in the standard interval. All the involved locations were stimulated within a time window of several hundred milliseconds.

The idea behind the SSTI paradigm is to test whether only three to four location–bound 'local clocks' were activated and maintained during a time window of several hundreds milliseconds. The perceived duration of the standard interval was determined by a summation of 'clock pulses' during the sequential presentation of the targets. The activation of any 'local clock' will trigger a production of a stream of 'clock pulses' at that location, and then contribute to the final duration estimation. However, the opening of a new 'clock' at a new location will not close the current 'clock' at a different location automatically and immediately. This is due to that the 'local clocks' operate independently in vision. The additional activity, due to the delayed closure of the previous 'local clocks', would enlarge the summation of the total 'clock pulses', as well as the final time estimation. Thus, our prediction is that the time perception of the standard interval in a SSTI paradigm should expand as the number of involved spatial locations increases in a time window of several
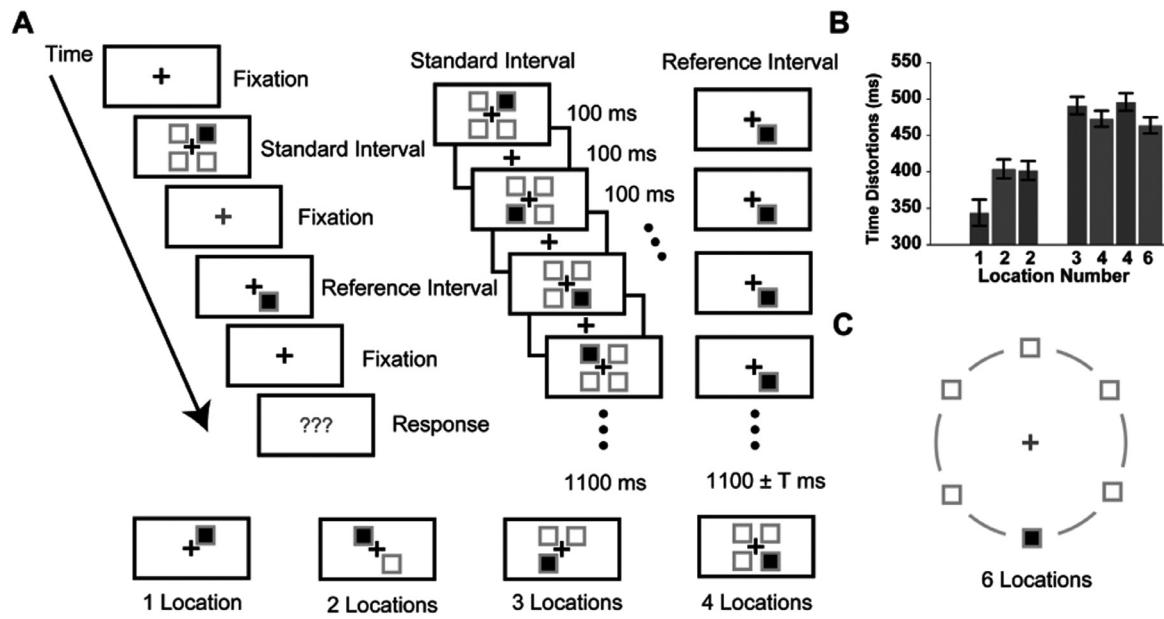
Figure 5. Experimental approach and results of Experiment 4 and 5. A, Duration comparison task of the SSTI paradigm. One typical trial of Experiment 4 (upper panel) and its four critical conditions (lower panel). Examples of a standard interval (upper–middle part) and a reference interval (upper–right part) of the 4Ls condition were given. Grey placeholders denoted the involved spatial locations in each critical condition and did not appear in the experimental display. **B**, Results of Experiment 4 (N=33, green bars) and Experiment 5 (N=30, red bars). **C**, Stimulus layouts of the six location experiment (a 6Ls condition). Grey circles did not appear in the experimental display. Other aspects were similar to Figure 5A lower panel. Error bars are within–subjects SEs.
doi:10.1371/journal.poen.0091797.g005

hundred milliseconds. Also, this effect should saturate at three to four regions of interest due to the limited number of location–bound 'local clocks' that can be activated and maintained concurrently in visual modality.

5.1 Materials and methods

Thirty three paid participants (21 female, 12 male; mean age = 21.4 years) from CCNU took part in Experiment 4. An additional 30 paid participants (11 female, 19 male; mean age = 21.7 years) from CCNU took part in Experiment 5. In the second paradigm shown in Figure 5A, participants viewed a central fixation followed by two temporal intervals. The standard interval was always 1100ms, consisted of 1, 2, 3 or 4 to–be–encoded spatial location(s). The reference interval always contained a single location with a variety of durations. Participants were asked to indicate which one of the two intervals was perceived longer in duration.

In the basic SSTI paradigm of Experiment 4, stimulus

viewing aperture was a square, subtending 12 deg wide × 12 deg high, against a white background. Each trial consisted of a standard interval (1L, 2Ls, 3Ls or 4Ls) and a reference interval.

Each standard interval included 6 sequentially–presented black squares (4 deg wide × 4 deg high, each lasted for 100ms) and 5 100ms–blank intervals between them. The duration between the onset of the first square and the offset of the last one, which was always 1100ms physically, indicated the to–be–judged standard interval. The 100ms blank interval was selected deliberately to minimum the effect of the 'persistence of vision' of the immediate–last square, which typically occurs with a blank interval less than 40 ms. Also, the Inhibition of Return (IOR) , an effect of impaired detection to objects appearing in previously cued locations relative to locations not previously cued, is unlikely to play a role in this paradigm because a typical IOR needs a SOA of at least

300 ms while the SOA is 200ms in the present study, and a typical IOR also needs a cue–response paradigm which is absent here. The spatial locations of 4Ls condition were taken from 4 corners of the viewing aperture. The minimum distance between the centers of any 2 squares was 8 deg, a large distance to reduce the effect of apparent motion. Neighboring squares in a standard interval always had different spatial locations. The SOA between any 2 neighboring squares was 200ms, resulting in a sequential location–stimulation of 5 Hz. In the standard interval, all the involved locations went through once with random order before next going–through. By this arrangement, all the involved locations were presented within the first 100ms, 300ms, 500ms, and 700ms for the 1L, 2Ls, 3Ls, and 4Ls conditions, respectively. In other words, all the involved locations were stimulated within several hundred milliseconds. In this paradigm, although multiple spatial locations were included, a single object and its occupied single spatial location were involved at any give moment to indicate the continuation of the standard interval. This feature is a critical departure from previous others' paradigm, for example a study which demonstrated a perceived time expansion induced by simultaneously–presented multiple objects on the display.

The reference interval was composed of one static black square, presenting continuously at one of four locations (Figure 5A, upper–right panel). The spatial position of the reference interval was counterbalanced across different trials. The duration of the reference interval was equally sampled from 7 options, e.g., 200, 800, 1100, 1400, 1700, 2000 or 2600 ms. Here, four of the reference intervals were longer than the standard interval, one was equal to the standard interval, and two were shorter than the standard interval. The reference intervals were 300ms longer, on average, than the standard interval (1100ms) because our pilot testing revealed that the 1100ms standard (dynamic) interval was perceptually overestimated at least 300ms

relative to a 1100ms (static) reference interval, which is consistent with previous literature demonstrating a time dilation for dynamic stimuli relative to static stimuli. Similar approach of asymmetrical arrangement of the reference stimuli around the standard stimulus was used in recent studies. It is important to note that although this may have affected the distribution of responses, the effect would be constant across all conditions used here. The usage of 4 possible locations within each critical condition (1L, 2Ls, 3Ls or 4Ls) of both standard and reference intervals was counterbalanced across trials to make sure that each location had equal chance to be used. Thus, the effect of apparent motion was equalized across the conditions of 2Ls, 3Ls and 4Ls. Participants were instructed to keep gazing at a fixation cross throughout the experiment.

The standard and reference stimuli were sequentially presented within the two intervals, with a random ISI (Inter–Stimulus Interval) varying between 500ms and 1000ms. The order of the standard and the reference intervals in a trial was counterbalanced across different trials. The participants' task was to indicate which interval (the first or the second) contained duration that seemed to last longer by pressing one of two response keys. Each level of duration of the reference interval was repeated for 24 times, resulting in a total of 672 trials (168 trials for each condition of 1L, 2Ls, 3Ls or 4Ls). Different conditions were randomly mixed during testing.

All aspects of Experiment 5 were identical as Experiment 4, except that 1L and 3Ls conditions were replaced by a new condition, i.e., 6Ls, where six potential locations were used and all of them positioned along a virtual circle centered in the fixation with a radius of 5.66 deg (Figure 5C). The minimum angular distance between any two locations was 60 o due to that all the six locations were equally distributed along the circle. In any new trial, the six positions were given a random jitter (with a

rotation angle between 0° to 60°) to avoid expectation from the previous trial. There were 3 critical conditions in Experiment 5, i.e., 2Ls, 4Ls and 6Ls.

5.2 Results

In Experiment 4, and 5 that involved the measurement of perceived time distortions, a sigmoid curve was fitted on each individual's data to calculate the PSE (Point of Subject Equality) at each critical condition (1L, 2Ls, 3Ls, 4Ls, or 6Ls). Participants had a 50 % percentage to report a 'longer' standard interval at the PSE, indicating that the standard and the reference intervals were perceptually equal in duration. The perceived time distortions were subsequently defined as the differences between the PSEs and the standard intervals (1100 ms).

As predicted, Experiment 4 demonstrated an effect of number of encoded spatial locations on time estimation ($F_{(3, 96)} =21.932$, $p<0.001$). With the number of the encoded spatial locations increased, the standard interval was perceived longer in duration (Figure 5B). Particularly, the 1L and 2Ls conditions were qualitatively different from the 3Ls and 4Ls conditions in that the mean perceived time distortion of 1L and 2Ls were significantly smaller (Difference= −121ms, $t_{(32)}= −7.424$, $p< .001$) than that of 3Ls and 4Ls conditions. The hypothesis that the location−number−related effect reaches a plateau at around three to four spatial locations received further support from Experiment 5, where no significant difference was found between 4Ls and a condition with more location number, i.e., 6Ls (Figure 5C), though time distortions of both conditions were larger than that of 2Ls condition ($p <.008$ for 4Ls and $p <.03$ for 6Ls, Bonferroni corrected). This result supported the idea that only 3 to 4 location−bound 'local clocks' are activated and maintained during a time window of several hundreds milliseconds. Our finding implicates that the hypothetic 'object individuation' might need a modification, e.g., to include the 'multiple clocks' maintained during a time window of several hundreds milliseconds, in order to explain the capacity of simultaneous temporal processing.

6 General discussion

Our results, based on two novel and complementary paradigms, indicate consistently that the capacity limit of simultaneous temporal processing in vision is around 3 to 4 spatial locations and the capacity of STP is not constrained by the capacity of visual working memory (VWM). Moreover, the binding of the 3 to 4 'local clocks' and their specific location is undermined by bottom−up competition of spatial attention, indicating that the time−space binding is resource−consuming. The second paradigm confirms further that the limited number of location−bound 'local clocks' are activated and maintained during a time window of several hundreds milliseconds.

6.1 Distributed attention vs. Focal attention

In an earlier study, Morgan and colleagues (2008) explored whether the different thresholds of temporal oddball−discrimination was related with set size in a visual search paradigm. They asked their participants to report whether a single "odd duration" was shorter or longer than the other distractor durations inside a visual search array with a set size of 2, 4, 6 or 8. The temporal orders of the onsets of the odd duration and the distractors were randomized, similar to the onset timing of the oddball and standard squares in our TODP paradigm, though their stimuli were all horizontal lines staggered in space. The usages of the temporal oddball and multiple distractors were very similar in these two paradigms. However, Morgan and colleagues found that the precisions, i.e. Weber fractions, of discriminating an odd duration from a set of distractors were affected by the set size and concluded that a centralized supramodal clock, i.e. a single "stopwatch", was used for duration estimation. Although our results suggested 3−4 independent 'clocks', our conclusion is not

necessarily contradicted with their conclusion because our TODP paradigm differs from Morgan and colleagues' visual search paradigm (2008) in significant ways.

First, by using an eye tracker our TODP paradigm controlled eye movements and elicited a broadly *distributed attention* over the entire stimulus viewing aperture, whereas the visual search paradigm of Morgan and colleagues allowed multiple fixations and shifts of *focal attention* between different spatial locations during the search to the odd duration (0.5s, 2s and 8s). Here, focal attention means to concentrate on small areas for visual processing. It has been well documented in the visual search literature that visual search performance was qualitatively different between *distributed* attention and *focal* attention. For example, it was revealed that a broadly distributed attentional allocation is sufficient for participants to search an odd–coloured target in a parallel way with unlimited capacity, i.e. a parallel process is operating over large areas. In contrast, when participants were asked to use saccades to indicate the odd–coloured target, a serial search by shifting focal attention to each target was performed. This implied that the goal–directed saccades are concurrent with the shifts of focal attention even in simplest visual search tasks. Thus, it is not surprising that a serial–like process was found in Morgan and colleagues' paradigm, while a parallel–like process was elicited in our TODP paradigm. Second, task difficulty or more specifically the feature distance between a target and its background might also contribute to the differences between these two paradigms. By using the oddball duration experiment we selected a large feature distance between the oddball (1320ms) and the standard intervals (600ms), i.e. corresponding to correctness in–between 95% to 100%. Morgan and colleagues' study measured a discrimination threshold of 82.5% correct response. To reduce performance to threshold, their paradigm used very similar targets and distractors, whereas the targets and

distractors in our paradigm were more different. Based on the suggestion of previous literature, a task of a larger distance in feature space, such as the task of our TODP paradigm, is more likely to elicit a parallel process than the task of Morgan and colleagues.

Another factor that might be relevant here is the difference between a *detection* task and a *discrimination* task. Sagi & Julesz (1985) proposed that to detect a feature of a target that is *different* from its background involves different attentional process compared with to discriminate the *direction* of the feature gradient of a target from its background. The former involved a parallel (preattentive) process, while the latter involved a serial attentive process. In its definition, our TODP paradigm was a typical detection task (to detect the occurrence of a 'longer' duration), while Morgan and colleagues' paradigm was more likely to be a discrimination task involving feature directions (to discriminate a 'shorter' or 'longer' oddball) in each trial.

Taken together, we proposed that the differences in task demands, task difficulties, and performance measurements, might interact in a complex way and elicit essentially different attentional processes for these two paradigms, i.e. distributed attention in our TODP paradigm and focal attention in Morgan and colleagues' paradigm. The former might induce parallel–like process by *distributed* attention, while the latter might elicit serial–like process by *focal* attention. In that sense, our conclusion of the 3 to 4 concurrent 'local clocks' in vision are not necessarily incompatible with Morgan and colleagues' claim that a single "stopwatch" is operating for duration estimation. In fact, it is very likely that both claims revealed different aspects of how same amount of a limited–capacity is dynamically allocated under different situations, i.e. *distributed* attention vs. *focal* attention.

6.2 Constraints from spatial processing

In visual modality, spatial encoding and temporal

processing are closely linked. The classic Kappa effect and Tau effect indicated dynamic interactions between spatial judgment and time perception. A major theoretic development in the past decade is Walsh's theory which sheds new light on the relationship between spatial encoding and temporal processing. According to this theory called ATOM (A Theory of Magnitude), processing of time, space, or quantity all shares a common magnitude system and overlapped neutral substrate, e.g. the inferior parietal cortex. The present study adds new evidence to a growing body of literatures that the temporal processing in visual modality is constrained by spatial processing.

The precise estimation to the duration of a visually-presented object, such as the square in the present study, relies on continuous online updating and monitoring to the object's spatial feature, such as the brightness. Thus, it is justifiable that the spatial processing, particularly the quality and frequency of the 'spatial feature monitoring' can constrain the object's temporal representation, i.e., the accumulated time 'pulses'. The construction of the temporal representation relies on a successful conversion of encoded information from spatial domain into temporal domain. The exact nature of this information conversion is not yet clear, partially due to the debate on how temporal pulses are accumulated, i.e., by a linear or nonlinear way Adaptation studies implied the existence of location-based timing system in vision. However, it is not yet known how spatial processing imposes constraints on this temporal processing. Our findings of the present study demonstrate clearly that this location-based timing system is subject to a spatial-processing-related bottleneck, i.e., at any given moment or during any interval of hundred milliseconds the information conversion from spatial domain into temporal domain can only be maintained concurrently at 3 to 4 spatial locations.

It is worth to note that this bottleneck is unlikely solely determined by low-level factors, such as crowding and eccentricity-related effects, since our Experiment 2 demonstrated a relatively stable capacity of STP after controlling those low-level spatial properties. Instead, this bottleneck is most likely located at relatively higher level of visual hierarchy, such as those linked with limited resource of spatial attention. This is confirmed by our Experiment 3, which indicates that certain amount of the allocated spatial attention is necessary for the successful conversion of information from spatial domain into temporal domain. Considering the change of the general pattern of distributed attention from Experiment 2 to 3, when the averaged amount of allocated attention at each of the relevant spatial locations was reduced by the introduction of a group of should-be-ignored static distractors, i.e. with the presence of bottom-up attentional competition, the temporal estimations of those relevant locations were greatly undermined. It is an open question for future study to explore whether the attentional competition also invokes interference to spatial processing. Similar effect was observed in the situation of multiple-object-tracking in that the changes of the moving objects' spatial features, such as color or shape, is surprisingly blinded and miss-reported when attentional resource is directed to the tracking of the objects' identities.

6.3 Multiple capacities in vision?

An interesting observation in the present study is that the capacity of STP is not constrained by an individual's capacity of VWM. This implies that multiple capacities, rather than a universal single one, exist in visual modality. Indeed, similar dissociations of separate capacities in vision had been reported before. For example, Hyde & Wood suggested that in nonverbal numerical processing the approximate number system (ANS)-based numerical representations are not constrained by the capacity of VWM since those representations either only take one VWM slot or do not enter into VWM automatically. Other literature demonstrated that the attention-

related capacity, including multiple object tracking (MOT), detection of RSVP and nonverbal enumeration etc, is qualitatively different from a VWM—related capacity, such as that measured in a change detection task of VWM. The attention—related capacity can not be reducible to a VWM—capacity. Instead, VWM—capacity relies on multiple factors, including visuospatial attention; a central, amodal supervisory process; and local stage—specific operations. One critical similarity between STP of our task and nonverbal numerical processing is that both of them are subject to a spatial attention—based bottleneck. This leads to why the capacity of STP in the present study is not constrained by an individual's capacity of VWM.

As suggested by our second paradigm, the STP capacity is unlikely dependent on the object individuation, i.e., the activation of a parallel individuation system. This is further confirmed by the fact that we found reliable set size—correctness relationships in Experiment 1a and 2. In those two experiments, the relationship between the correct response and the set size all obeyed Weber's laws (see Figure 3). That is to say, the size of the variance introduced into the temporal estimation in our STP task is positively correlated with the number of spatial locations that need to be attended to. This indicates that at all levels of different spatial locations our participants did not simply select 3 to 4 objects first and mask the rest of all the other objects. Instead, the cognitive resource is distributed across all the possible spatial locations in a parallel way with limited capacity. Particularly, when spatial attention is further spread over the visual scene by introducing more to—be—processed spatial locations, more variance is added into the numerical estimation at each location and thus increases the response error. The obedience to Weber's law is a hallmark of the approximate number system in nonverbal numerical processing rather than that of the subitizing which is likely based on a parallel individuation

system. Our result adds new evidence that nonverbal numerical processing and temporal estimation share common cognitive principles and even neutral basis. This is supported by a recent neuroimaging study, indicating both numerosity and duration estimations share a common right fronto—parietal network. All above facts can be justified in the context of an accumulator model in that both numerosity and time are likely based on a common accumulator that is working under two different modes.

A recent study on nonverbal numerical processing demonstrated that it is the allocation of spatial attention rather than the absolute magnitude of quantity itself that determines which system, a parallel individuation system or a ANS—like 'numerical magnitude' system, should be deployed in a given visual scene. When objects are too numerous to be selected simultaneously or spatial attention is distributed across the visual field, the visual system will invoke the ANS system to perform approximate estimation to numerical information, even to small numbers (< 3), in the visual scene. A STP task in vision, such as the TODP paradigm here, requires constant attentional consumption and distributed attention. Thus, it is not surprising that visual system automatically invoke an ANS—like system to estimate the number of accumulated temporal pulses at each attended location. This explains why we did not get a Cowan's K close to 3, i.e., close to 100% correct performance (as predicted by a parallel individuation—like system), at the condition of set size 3 in Experiment 2 when this set size is well below the capacity of STP (3.86). Further data collections are needed in order to explore whether the counterpart of ANS in the domain of time perception, i.e., a plausible ATS (Approximate Time System)—like system, exists in simultaneous temporal processing.

6.4 Neural basis of STP

Previous neurophysiologic studies provided clues on the possible neural basis of the capacity of simultaneous

temporal processing. In one hand, a study demonstrated that two distinct neuron populations in the striatum (STR) of rats were activated differentially by rewards of two criterion durations (10s vs. 40s), implicating two 'clocks' in the striatum. On the other hand, other study showed that the prefrontal cortex (PFC) was necessary in terms of dynamically allocating attention to multiple timed stimuli. These results are compatible with our findings that 'multiple clocks' can coexist and the bindings between the 'multiple clocks' and specific locations are attention–dependent. A growing body of researches on human and monkey participants had demonstrated that a distributed network, including basal gangalia, cerebellum, PFC, pre–supplementary motor area (preSMA) and supplementary motor area (SMA), involves in multiple aspects of temporal processing depending on the task requirements, i.e., motor vs. perceptual and sub–second vs. supra–second judgments. Further neural imaging studies on human by using sub–second intervals are necessary to investigate where a potential bottleneck, corresponding to the capacity of simultaneous temporal processing, lies in the cortico–striatal circuits and its exact role in our ability to time multiple events simultaneously.

In conclusion, we found that the temporal processing in visual modality is constrained by spatial processing when time estimations are based on information of encoded spatial locations. Specifically, the present study provide converging evidence to support that our capacity of simultaneous temporal processing is limited at around 3 to 4 spatial locations in visual modality. This spatial capacity in vision is subject to the allocation of spatial attention but not constrained by the capacity of working memory and is unlikely be determined by the object individuation. This conclusion receives highly consistent supports from two novel and complementary paradigms.

A demonstration of stimuli used in this study can be found at http://visionlab.byethost7.com/.

References

Ayhan, I., Bruno, A., Nishida, S., & Johnston, A. (2009). The spatial tuning of adaptation–based time compression. *Journal of Vision, 9,* 1–12.

Bahrami, B. (2003). Object property encoding and change blindness in multiple object tracking. *Visual Cognition, 10.*

Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature, 303,* 696–698.

Bergen, J. R., & Julesz, B. (1983). Rapid discrimination of visual patterns. Systems, Man and Cybernetics, *IEEE Transactions SMC–13,* 857–863.

Bo, J., & Seidler, R.D. (2009). Visuospatial Working Memory Capacity Predicts theOrganization of Acquired Explicit Motor Sequences. *Journal of Neurophysiology, 101,* 3116–3125.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision, 10,* 433–436.

Bravo, M. J., & Nakayama, K. (1992). The role of attention in different visual–search tasks. *Perception & Psychophysics, 51.*

Brown, S. W., & West, A. N. (1990). Multiple timing and the allocation of attention. *Acta Psychologica, 75,* 103–121.

Buhusi, C. V, & Meck, W. H. (2009). Relativity theory and time perception: Single or multiple clocks? *PLOS ONE, 4,* e6268.

Burr, D., Tozzi, A., & Morrone, M. C. (2007). Neural mechanisms for timing visual events are spatially selective in real–world coordinates. *Nature Neuroscience, 10,* 423–425.

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences, 9,* 349–354.

Cicchini, G. M., & Morrone, M. C. (2009). Shifts in spatial attention affect the perceived duration of events. *Journal of Vision, 9,* 1 – 13.

Coltheart, M. (1980). The persistences of vision. *Philos Trans R Soc Lond B Biol Sci 8, 290,* 57 – 69.

Cousineau, D. (2005). Confidence intervals in within–subjects designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology, 1,* 42 – 45.

Cowan, N. (2001), The magical number 4 in short–term memory: A reconsideration of mental storage capacity. *Behavioral Brain Sciences, 24,* 87 – 114.

Dormal, V., Dormal, G., Joassin, F., & Pesenti, M. (2012). A common right frontoparietal network for numerosity and duration processing: An fMRI study. *Human Brain Mapping, 33,* 1490 – 1501.

Fougnie, D., & Marois, R. (2006), Distinct capacity limits for attention and working memory. *Psychological Science, 17,* 526 – 534.

Gibbon, J. (1977), Scalar expectancy theory and Weber's law in animal timing. *Psychological Review, 84,* 279 – 325.

Hagit, M., Emmanouil, T–A., McMains, S. A., Kastner, S., & Treisman, A. (2009). Attentional demands predict short–term memory load response in posterior parietal cortex. *Neuropsychologia, 47,* 1790 – 1798.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the 'number sense': The approximate number system in 3–, 4–, 5–, and 6–year–oldsand adults. *Developmental Psychology, 44,* 1457–1465.

Hartshorne, J. K. (2008). Visual Working Memory Capacity and Proactive Interference. *PLoS ONE, 3,* e2716.

Helson, H. (1930). The tau effect–an example of psychological relativity. *Science, 71* (1847): 536 – 537.

Hyde, D. C, & Wood, J. N. (2011). Spatial attention determines the nature of nonverbal number representation. *Journal of Cognitive Neuroscience, 23,* 2336 – 2351.

Ikkai, A., McCollough, A. W., & Vogel, E. K. (2010). Contralateral delay activity provides a neural measure of the number of representations in visual working memory. *Journal of Neurophysiology, 103,* 1963 – 1968.

Johnston, A., Bruno, A., & Ayhan, I. (2011). Retinotopic selectivity of adaptation–based compression of event. *Journal of Vision, 10,* 1 – 3.

Johnston, A., Arnold, D. H., & Nishida, S. (2006). Spatially localized distortions of event time. *Current Biology, 16,* 472 – 479.

Kanai, R., Paffen, C. L., Hogendoorn, H., & Verstraten, F. A. (2006). Time dilation indynamic visual display. *Journal of Vision, 6,* 1421 – 1430.

Klapproth, F. (2011). Temporal decision making in simultaneous timing. *Frontiers in Integrative Neuroscience, 5.*

Lewis, P. A., & Miall, R. C. (2003). Brain activation patterns during measurement of sub– and supra–second intervals. *Neuropsychologia, 41,* 1583 – 1592.

Lewis, P. A., & Miall, R. C. (2006). A right hemispheric prefrontal system for cognitive time measurement. *Behavioural Processes, 71,* 226 – 234.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279 – 281.

Matell, M. S., King, G. R., & Meck, W. H. (2004). Differential modulation of clock speed by the administration of intermittent versus continuous cocaine. *Behavioral Neuroscience, 118,* 150 – 156.

Matell, M. S., Meck, W. H., & Nicolelis, M. A. (2003). Interval timing and the encoding of signal duration by ensembles of cortical and striatal neurons. *Behavioral Neuroscience, 117,* 760 – 773.

McPeek, R. M., Maljkovic, V., & Nakayama, K. (1999).

Saccades require focal attention and are facilitated by a short-term memory system. *Vision Research, 39*, 1555–1566.

Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes, 9,* 320–334.

Meck, W. H., & Church, R. M. (1984). Simultaneous temporal processing. *Journal of Experimental Psychology: Animal Behavior Processes, 10*, 1–29.

Meck, W. H., & Williams, C. L. (1997). Simultaneous temporal processing is sensitive to prenatal choline availability in mature and aged rats. *Neuroreport, 8*, 3045–3051.

Merchant, H. , Harrington, D. L., & Meck, W. H. (2013). Neural Basis of the Perception and Estimation of Time. *Annual Review of Neuroscience, 36,* 313–336.

Morgan, M. J., Giora, E., & Solomon, J. A. (2008). A single "stopwatch" for duration estimation: A single "ruler" for size. *Journal of Vision, 8,*1–8.

Nakayama, K., & Mackeben, M. (1989), Sustained and transient components of focal visual attention. *Vision Research, 29,* 1631–1647.

New, J. J., & Scholl, B. J. (2009). Subjective time dilation: Spatially local, object-based,or a global visual experience? *Journal of Vision, 9*(2), 1–11.

Olton, D. S., Wenk, G. L., Church, R. M., Meck, W. H. (1988). Attention and the frontal cortex as examined by simultaneous temporal processing. *Neuropsychologia, 26*, 307–318.

Pagano, S., & Mazza, V. (2012). Individuation of multiple targets during visual enumeration: New insights from electrophysiology. *Neuropsychologia, 50,* 754–761.

Palmer, J., Ames, C. T., & Lindsey, D. T. (1993). Measuring the effect of attention on simple visual search. *Journal of Experimental Psychology: Human Perception &Performance, 19,* 108–130.

Pang, K. C., Yoder, R. M., & Olton, D. S. (2001). Neurons in the lateral agranular frontal cortex have divided attention correlates in a simultaneous temporal processing task. *Neuroscience, 103*, 615–628.

Pashler, H. (1988). Familiarity and visual change detection. *Perception & Psychophysics, 44,* 369–378.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision, 10*, 437–442.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In: Bouma H,Bouwhuis DG, editors. Attention and performance x: Control of language processes. *NJ: Erlbaum*. 531–556.

Price-Williams, D. R. (1954). The kappa effect. *Nature, 173* (4399), 363–364.

Pylyshyn, Z. W., & Storm, R. W. (1988).Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial vision, 3*, 179–197.

Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science, 19*, 607–614.

Rousseau, L., & Rousseau, R. (1996). Stop-reaction time and the internal clock. *Perception & Psychophysics, 58*, 434–448.

Sagi, D., & Julesz, B. (1985). Detection versus discrimination of visual orientation. *Perception, 14*, 619–628.

Sears, C. R., & Pylyshyn, Z. W. (2000). Multiple object tracking and attentional processing. *Canadian Journal of Experimental Psychology, 54,* 1–14.

Spencer, R. K., & Catherine, L. R. (2013). Contextual influences of dimension, speed, and direction of motion on subjective time perception. Attention, *Perception, & Psychophysics, 75,* 161–167.

Staddon, J. E. R., & Higa, J. J. (1999). Time and memory: Towards a pacemaker-free theory of interval timing. *Journal of the Experimental Analysis of Behavior, 71,*

215–251.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerateddifferently? A limited-capacity preattentive stage in vision. *Psychological Review, 101*, 80–102.

Van, R. H., & Taatgen, N. A. (2008). Timing of multiple overlapping intervals: How many clocks do we have? *Acta Psychologica, 129*, 365–375.

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature, 428*, 748–751.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends in cognitive sciences, 7*, 483–488.

Wienera, M., Turkeltaubb, P., Coslett, H. B. (2010). The image of time: A voxel-wise meta-analysis. *NeuroImage, 49*, 1728–1740.

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual shortterm memory for objects. *Nature, 440*, 91–95.

Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of Vision, 7*(10), 1–5.

Yumoto, N., Lu, X., Henry, T. R., Miyachi, S., Nambu, A. , et al. (2011) A Neural Correlate of the Processing of Multi-Second Time Intervals in Primate Prefrontal Cortex. *PLoS ONE, 6,* e19168.

Are past and future symmetric in mental time line?

Xianfeng Ding[1,2], Ning Feng[1,2], Xiaorong Cheng[1,2], Huashan Liu[1,2], Zhao Fan[1,2,*]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior, Ministry of Education, Wuhan, China)

([2] School of Psychology, Central China Normal University, Wuhan, China)

Abstract A growing body of evidence has suggested that time, from early to late, or from past to future, was represented in a spatially oriented mental time line. However, little is known about its characteristics. The present study provided the first empirical evidence to explore the symmetry of spatial representations of past and future in the mental time line. Specifically, we compared the Spatial−Temporal Association Response Codes (STARC) effects and distance effects of past and future in four experiments. Results showed that for near past and near future, STARC effects were similar (Experiment 1). For distant past, the STARC effect was significant, but not for distant future (Experiment 2). Furthermore, the distance effect in the past was significantly stronger than in the future (Experiments, 3,4). These findings supported the idea that time points are not evenly distributed in mental time line. Spatial representations of the past and the future are asymmetric, and the spatial representation of past seems stronger than future. The logarithmic pattern of internal spatial representation of past or future is also discussed.

Keywords mental time line; asymmetry; STARC effect; distance effect

1 Introduction

Human beings often represent abstract concepts in concrete visual−spatial images. The spatial representation of number is a typical instance. It was suggested that numbers are represented in a continuous mental number line based on the extensive research on the Spatial−Numerical Association of Response Codes (SNARC) effect (Dehaene et al., 1993; Fischer et al., 2003; Schwarz and Keus, 2004; Hubbard et al., 2005; Nuerk et al., 2005 a,b;). Small numbers are represented at the left side of the line, while large numbers are represented at the right side. Time is also tightly connected with space. Specifically, researchers recently observed a SNARC like effect with time, which was labeled as the Spatial−Temporal Association of Response Codes (STARC) effect (Ishihara

et al., 2008;Vallesi et al., 2008). Therefore, time was analogically thought to be represented in a mental time line similar to the mental number line. In other words, time is represented in a continuous spatial line with a left–to–right orientation, where time flows from early to late, or from past to future (Bonato et al., 2012).

The mental time line hypothesis was supported by three categories of spatial–temporal congruency effects. The first type of congruency effect was based on temporal duration or interval. Vallesi et al.(2008) found that left responses were faster when associated with a short duration, while right responses were faster when associated with a long duration. The authors thought this compatible effect was a result of the spatial representation of elapsing time. When a temporal duration has to be estimated, elapsing time may be represented progressively from the left to the right. Then a short duration would be represented relatively to the left, and a long duration relatively to the right (See also Vallesi et al.,2011; Fabbri et al.,2012). Furthermore, duration estimation or judgment can also be influenced by spatial attention. Time duration would be underestimated when attention was directed to the left space, and be overestimated when attention was directed to the right space (Vicario et al., 2007; Frassinetti et al., 2009). A short duration was responded faster when a visual prime was in the left space, whereas a long duration was responded faster when a visual prime was in the right space (Di Bono et al.,2012). These findings suggested that elapsing time was represented in a mental time line from the left to the right.

The second type of congruency effect was based on temporal order. Santiago et al.(2010) found a space–time congruency effect when meaningful event sequences were presented by means of naturalistic movie clips or picture sequences. Order judgments between two events were faster when the left hand was used to respond "before" and the right hand to respond "after" than when

responded with the opposite mapping (see also Fuhrman & Boroditsky, 2010; Boroditsky et al., 2011). However, inherent and logical associations between successive stimuli may be confounded with temporal order in these studies. Some researchers found the STARC effect when using stimuli without logical or internal links et al. (2010) used nine words to explore the congruency effect of order and space in a serial learning paradigm. After an over–learned training phase, these nine words showed a SNARC similar effect for both order–relevant and order–irrelevant tasks. Moreover, a STARC effect based on mere temporal order was also found in working memory paradigms (Ding et al., 2014). These findings indicated that we could represent temporal order information in a spatial line.

The third type of congruency effect was based on abstract time words. Gevers et al. (2003, 2004) found that early months of a year or early days of a week were responded faster with the left key, whereas late months or days were responded faster with the right key. In addition, words referring to the past were responded faster with the left hand; words referring to the future were responded faster with the right hand (Santiago et al., 2007). This effect was found in both visual and auditory modalities (Lakens et al., 2011; Kong & You, 2012). Furthermore, time words can shift attention. Words related to the past can shift attention to the left and words related to the future can shift attention to the right in priming tasks (Weger and Pratt, 2008, 2009; Ouellet et al., 2010a). These findings indicated that the abstract concept of time could also be represented in a spatially oriented mental time line. Past is at the left side of this line, and future is at right side of this line. Time flows from past (left) to future (right) in the mental time line.

Taken together, three categories of evidence strongly supported that time can be represented in a mental time line. Time, from early to late, or from past to future, appears

to be represented in a left−to−right spatially continuous line. Previous studies have provided a lot of evidence for the existence of the mental time line, however, little is known about its characteristics. Are time points distributed evenly in this mental time line? Specifically, are past and future symmetric in the mental time line? The past is time we have actually experienced while the future is time that we have never experienced. Could this difference in reality for past and future lead to different spatial representations? A temporal asymmetry of past and future was suggested by evidence in some other paradigms. Representations of past events were associated with more specific details than representations of future events (D'Argembeau and Van der Linden, 2004, 2006; Addis and Schacter, 2008; Wang et al., 2011), and future events were more prototypical than past events (Kane et al., 2012). Thus we hypothesized that the spatial representations of past and future with the same temporal distance from the present are not identical but asymmetric in the mental time line. Examining symmetry would provide the first empirical evidence of characteristics of the mental time line, which is important to the construction of a theory of the spatial representation of time. Moreover, it will enhance the understanding of difference or similarity of past and future from the aspect of spatial linear representations and further provide a more specific spatial frame for past−future related theories.

To investigate this issue, we compared the spatial representations of past and future in the mental time line. According to previous studies, the STARC effect is the most important index of the spatial representation of time. Thus, we explored the symmetry of past and future in the mental time line by comparing the STARC effects of past and future. Another typical index of the spatial representation is the distance effect, in which the distance discrimination of two points located on a spatial line would be faster when the two points are far from each other than when they are near from each other (Moyer and Landauer,

1967; Dehaene et al.,1990; Dehaene, 1997). So we also compared the distance effects between past and future. We hypothesized that if the past and future are symmetric in the mental time line, there should be no differences on STARC effects and distance effects between past and future.

2 Experiment 1

Experiment 1 was designed to examine whether the STARC effects for near past (yesterday) and near future (tomorrow) were different. If the STARC effects were the same, it would support that spatial representations of near past and near future were symmetric in the mental time line.

2.1 Methods

2.1.1 Participants

Thirty six undergraduate students (13 male and 23 female) from Central China Normal University participated in the experiment for course credits. All participants signed a consent form according to the requirements of Institutional Review Board of CCNU. They were 19.6 years old on average (range 18 to 21). All participants were naive to the purpose of the experiment.

2.1.2 Stimuli and apparatus

Sixteen Chinese time words were used, 8 referring to the past time of yesterday (e.g., yesterday morning, yesterday afternoon, yesterday evening, etc.), and the other 8 referring to the future time of tomorrow (e.g., tomorrow morning, tomorrow afternoon, tomorrow evening, etc.). Time for the two groups of words were same except that past time was labeled with yesterday, and future time was labeled with tomorrow.

Participants viewed words on a 17−in. CRT screen (refresh rate 75 Hz and resolution 1280 × 1024 pixels) from a distance of 70 cm. The experiment procedure was programmed in Visual C++.

2.1.3 Experimental design

We used a $2 \times 2 \times 2$ mixed design. A between-subjects factor was Type of Time Words (yesterday vs. tomorrow) and two within-subjects factors were Temporal Position (early vs. late), and Response Congruence (congruence vs. incongruence, congruence means early stimuli responded with the left key and late stimuli responded with the right key; incongruence means early stimuli responded with the right key and late stimuli responded with the left key). Response times (RTs) and accuracy rates were dependent variables.

2.1.4 Procedure

Half of participants took part in the past or yesterday condition and the other half in the future or tomorrow condition. In the past condition, a trial started with a central fixation cross, lasting for 500 ms. Following that cross, a time word of yesterday was presented for 300 ms. Participant were required to judge whether the time of word was earlier or later than yesterday noon. For example, yesterday evening was later than yesterday noon. In one session, the participant pressed the left key (left arrow on the keyboard) if earlier and pressed the right key (right arrow on the keyboard) if later. In the other session participants responded in the opposite way. The order of the two sessions was counterbalanced across participants. The participants were required to respond as fast and accurately as possible using two fingers of the right hand only. After responses to stimuli, a 1000 ms blank separated one trial from another. Each session included 10 trials of practice and 4 blocks of 160 trials in the formal experiment.

In the future condition, the procedure was the same as in the past condition, except that the stimuli were time words of tomorrow and participants were required to judge whether the time was earlier or later than tomorrow noon.

2.1.5 Data analysis

Trials were treated as errors and discarded from the RT analyses if a response was made during the first 100 ms after the stimuli onset (anticipated responses), if the RT was slower than 2000 ms or no response was detected (delayed and null responses), or if the judgment was incorrect. RT outliers of correct trials (out of 3 standard deviations) were also filtered on a per-participant basis and excluded from analyses. A $2 \times 2 \times 2$ repeated measures MANOVA was performed both for accuracy rates and mean RTs of correct trials.

2.2 Results and discussion

The mean error rate in judging the time words was 2.31%. No significant effect was observed in the MANOVA concerning accuracy. The results of RTs indicated that the main effect of response congruence was significant (See Figure.1), $F_{(1,34)}=29.33$, $p<0.001$, partial $\eta^2=0.46$. The main effect of temporal position was significant, $F_{(1,34)}=8.13$, $p<0.01$, partial $\eta^2=0.19$. The main effect of type of time words was not significant, $F_{(1,34)}=0.94$, $p>0.05$. The interaction between response congruence and type of time words was not significant, $F_{(1,34)}=2.99$, $p>0.05$. The interaction between temporal position and response congruence was not significant, $F_{(1,34)}=0.01$, $p>0.05$. The interaction between temporal position and type of time words was not significant, $F_{(1,34)}=3.06$, $p>0.05$. The three-way interaction was not significant either, $F_{(1,34)}=0.32$, $p>0.05$.
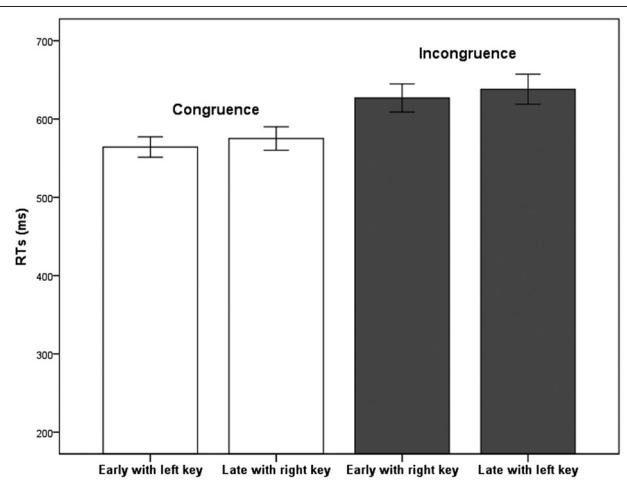
Figure1: Mean RTs in Experiment 1 as a function of response congruence and temporal position. Error bars represent standard errors of the means.

The results of Experiment 1 revealed a typical STARC effect. The RT of congruence (M=569 ms, SD=83) was significantly shorter than the RT of incongruence (M= 632 ms, SD=110). In other words, early time of a day was responded faster with the left key and late time was responded faster with the right key. However, there was no significant interaction effect between response congruence and type of time words. The STARC effects were the same for near past (yesterday) and near future (tomorrow). This result is consistent with the idea that the spatial representations of past and future in near space are symmetric in the mental time line.

3 Experiment 2

The STARC effects for near past and near future did not show any difference in Experiment 1, 2 was designed to further examine whether the STARC effects of distant past (last year) and distant future (next year) were different.

3.1 Methods

3.1.1 Participants

Thirty six undergraduate students (16 male and 20 female) from Central China Normal University participated in the experiment for course credits. All participants signed a consent form according to the requirements of Institutional Review Board of CCNU. They were 20.1 years old on average (range 18 to 21). All participants were naive to the purpose of the experiment.

3.1.2 Stimuli and apparatus

Sixteen Chinese time words of festivals were used，8 referring to past time of last year (e.g., Lantern Festival of last year, Labor Day of last year, National Day of last year, etc.) and the other 8 referring to future time of next year (e.g., Lantern Festival of next year, Labor Day of next year, National Day of next year, etc.). Times of the words were same except that past time was labeled with last year, and future time was labeled with next year.

3.1.3 Experimental design

The design was similar to the design in Experiment 1. The only change was the time words. Distant time words were used: last year vs. next year instead of yesterday vs. tomorrow.

3.1.4 Procedure

The procedure was the same as in Experiment 1. The task was to judge whether the time of word was earlier or later than July of last year or July of next year. For example, National Day of last year was later than last July. Half of participants took part in the past or last year condition and the other half in the future or next year condition.

3.1.5 Data analysis

Data analysis was the same as in Experiment 1.

3.2 Results and discussion

The mean error rate in judging the time words was 3.74%. No significant effect was observed in the MANOVA concerning accuracy. The results of RTs indicated that the only significant main effect was for response congruence, $F_{(1,34)}$=12.59, p<0.001, partial η^2=0.27. The interaction between response congruence and type of time words was significant, $F_{(1,34)}$=6.64, p=0.014, partial η^2=0.16 (See **Figure 2**). No other interaction was significant, ps>0.05. Simple effect analysis revealed that the response congruence effect was significant only for last year, $F_{(1,34)}$=18.76, p<0.001; but not significant for next year, $F_{(1,34)}$=0.47, p=0.49.

The results of Experiment 2 revealed a significant STARC effect as in Experiment 1. However, there was an interaction between response congruence and type of time words. The STARC effect was only found in distant past (last year) condition, but not in distant future (next year) condition. Early times of last year were responded faster with the left key and late times of last year were responded faster with the right key, while this was not true for next year. Thus, it suggested that spatial representations of
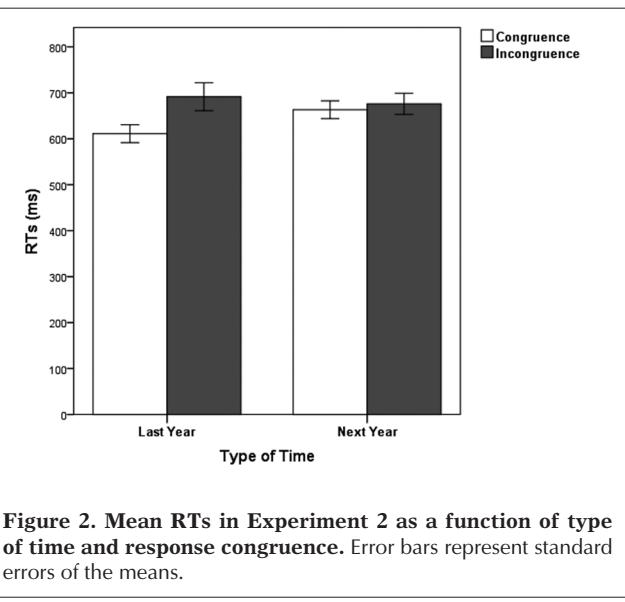
Figure 2. Mean RTs in Experiment 2 as a function of type of time and response congruence. Error bars represent standard errors of the means.

distant past and future were asymmetric in the mental time line, and the spatial representation of distant past was stronger than that of distant future.

4 Experiment 3

The results of Experiment 1, 2 showed that spatial representations of past and future in the mental time line were symmetric in near space, but asymmetric in distant space. If this was the case, a specific distance of past or future (two time points from either near space or distant space with same temporal distance) might be represented asymmetrically in the mental time line. Consequently, the distance effect for past and future might be different. Experiment 3 was designed to further examine whether the distance effects of past and future were the same.

4.1 Methods

4.1.1 Participants

Thirty six undergraduate students (14 male and 22 female) from Central China Normal University participated in the experiment for course credits. All participants signed a consent form according to the requirements of Institutional Review Board of CCNU. They were 21.2 years old on average (range 18 to 22). All participants were naive

to the purpose of the experiment.

4.1.2 Stimuli and apparatus

Sixteen Chinese time words were used. Eight words were near distance time words, 4 of them referring to past time of yesterday (e.g., yesterday morning, yesterday evening, etc.) and the other 4 referring to future time of tomorrow (e.g., tomorrow morning, tomorrow evening, etc.). Eight words were far distance time words, 4 of them referring to past time of last year (e.g., Labor Day of last year, National Day of last year, etc.) and the other 4 referring to future time of next year (e.g., Labor Day of next year, National Day of next year, etc.).

4.1.3 Experimental design

We used a $2 \times 2 \times 2$ mixed design. Three independent variables were temporal distance (near vs. far), between–subjects factor; type of time words (past vs. future), within–subjects factor; response congruence (congruence vs. incongruence, congruence means past time with left key and future time with right key; incongruence means past time with right key and future with left key), within–subjects factor. RTs and accuracy rates were dependent variables.

4.1.4 Procedure

Half of participants were in the far distance condition (last year and next year). A trial started with the central fixation cross, lasting for 500 ms. Following that cross, the time word of last year or next year was presented, lasting for 300 ms. Participants were required to judge whether the time of word was earlier or later than present. For example, National Day of last year was earlier than present. In one session, the participant pressed the left key (left arrow in keyboard) if earlier and pressed the right key (right arrow in keyboard) if later. In the other session participants were required to respond in the opposite way. The order of the two sessions was counterbalanced across participants. The participants were required to respond as fast and accurately as possible using two fingers of the right hand only. After responding to stimuli, a 1000

ms blank separated one trial from another. Each session included 10 trials of practice and 4 blocks of 160 trials in the formal experiment. The other half participants were in the near distance condition (yesterday and tomorrow). The procedure was the same as in the far distance condition.

4.1.5 Data analysis

Data analysis was the same as in Experiment 1.

4.2 Results and discussion

The mean error rate in judging the time words was 2.56%. No significant effect was observed in the MANOVA concerning accuracy. The results of RTs indicated that the main effect of response congruence was significant, $F_{(1,34)}=13.09$, $p<0.001$, partial $\eta^2=0.29$. The main effect of temporal distance was significant, $F_{(1,34)}=7.17$, $p=0.011$, partial $\eta^2=0.17$. The main effect of type of time words was not significant, $F_{(1,34)}=1.64$, $p>0.05$. The interaction between temporal distance and type of time words was significant (See **Figure3**), $F_{(1,34)}=5.51$, $p=0.025$, partial $\eta^2=0.14$. No other interactions were significant, ps>0.05 Simple effect analysis revealed that the distance effect was greater for past time, $F_{(1,34)}=9.69$, $p=0.004$; smaller for future time, $F_{(1,34)}=4.74$, $p=0.036$.

The results of Experiment 3 revealed a typical STARC effect as in Experiment 1, 2. Past time words were responded faster with the left key, whereas future time words were responded faster with the right key. Moreover, significant distance effect was observed. The time words of near distance (yesterday and tomorrow) were responded slower than far distance (last year and next year). Most importantly, there was an interaction between temporal distance and type of time words. The distance effect was greater for the past than for the future. These results further suggest that the spatial representations of past and future are asymmetric in the mental time line and that spatial representation of the past seems to be stronger than that of future.

However, some characteristics of time words, such as familiarity, could be confounded with temporal distance
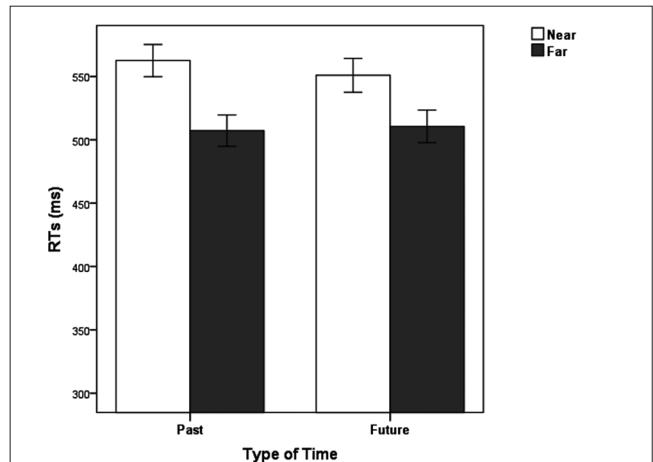
Figure 3. Mean RTs in Experiment 3 as a function of type of time and distance. Error bars represent standard errors of the means.

in Experiment 3. Separately we found that time words of near distance (yesterday or tomorrow) were more familiar than time words of far distance (last year or next year) in Experiment 3 through a questionnaire. Since familiar words were usually responded faster than unfamiliar words, it seemed that this distance effect could not be explained by familiarity of words. Nonetheless, we ran another experiment to balance the familiarity of time words in different distance.

5 Experiment 4

Experiment 4 was designed to balance the familiarity of time words in different distance. Both familiar and unfamiliar time words were chosen in near and far distance condition. Moreover, we changed the response way from left–right direction to an orthogonal up–down direction in the keyboard. If distance effects of past and future were actually different, the way of response would not affect it.

5.1 Methods

5.1.1 Participants

Eighteen undergraduate students (6 male and 12 female) from Central China Normal University participated in the experiment for course credits. All participants

signed a consent form according to the requirements of Institutional Review Board of CCNU. They were 19.2 years old on average (range 18 to 20). All participants were naive to the purpose of the experiment.

5.1.2 Stimuli and apparatus

Thirty two Chinese time words were used. Sixteen words were near distance time words, 8 referring to past time of yesterday (e.g., yesterday morning, yesterday evening, etc.) and the other 8 referring to future time of tomorrow (e.g., tomorrow morning, tomorrow evening, etc.). Sixteen words were far distance time words, 8 referring to past time of last year (e.g., Labor Day of last year, National Day of last year, etc.) and the other 8 referring to future time of next year (e.g., Labor Day of next year, National Day of next year, etc.). In a pre-experimental questionnaire investigation, 46 subjects rated the familiarity of 32 time words from 1 (unfamiliar) to 5 (familiar). The results showed that the familiarity of time words in different temporal distances were not significantly different ($M_{near}=3.48$ vs. $M_{far}=3.31$), $F_{(1,45)}=1.81$, $p=0.19$.

5.1.3 Experimental design

We used a $2 \times 2 \times 2$ within-subjects design. Three independent variables were time distance (near vs. far), type of time words (past vs. future), response key (up arrow vs. down arrow). RTs and accuracy rates were dependent variables.

5.1.4 Procedure

The procedure was similar to the previous experiments. Participants were required to judge whether the time of word was earlier or later than present. For example, Yesterday morning or National Day of last year was earlier than present. In one session, the participant pressed up key (up arrow in the keyboard) if earlier and pressed the down key (down arrow in the keyboard) if later. In the other session participants were required to respond in the opposite way. The order of the two sessions was counterbalanced across participants. The participants were required to respond as fast and accurately as possible using the middle finger of the right hand only. Each session included 10 trials of practice and 8 blocks of 320 trials in the formal experiment.

5.1.5 Data analysis

Data analysis was the same as in Experiment 1.

5.2 Results and discussion

The mean error rate in judging the time words was 3.13%. No significant effect was observed in the MANOVA concerning accuracy. The results of RTs indicated that the main effect of time distance was significant, $F_{(1,17)}=29.04$, $p<0.001$, partial $\eta^2=0.63$. The interaction between time distance and type of time words was significant (See **Figure 4**), $F_{(1,17)}=8.05$, $p=0.011$, partial $\eta^2=0.32$. All other main effects and interactions were not significant, ps>0.05. As the interaction between time distance and type of time words was significant, a simple effect analysis revealed that distance effect for past time, $F_{(1,17)}=24.57$, $p<0.001$; and a smaller effect for future time, $F_{(1,17)}=4.58$, $p=0.047$.

The results of Experiment 4 were similar as in Experiment 3. After controlling the familiarity of time words, the distance effect was still observed. The time words of near distance (yesterday and tomorrow) were responded slower than words of far distance (last year and next year). More important, the distance effect of past was also greater than that of future. Again, this result suggest that spatial representations of past and future were asymmetric in the mental time line and the spatial representation of the past seemed stronger than that of the future.

6 General discussion

Previous findings supported that representation of time flows from past to future in a continuous spatial line with a left-to-right orientation. The present study provided the first empirical evidence for a fundamental characteristic of the mental time line: Are the spatial representations of past
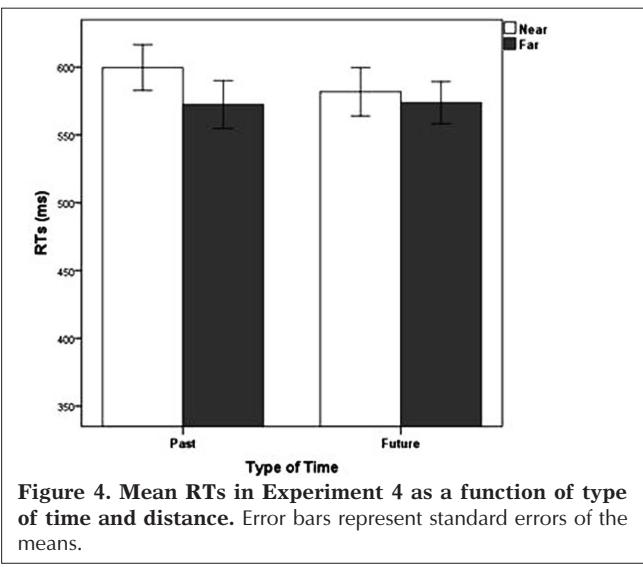
Figure 4. Mean RTs in Experiment 4 as a function of type of time and distance. Error bars represent standard errors of the means.

and future symmetric in the mental time line?

In Experiment 1, we compared STARC effects under near past and near future conditions. As expectely, a typical STARC effect was observed. Early time was responded faster with the left key, whereas late time was responded faster with the right key. Moreover, STARC effects were the same between yesterday and tomorrow. This result indicated that spatial representations of past and future were symmetric in near past and near future in the mental time line. In Experiment 2, STARC effects were further compared under distant past and distant future condition. However, the STARC effect was only observed in the distant past condition, not in the distant future condition. This result showed that spatial representations of distant future and distant past were asymmetric in the mental time line and the spatial representation of past seemed stronger than that of future, as the STARC effect disappeared in distant future. Therefore, it seemed that past and future in the mental time line were symmetric in near space, but not in distant space.

In Experiment 3, 4, distance effects were compared under past and future conditions. Results showed that there were both a significant STARC effect and a distance effect in Experiment 3. Past time words were responded faster with the left key, whereas future time words were responded faster with the right key. When compared with

the present, time points in the far distance (last year or next year) were responded faster than in the near distance (yesterday and tomorrow). Moreover, the distance effect in the past was greater than in the future. The same result was observed even when the response was changed to an orthogonal direction in Experiment 4. These results about distance effects support the idea that past and future are represented asymmetrically in the mental time line. Again, the spatial representation of the past seemed stronger than that of the past, as the distance effect for the future was smaller than that of the past.

These findings revealed that the mental time line is not evenly distributed and the past and future were asymmetric in the mental time line. According to our results, the STARC effect was significant for the distant past, but not for the distant future. And the distance effect was stronger for the past than for the future. Why is past different from future? A possible reason is that the past is more concrete or clear than the future. The past is time we have actually experienced. It is true and available for us. We can store the past information in our memory and retrieve it. The construction of the representation of past could be based on real events. However, the future is not yet true. It is obscure, abstract and fictional for us. The construction of the representation of future could only be based on fictional events.

This temporal asymmetry was in line with findings in some other paradigms. For example, Vallesi et al. (2008) found a stronger leftward representation of short durations than rightward representation of long durations in their fourth experiment at least numerically, suggesting a similar asymmetrical effect though no further statistics were provided in this literature. In a neuroimaging study, Okuda et al. (2003) found that anteromedial frontal pole and medial temporal areas showed a significant effect of temporal distance from the present. Specifically, the increase in brain activity in the left parahippocampal gyrus (BA 36) from the near future task to the far future task was

smaller than that from the near past task to the far past task. It suggested that distance effect was smaller for the future than for the past task. Addis and Schacter (2008) found that representations of past events were associated with more specific details than representations of future events (See also D'Argembeau and Van der Linden, 2004, 2006; Wang et al., 2011). Future events were also found more prototypical than past events (Kane et al., 2012). These findings supported that the representation of past was more concrete or clear than that of future.

Although the future is different from the past, it is similar in that the past and future are represented as a spatial line. STARC effects and distance effects were found for both past and future. These findings were consistent with constructive episodic simulation hypothesis (Schacter and Addis, 2007a,b). Since future is what we have never experienced before, how do we construct representations that we never truly experienced? Schacter and Addis (2007a,b) thought that one of important functions of constructive episodic memory is to allow individuals to simulate or imagine future episodes. We construct future based on past that we have experienced. Therefore, there should be considerable overlap in the psychological and neural processes involved in remembering the past and imagining the future. Neuroimaging evidence from Mental Time Traveling (MTT) supported that underlying neural mechanisms for past and future were similar. Remembering the past and imaging the future may activate same brain areas (Okuda et al., 2003; Schacter et al., 2007; Szpunar et al., 2007).These findings have led to the concept of the prospective brain and an idea that a crucial function of the brain is to use stored information to imagine, simulate and predict possible future events (Schacter et al., 2007).

The present study further indicated that the internal spatial representation of past or future seemed to be unevenly−distributed in the mental time line. Taken the results of Experiment 1,2 together, the STARC effect

was significant in tomorrow (near future) condition but not in next year (distant future) condition. Thus, the spatial representation of the near future seemed stronger than that of distant future in the mental time line. This finding was consistent with some studies about a loglinear characteristic of mental time. When participants were asked to judge whether an event of past or future was before or after an imagined "location" on the time line, the reaction time of this "self−projection" decreased logarithmically as the temporal distance between this imagined location and the location of another imagined event from the time line increased (Arzy et al., 2009). In addition, logarithmic curves were also found to fit the relation between temporal distance and memory, as the distribution of the correct recall of events from different points in time was logarithmic (Rubin and Schulkind, 1997; Spreng and Levine, 2006).

This logarithmic pattern suggested that time points (past or future) near the present were relatively sparse, and time points far from the present were relatively dense in the mental time line. In other words, if a set of two time points is near the present, the spatial distance would be larger, as the reaction times decrease sharply with the increase of the temporal distance; if the set of two time points was far from the present, the spatial distance would be smaller, as the reaction times decrease slowly with the increase of the temporal distance. Interestingly, a logarithmic pattern was also found in the mental number line. Humans map numbers into space line in logarithmic scaling (Siegler and Booth, 2004; Dehaene and Cohen, 1995; Dehaene et al., 2008). The mechanisms of processing time and number are similar, in line with the A Theory of Magnitude, i.e. ATOM (Walsh, 2003; Bueti and Walsh, 2009). Nonetheless, we should be cautious with these inferences and further research is needed on this logarithmic pattern of spatial representation in the mental time line.

Finally, it was worth noting that culture may play an

important role on the representations of past and future. For instance, reading and writing habits can change the direction of the mental time line. The direction is from left to right in English or Italian speakers, whereas the direction is from right to left for Arabic or Hebraic speakers (Fuhrman and Boroditsky 2010; Ouellet et al., 2010b; Vallesi et al., 2014). Most importantly, culture may shape the characteristics of the mental time line. Westerners exhibit greater episodic specificity than East Asians (Wang et al., 2011). The spatial representations of mental time for Westerners might be stronger than those of East Asians. Age and gender may also influence the representations of past and future. Older adults generated fewer internal details than younger adults for both past and future events (Addis and Schacter, 2008). Women exhibit greater episodic specificity than men for both past and future events (Wang et al., 2011).

In summary, the present study provided the first empirical evidence for the characteristics of the mental time line. Time points are not evenly distributed in the mental time line. The differences on STARC effect and distance effect supported that the spatial representations of past and future are asymmetric in the mental time line. And the spatial representation of past seemed stronger than that of future. Importantly, future studies should focus more on the characteristics of internal spatial representation of past or future (e.g. logarithmic pattern) and how the culture and some other factors shape the characteristics of the mental time line.

References

Addis,D.R.,and Schacter,D.L.(2008).Effects of detail and temporal distance of past and future events on the engagement of a common neural network. *Hippocampus 18*, 227–237.doi:10.1002/hipo.20405

Arzy,S.,Adi–Japha,E.,and Blanke,O.(2009).The mental time line:an analogue of the mental number line in the mapping of life events. *Conscious.Cogn. 18,* 781–785. doi:10.1016/j.concog.2009.05.007

Bonato,M.,Zorzi,M.,andUmilt à ,C.(2012).When time is space:evidence for a mental time line. *Neurosci. Biobehav.Rev. 36*, 2257–2273.doi: 10.1016/ j.neubiorev.2012.08.007

Boroditsky,L.,Fuhrmana,O.,and McCormicka,K.(2011). Do Englisha nd Mandarin speakers think about time differently? *Cognition 118*, 123–129.doi: 10.1016/ j.cognition.2010.09.010

Bueti,D.,and Walsh,V.(2009).The parietal cortex and the representation of time, space,number and other magnitudes. *Philos.Trans.R.Soc.Lond.B Biol.Sci. 364,* 1831–1840.doi:10.1098/rstb.2009.0028

D'Argembeau,A.,and van der Linden,M.(2004). Phenomenal characteristic sassociated with rojecting oneself back into the past and forward into the future: influence of valence and temporal distance. *Conscious. Cogn.13*, 844–858.doi: 10.1016/j.concog.2004.07.007

D'Argembeau,A.,and van der Linden,M.(2006). Individual differences in the phenomenology of mental time travel:the effects of vivid visual imagery and emotion regulation strategies. *Conscious.Cogn. 15*, 342–350.doi: 10.1016/j.concog.2005.09.001

Dehaene, S.(1997). *The Number Sense:How the Mind Creates Mathematics*. New York,NY:Oxford University Press.

Dehaene, S.,Bossini,S.,and Giraux,P.(1993).The mental representation of parity and number magnitude. *J. Exp.Psychol.Gen. 122*, 371–396.doi:10.1037/0096– 3445.122.3.371

Dehaene, S.,and Cohen,L.(1995).Towards an anatomical and functional model of number processing. *Math. Cogn. 1,* 83–120.

Dehaene, S.,Dupoux,E.,and Mehler,J.(1990).Is numerical comparison digital? Analogical and symbolic effects in two–digit number comparison. J. Exp.Psychol.Hum.

Percept.Perform.16,626–641.doi:10.1037/0096–1523. 16.3.626

Dehaene, S.,Izard,V.,Spelke,E.,and Pica,P.(2008).Log or linear?Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science 320,* 1217–1220.doi:10.1126/science.1156540

Di Bono,M.G.,Casarotti,M.,Priftis,K.,Gava,L.,Umilt à ,C .,and Zorzi,M. (2012). Priming the mental time line. J. Exp.Psychol.Hum.*Percept.Perform. 38*, 838–842. doi:10.1037/a0028346

Ding,X.,Cheng,X.,Fan,Z.,and Liu,H.(2015).Is elapsing time really recoded into spatial linear representation in working memory? *Exp.Psychol. 62*,11–19.doi: 10.1027/1618–3169/a000269

Fabbri,M.,Cancellieri,J.,and Natale,V.(2012).The A Theory Of Magnitude (ATOM)model in temporal perception and reproduction tasks. *Acta Psychol. 139,* 111–123. doi:10.1016/j.actpsy.2011.09.006

Fischer,M.H.,Castel,A.D.,Dodd,M.D.,and Pratt,J.(2003). Perceiving numbers causes spatial shifts of attention. *Nat.Neurosci. 6*, 555–556.doi:10.1038/nn1066

Frassinetti,F.,Magnani,B.,and Oliveri,M.(2009).Prismatic lenses shift time perception. *Psychol.Sci. 20*, 949–954. doi:10.1111/j.1467–9280.2009. 02390.x

Fuhrman,O.,and Boroditsky,L.(2010).Cross–cultural differences in mental representations of time:evidence from an implicit nonlinguistic task. *Cogn.Sci. 34*, 1430–1451. doi:10.1111/j.1551–6709.2010.01105.x

Gevers,W.,Reynvoet,B.,and Fias,W.(2003).The mental representation of ordinal sequences is spatially organized. *Cognition 87*, 87–95.doi:10.1016/S0010–0277(02)00234–2

Gevers,W.,Reynvoet,B.,and Fias,W.(2004).The mental representation of ordinal sequences is spatially organized:evidence from days of the week. *Cortex 40*, 171–172. doi:10.1016/S0010–9452(08)70938–9

Hubbard,E.M.,Piazza,M.,Pinel,P.,and Dehaene,S.(2005).

Interactions between number and space in parietal cortex. *Nat.Rev.Neurosci. 6*, 435–448.doi: 10.1038/ nrn1684

Ishihara,M.,Keller,P.E.,Rossetti,Y.,and Prinz,W.(2008). Horizontal spatial rep representations of time:evidence for the STEARC effect. *Cortex 44*, 454–461.doi: 10.1016/j.cortex.2007.08.010

Kane,J.,Van Boven,L.,and Mcgraw,A.P.(2012).Prototypical prospection:future events are more prototypically represented and simulated than past events. Eur. *J. Soc.Psychol. 42*, 354–362.doi:10.1002/ejsp.1866

Kong,F.,and You,X.(2012).Space–time compatibility effects in the auditory modality. *Exp.Psychol. 59*, 82–87.doi:10.1027/1618–3169/a000129

Lakens,D.,Semin,G.R.,and Garrido,M.V.(2011).The sound of time:cross– modal convergence in the spatial structuring of time. *Conscious.Cogn. 20*, 437–443. doi:10.1016/j.concog.2010.09.020

Moyer,R.S.,and Landauer,T.K.(1967).Time required for judgments of numerical inequality. *Nature 215*, 1519–1520.doi:10.1038/2151519a0

Nuerk,H–C.,Bauer,F.,Krummenacher,J.,Heller,D.,and Willmes,K.(2005a).The power of the mental number line:how the magnitude of unattended numbers affects performance in an Eriksen task. *Psychol.Sci. 47*, 34–50.

Nuerk,H.C.,Wood,G.,and Willmes,K.(2005b).The universal SNARC effect. *Exp. Psychol. 52,* 187–194. doi:10.1027/1618–3169.52.3.187

Okuda, J.,Fujii,T.,Ohtake,H.,Tsukiura,T.,Tanji,K.,Suzuki ,K.,etal.(2003). Thinking of the future and the past:the roles of the frontal pole and the medial temporal lobes. *Neuroimage 19*, 1369–1380.doi:10.1016/S1053–8119(03)00179–4

Ouellet, M.,Santiago,J.,Funes,M.J.,and Lupianez,J.(2010a). Thinking about the future moves attention to the right. J. Exp.Psychol.Hum.*Percept.Perform. 36*, 17–24.

doi:10.1037/a0017176

Ouellet, M.,Santiago,J.,Israeli,Z.,and Gabay,S.(2010b).Is the future the right time? *Exp.Psychol. 57*, 308–314. doi:10.1027/1618–3169/a000036

Previtali,P.,deHevia,M.D.,and Girelli,L.(2010).Placing order in space: the SNARC effect in serial learning. *Exp.BrainRes. 201*, 599–605.doi: 10.1007/s00221–009–2063–3

Rubin, D. C., & Schulkind, M. D. (1997). The distribution of autobiographical memories across the lifespan. *Mem. Cognit. 25,* 859–866. doi: 10.3758/BF03211330

Santiago, J., Lupianez, J., Perez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychon. Bull. Rev. 14*, 512–516. doi: 10.3758/BF03194099

Santiago, J., Roman, A.,Ouellet, M., Rodriguez, N., & Perez–Azor, P. (2010). In hindsight, life flows from left to right. *Psychol. Res. 74,* 59–70. doi: 10.1007/s00426–008–0220–0

Schacter, D. L., & Addis, D. R. (2007a). The cognitive neuroscience of constructive memory: remembering the past and imagining the future. *Philos. Trans. R. Soc. Lond B Biol. Sci. 362,* 773–786. doi: 10.1098/rstb.2007.2087

Schacter, D. L., & Addis, D. R. (2007b). The ghosts of past and future. *Nature 445*, 27. doi: 10.1038/445027a

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: the prospective brain. *Nat. Rev. Neurosci. 8*, 657–661. doi: 10.1038/nrn2213

Schwarz, W., & Keus, I. M. (2004). Moving the eyes along the mental number line comparing SNARC effects with saccadic and manual response. *Percept. Psychophys. 66*, 651–664. doi: 10.3758/BF03194909

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev. 75,* 428–444. doi: 10.1111/j.1467–8624.2004.00684.x

Spreng, R. N., & Levine, B. (2006). The temporal distribution of past and future autobiographical events across the lifespan. *Mem. Cognit. 34*, 1644–1651. doi: 10.3758/BF03195927

Szpunar, K. K., Watson, J. M., & McDermott, K. B. (2007). Neural substrates of envisioning the future. *Proc. Natl. Acad. Sci. U.S.A. 104,* 642–647. doi: 10.1073/pnas.0610082104

Vallesi, A., Binns, M. A., & Shallice, T. (2008). An effect of spatial–temporal association of response codes: understanding the cognitive representations of time. *Cognition 107,* 501–527. doi: 10.1016/j.cognition.2007.10.011

Vallesi, A., McIntosh, A. R., & Stuss, D. T. (2011). How time modulates spatial responses. *Cortex 47,* 148–156. doi: 10.1016/j.cortex.2009.09.005

Vallesi, A., Weisblatt, Y., Semenza, C., & Shaki, S. (2014). Cultural modulations of space–time compatibility effects. *Psychon. Bull. Rev. 21*, 666–669. doi: 10.3758/s13423–013–0540–y

Vicario, C. M., Caltagirone, C., & Oliveri, M. (2007). Optokinetic stimulation affects temporal estimation in healthy humans. *Brain Cogn. 64*, 68–73. doi: 10.1016/j.bandc.2006.12.002

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *Trends Cogn. Sci. 7*, 483–488. doi: 10.1016/j.tics.2003.09.002

Wang, Q., Hou, Y., Tang, H., & Wiprovnick, A. (2011). Travelling backwards and forwards in time: culture and gender in the episodic specificity of past and future events. *Memory 19*, 103–109. doi: 10.1080/09658211.2010.537279

Weger,U.W.,& Pratt,J.(2008).Time flies like an arrow: space–time compatibility effects suggest the use of a mental timeline. *Psychon. Bull. Rev. 15,* 426–430. doi: 10.3758/PBR.15.2.426

Weger, U. W., & Pratt, J. (2009). Time–words guide spatial attention. *J. Vis. 7*, 447. doi: 10.1167/7.9.447

The role of sustained posterior brain activity in the serial chaining of two cognitive operations: A MEG study

Zhao Fan[1,2], Suresh D. Muthukumaraswamy[3], Krish D. Singh[3], Kimron Shapiro[4]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan, China)

([2] School of Psychology, Central China Normal University, Wuhan,China)

([3] CUBRIC, School of Psychology, Cardiff University, Cardiff,UK)

([4] School of Psychology, Bangor University, Bangor, UK)

Abstract   A fundamental necessity in human cognition is to link sequential mental operations where appropriate execution of the second task requires input from the first. The present study explores the neural basis of such 'chaining' using a novel psychological refractory period(PRP) task. Participants were required to make speeded responses to two sequential visual tasks that were chained or independent. Magnetoencephalography (MEG) signals were recorded simultaneously to reveal the brain's response to these similar but fundamentally different conditions. RTs to Task 1 and 2 were slower in the Chained condition, and their temporal coupling weakened, relative to the Independent condition. MEG analysis of the accompanying event−related fields (ERFs) revealed an increased sustained posterior component in the Chained condition beginning approximately 350 ms after Task 2 onset and lasting for 450 ms. Beamformer localisation of this ERF effect revealed a left hemisphere source near the junction of the temporal, parietal and occipital lobes. These results extend our understanding of the behavioural and corresponding neural mechanisms required by everyday decision making.

Keywords   Chained sequential operations; Dual task; Magnetoencephalography (MEG); PRP

1 Introduction

Experimental investigations of making a simple decision, for example, which of two buttons to press to signify a chosen response in a two−alternative situation, have yielded detailed behavioural and neural insights into the cognitive operations that underlie such a task. But rarely do real world decisions involve such few choices, indeed often they involve antecedent or consequent decisions that form a chain of cognitive operations. The purpose of the present report is to understand the neural basis of such mental processes, which are crucial to many forms of decision making.

The consequences of performing two temporally adjacent cognitive operations have been studied for many years. Indeed, several lines of behavioural observations, such as the psychological refractory period (PRP; cf. Pashler, 1984; Sigman & Dehaene, 2005, 2008) and the attentional blink (AB; Raymond, Shapiro, & Arnell, 1992), converge to highlight the 'bottleneck' that arises when we attempt to perform multiple sequential operations. The term 'bottleneck' refers to the finding that aspects of the second task are prevented from execution until aspects of the first are completed. Although both PRP and the AB have been studied extensively, both share the commonality that each of the sub-tasks comprising them is independent. By 'independent' we refer to the fact that the decision from the first sub-task has no informational bearing on the second. Whereas the AB and PRP phenomenon do characterise many real-world cognitive tasks, they fall short in yielding insight into the more complex, yet frequently encountered tasks, which require the input from one to direct decision-making on the second. We refer to such tasks as 'chained', an example of which arises when a tennis player is required to judge whether to rush the net or remain at the baseline, then having decided to rush the net now must decide whether to prepare for the returned ball to come high or low.

In 1936, Alan Turing published his seminal essay "On computable numbers" and introduced the Turing machine (1936), which served as a computational blueprint for the subsequent invention of the electronic computer. Turing explicitly aimed to capture the basic operations of the human mind. Yet surprisingly, 70 years after Turing's paper, we still have little knowledge of the brain mechanisms by which humans carry out even the simplest of sequential computations postulated in Turing's framework. The present report is an attempt to understand the neural basis of the cost of performing two sequential chained operations over and above the cost of performing the same two independent operations as assessed by the standard PRP paradigm.

A recent behavioural study (Fan, et al., 2011) reveals a number of important aspects of this additional cost using a novel PRP paradigm. In two experiments, we required participants to give speeded responses to two sequential visual tasks separated by variable SOAs (Stimulus Onset Asynchrony). By requiring information to be passed from Task 1 to Task 2 in one condition and comparing it to another condition where the same two tasks did not require information from the first to perform the second, we explored the behavioural costs of chaining. Our results revealed two costs in the chained condition over and above classical dual-task PRP costs: (1) an altered distribution of response times RTs (particularly for the second task) in terms of an increased mean and variance; and (2) a disrupted temporal coupling between the first and the second tasks, as evidenced by a reduced correlation between the response times of both tasks. Based on these findings, we proposed that serial chaining costs originate from two sources: the first is an early stage of task setting, and the second, a later stage of information buffering and result-passing. Our results are consistent with a recent report (Sackur & Dehaene, 2009), which provides evidence that serial chaining is a slow and effortful process that consumes central processing resources and requires conscious control. Together these two behavioural reports shed light on the behavioural correlates of serial chaining induced by limits on human information processing.

Interest in the neural correlates of dual-task processing has been intense in recent years, thus we will attempt only a brief overview to set the stage for the current investigation. Brisson & Jolicoeur (2007a, 2007b) revealed the attenuation of an ERP component (N2pc) known to index the deployment of spatial attention as the SOA was reduced between the two tasks comprising a standard PRP paradigm. In addition to showing that this reduction was

due to central bottleneck constraints as are indexed by the PRP paradigm (see also Lien, Croswaite, & Ruthruff, 2011), Brisson & Jolicoeur. also found the onset latency of another ERP component (sustained posterior contralateral negativity, or SPCN[1]) to be delayed with decreasing Task1–Task2 SOA. The SPCN has been shown to index visual short–term memory (VSTM) capacity limits and probably reflects maintenance of the array in VSTM. As we shall show here, the present investigation finds the SPCN distinguishes between chained and independent tasks.

The SPCN component remains a current topic of investigation, as evidenced by recent reports from Jolicoeur's lab (Robitaille, Grimault, & Jolicoeur, 2009; Robitaille, et al., 2010) who examined the SPCN using different approaches, including for the present purpose magnetoencephalography (MEG). The MEG equivalent, referred to as SPCM, was determined to arise from a cortical network including bilateral parietal loci, likely intraparietal/intraoccipital cortex, and contralateral parietal sources, paralleling findings from neuroimaging (e.g., Xu & Chun, 2006). Moreover, as its equivalent (SPCN) in the electrophysiological domain, it was elicited by VSTM maintenance.

Although we have an understanding of the behavioural costs, we have surprisingly little knowledge of the brain mechanisms underlying the process by which multiple cognitive operations are assembled to implement serial chaining, in spite of the fact that such operations are vital to support the daily requirements of human decision making. Attempts to understand the locus and function of the central bottleneck are still being sought, for example, recent reports have specified brain areas involved in a "unified central bottleneck" (Tombu et al., 2011) and have even shown the bottleneck can be overcome

with practice (Oberauer & Kliegl, 2004). The present report seeks to characterise the cognitive and neural mechanisms of serial chaining by comparing a variant of the PRP task, which requires transferring information from the first to the second task, to a traditional PRP task, which does not require information passing. We achieve this goal by conducting simultaneous behavioural and magnetoencephalographic (MEG) measurements to test the hypothesis that serial chaining requires additional resources and recruits additional brain areas relative to serial independent tasks. We employ the same PRP paradigm as was used successfully in a prior behavioural report (Fan et al., 2011).

2 Methods

2.1 Participants

Twenty–six right–handed participants took part in this experiment, with five excluded due to excessive head / body movement during MEG recording and thus lack of enough valid trials per cell during subsequent data analysis. Twelve females and nine males were included in the data analyses, aged from 19 to 38. All participants had normal or corrected–to–normal vision, with no history of visual disorders and gave their informed consent to participate.

2.2 Stimuli

The participant was seated in a shielded–MEG recording room that was dark except for the display. All stimuli were produced within MATLAB (The Mathworks, Inc.) using the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), and displayed on a Mitsubishi Diamond Pro 2070 colour monitor running at 60 Hz and located 208 cm from the participant's eyes (10.7 ° × 8.0 ° visual angle).

---

[1] The SPCN appears to be the same component as the contralateral delay activity (CDA) ERP component (Vogel & Machizawa, 2004).
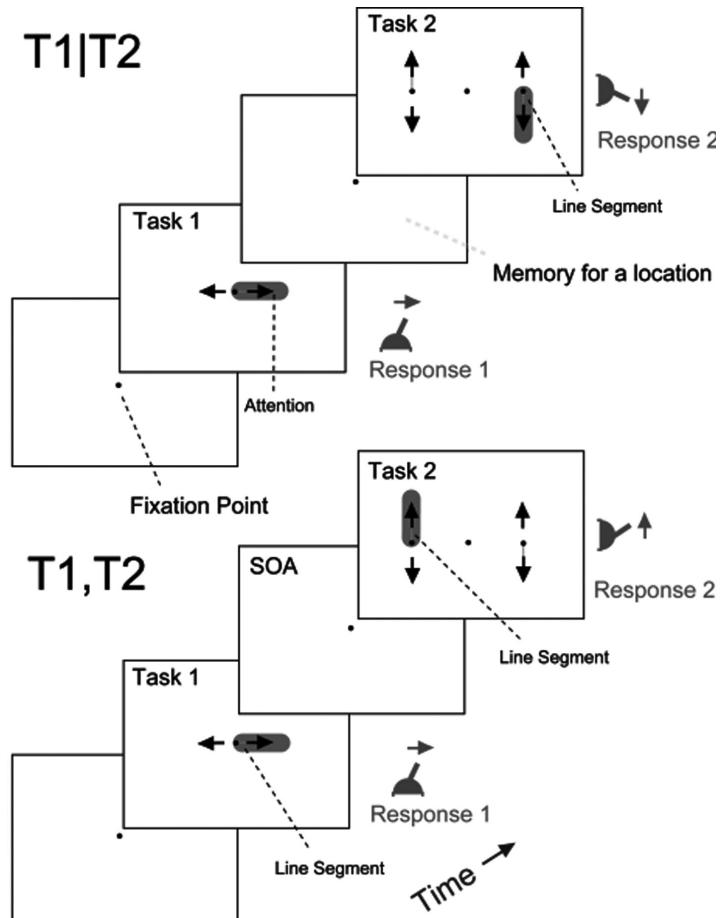
Figure 1. Schematic representation of the tasks used to study serial processing. Upper: in Task 1 of the Chained task (T1|T2), the participant saw two horizontal arrows; one of these arrows was connected by a low contrast line segment to the fixation point. The participant was instructed to push a joystick as quickly as possible with the left hand to the corresponding side of the connected line segment and remember this location. For Task 2 two vertical arrows and one line segment appeared on both sides simultaneously and the participant indicated the direction of the arrow connected to the T2 line segment of the remembered side by pushing a joystick with his/her right hand. The grey region (figure only) indicates the anticipated distribution of attention as well as the location that had to be remembered. Lower, in the independent task (T1,T2), Task 1 is identical. Task 2 was independent from Task 1 because the participant reported either the left or right side T2 line segment discrimination for a block of trials. In this example, the participant must report the left side T2 line segment and it involves a switch of spatial location from the left side to the right side (a trial in the Independent−switch condition).

Participant's responses were recorded by two MEG−compatible fORP fibre−optic joysticks (Current Designs Inc.).

In the first task (Task 1; T1), two horizontal arrows and a single (low−contrast) line segment were presented (see Figure 1). Each arrow contained an equal−sided triangle head (0.91° length for the sides and 0.82 ° length for the base) and a horizontal arrow line (1.1 ° length). The apex of each arrow head was 2.91° away from the central fixation point horizontally. The luminance of the background and that of the arrows were 80 and 1.2 cd/m^2, respectively. The contrast (C) of the (low−contrast) line segment was −15.

This contrast was defined as Weber contrast (C=100*(L−Lb)/Lb), that is, with reference to the mean luminance of the background, Lb. For each trial, Target 1 was to be judged "left" or "right" by virtue of which end of the arrow line was connected to the central fixation point by a horizontal low−contrast line segment (1° length). For T1, participants were instructed to push the left−hand joystick horizontally (left or right) according to the location of the target as quickly and also as accurately as possible after the target appeared.

The stimuli in the second task (Task 2; T2) were composed of four vertically oriented arrows and two vertical

line segments/targets (see Figure 1). The size of the arrows and line segments were the same as those in T1, however the orientation and presentation locations were different. There were two targets in T2, each with two possibilities for their locations (either top-left and bottom-right quadrants or top-right and bottom-left quadrants of the visual field). By this arrangement, one target was always on the left and the other on the right side of the display. For T2 participants were instructed to respond to either the left- or right- side target and push the right-hand joystick vertically up or down, depending on whether the selected target was in the upper or lower visual field.

The additional information about which target (left or right) to which a response was required in T2 varied. In the Chained condition (T1|T2; upper display in Figure 1), this extra precuing information was derived from the result of T1. For example, if the single target in T1 was located on the left side of the display, then a discrimination had to be made on the left side of the display in T2, and vice versa. By contrast, in the Independent group, (T1,T2; lower display in Figure 1), this precuing information (left or right), was provided at the beginning of each (Independent) block. For example, in the beginning of an Independent-left block, participants were told through screen instructions to respond to the left side target in T2 during this block, and vice versa for an Independent-right block.

Half of the trials were dual-task trials requiring two tasks to be performed sequentially. The other half were single-task trials requiring the first task only. This allowed us in the MEG analysis to isolate the effects due to T2 from T1.

2.3 Design

A two-factor within-snbjects design was used in this experiment. One factor was whether a trial was a single-task trial (T1 only) or dual-task trial (T1 and T2). The other factor was the relationship (Chained vs. Independent) between the two tasks comprising a dual-task trial. The SOA between T1 and T2 stimuli was kept constant at 100ms. The location(s) of the target(s) in both tasks was counterbalanced across trials. Each combination of trial type (single or dual task) and dual-task relationship (Independent or Chained) was repeated 80 times, resulting in a total of 320 trials in each experimental session.

The experimental session was divided into eight blocks with 40 trials in each block. Half of the trials in each block were single-task trials and the other half were dual-task trials (randomly mixed) and the participant did not know in advance whether a trial was a single- or dual- task trial. Four out of eight blocks were Chained blocks, each containing 20 dual-task trials of the Chained condition (A|B) and 20 single-task trials. The other four blocks were Independent blocks, each containing 20 dual-task trials of the Independent condition (A, B) and 20 single-task trials. The left target of T2 was predefined in two of four Independent blocks, and in the other two blocks the right target of T2 was predefined. The order of Chained block, left pre-defined Independent block, and right pre-defined Independent block was counterbalanced across participants.

Unlike in the Chained condition where T1 and T2 by definition occurred in the same spatial location, in the Independent condition (T1,T2) the two tasks occurred on half the trials in the same location and on the other half in a different location. Thus the Independent condition had two subconditions, Independent-stick (without a location switch-called "Stick") and Independent-switch (with a location switch-called "Switch"), both of which were randomly mixed within each Independent block of trials. Since the Chained condition involved no switch of spatial attention, the Independent-stick condition is a more appropriate comparison to the Chained condition.

2.4 Procedure

Participants were instructed to respond as quickly and also as accurately as possible. Response order in the dual-task trials was also emphasized in that a T1 response

should be always followed by a T2 response. A verbal instruction was presented on the display in the beginning of each block to inform participants in which block they were (Chained, Independent Left, or Independent Right). At the beginning of each trial, participants pressed a button on the joystick to initiate the trial whenever they were ready. A black fixation point (radius 0.16 °) was then presented immediately in the centre of the display and remained until the feedback was presented at the end of each trial. Participants were instructed to perform the tasks while maintaining the central fixation point. The T1 stimulus appeared 1000 ms after the onset of the fixation point and lasted for 100 ms. For a dual-task trial, the onset of the T2 stimulus was always 100ms after onset of T1, i.e., immediately after T1 offset. T2 stimuli remained on the display until participants made their response. Feedback was provided 1000ms after the occurrence of the T2 response and the fixation point was replaced by the feedback. Feedback was provided by two coloured dots appearing to the left and right side of the location previously occupied by the fixation point. The dot on the left side indicated the accuracy of T1, and the dot on the right side indicated the accuracy of T2. If the participant's response to T1 was correct, the left dot was green (red if it was incorrect). The same rule applied for the right dot. The left dot was yellow if participants made no response to T1. The feedback dots remained on the display for 1000 ms before the initiation of the next trial. For a single-task trial, the fixation point remained on the display for another 1800ms from the offset of T1 but before the onset of the feedback. The right feedback point was green in default and only became red if participants mistakenly pushed the right-hand joystick either up or down.

Response Grouping[2] between T1 and T2 was

---

[2] Response qrouping is the tendency for participarts to wait for the reponding of T2 before initiating their response to T1.

discouraged by presenting a yellow warning point in the location of the fixation point if T2 response (RT2 + SOA) was made in less than 125% of the response to T1, measured from T1 onset. This operational definition of response grouping was adapted from Van Selst and Jolicoeur (1994). One practice block (containing 40 trials with mixed conditions) was given to participants before the formal test. All the aspects of the practice block were exactly the same as the subsequent experimental blocks. To reduce fatigue, self-controlled breaks between two continuous blocks were provided. The entire MEG recording took about 50 min.

2.5 Behavioural analysis

For RT analysis, only trials with correct T1 and T2 responses and correct response order were entered into the analysis. An outlier screen procedure (Van Selst & Jolicoeur, 1994) was used to exclude outliers of RTs in each cell for each participant. By using this approach, less than 4% of trials were labelled as outliers in the RT analysis. Post hoc analyses in each analysis of variance (ANOVA) of each experiment were conducted with the Bonferroni correction for multiple comparisons.

2.6 MEG acquisition and analysis

Whole-head MEG recordings were performed using a 275-channel axial gradiometer system (VSM MedTech Ltd, Port Coquitlam, Canada) with a 600 Hz sample rate and a 0-150Hz band-pass filter. Three out of the 275 channels were turned off due to excessive sensor noise. An additional 29 reference channels were used during the recording for the purpose of noise cancellation, and the primary sensors were analysed as synthetic third-order gradiometers (Vrba & Robinson, 2001). Three fiduciary markers (coils) were attached at fixed distances from anatomical landmarks (tragus, eye centre, which are identifiable in participants' anatomical MRIs) before the MEG data acquisition in order to achieve MRI/MEG co-registration in later source localization. Their location was

recorded with digital photographs before and after MEG acquisition. MEG data were first visually inspected, and trials with eye movement artefacts or sensor drifts were removed from the analysis.

2.7 ERF data preparation

To generate event-related field (ERF) time series, raw MEG data were first filtered using a 40 Hz low-pass filter. The averaging epoch was then defined from 200 ms before T1 onset (300ms before T2 onset) to 1200ms after T1 onset (1100ms after T2 onset), and data were baseline corrected based on a 200-ms pre-T1 onset interval. Since the CTF MEG system has 275 first-order axial gradiometer sensors that measure the gradient of the magnetic field in the radial direction, that is, orthogonal to the scalp, we also performed an axial to planar transformation using the Fieldtrip toolbox (neuroimaging.ruhosting.nl/fieldtrip/), which rearranged MEG data to a synthetic planar gradient configuration. One advantage of the axial to planar transformation is that the signal amplitude typically is then largest directly above a source.

In this experiment, half the correct T1 trials were on the left side of the display and half on the right side (Note that we did not have enough trials to further split the data into T1-left and T1-right conditions for further analysis). In order to avoid biasing the ERF result by an uneven number of valid T1 trials, we performed a counterbalancing procedure in the analysis for left-/right-T1 responses. For the ERF waveform analysis, from each condition of each individual participant the same number of T1-left and T1-right trials (depending on which condition had the minimum valid trial number) were randomly selected and fed into the ERF analysis, resulting in equal numbers of T1-left and T1-right trials. The total number of trials in the Chained dual condition was twice that of either the Stick or Switch dual conditions. In order to avoid a potential bias due to the uneven number of valid trials (particularly in the contrast of Chained vs. Independent-

stick), the valid trials of the Chained dual condition were randomly sampled (according to the minimum valid trial number of Stick and Switch conditions) to equalize trial numbers of Chained, Stick and Switch conditions for each individual participant. Similar random sampling procedures were performed for the other two contrasts (Stick vs. Switch; Chained single vs. Independent single) to equalize trial numbers for contrasts. Only trials with correct responses were used in ERF and source-localisation analyses. The minimum number of epochs used to produce any ERF waveform for any individual participant in any condition was 24. Otherwise this individual was excluded from analysis. On average, 32.1 epochs (out of 40) were used to produce an individual ERF for Chained dual, Independent-stick dual and Independent-switch dual conditions (SD=4.14), and 69.05 epochs (out of 80) were used to produce an individual single-task ERF for both Chained and Independent blocks (SD=6.65).

2.8 Multisensor ERF analysis

In order to avoid bias in selecting any particular sensor or sensor groups in the ERF analysis, we first performed a whole-head, multisensor ERF analysis with all the sensors included. To achieve this, a cluster-based permutation test was used to perform statistical group contrasts between three pairs of whole, 272-sensor, ERFs, (i.e., Chained dual ERF vs. Independent-stick dual ERF, Chained Single ERF vs. Independent Single ERF and Independent-stick ERF vs. Independent-switch ERF). By comparing activities induced by single-task trials within Chained blocks and that within Independent blocks, we can investigate whether other factors, such as different expectations and different T1 processing levels between Chained and Independent blocks contaminated our comparisons of interest (Chained dual task vs. Independent dual task). 1000 permutations were calculated for each statistical group analysis by using the Fieldtrip toolbox, and each resulting paired-samples Tstatistic was compared

with a primary threshold (0.05) to form clusters on the basis of temporal adjacency. A minimum neighbour sensor of 2 was used as an independent criterion in the formation of clusters to prevent all selected samples being included into a cluster if they had less than two neighbour sensors that also were selected. Cluster mass, that is, the sum of Tstatistics within each cluster, was used as the cluster-level statistic to assess the significance of suprathreshold clusters. The Monte Carlo method was used to calculate the distribution of the maximum of cluster mass in order to estimate the significance probability for each cluster. Only clusters with a significance probability (i.e., p value) less than a false alarm rate of .05 (two-tailed) were accepted to correct multiple comparisons (Maris & Oostenveld, 2007).

2.9 SAMerf analysis

For source localization, we used an event-related beamformer (Cheyne, Bakhtazad, & Gaetz, 2006), specifically SAMerf (Robinson, 2004) to explore the temporal sequence of cortical source activity underlying serial chaining. To perform the SAMerf analysis, unaveraged data (filtered 0 to 100Hz) from the whole of all trials (−200ms to 1200ms) were used to establish a global covariance matrix and set of SAM beamformer weights for all conditions (Robinson & Vrba, 1999). For headmodelling, a multiple, local-spheres-forward head model was derived by fitting spheres to each individual's brain surface extracted by BET(Brain Extraction Tool) (Huang, Mosher, & Leahy, 1999). Each individual had a structural image MRI scan (3D FSPGR [fast spoiled gradient echo] , 1 mm isotropic) acquired on a 3-T General Electric HDx scanner. From these beamformer weights, ERFs for individual conditions were passed through the beamformer weights so that SAMerf images of source power could be computed. Source power is defined as mean squared amplitude in a particular response time window (Robinson, 2004). Since we were interested in the stimulus-locked sustained activity (based on the multi-

sensor ERF results, see below), we used a time window covering from 450ms to 850ms after T1 onset to create SAMerf images, referenced to a baseline period of −200 to 0ms.

For group analysis, SAMerf images were spatially normalized onto the MNI (T1) average brain using FMRIB's Linear Affine Registration Tool (FLIRT; Jenkinson & Smith, 2001). This was achieved by first obtaining a set of warping parameters by registering each participant's anatomical MRI with the MNI template brain and thereafter applying these parameters to that individual's SAMerf maps. Nonparametric permutation tests (Singh, Barnes, & Hillebrand, 2003) (corresponding to parametric paired-sample t tests) were conducted in FSL (www.fmrib. ox.ac.uk/fsl/) using Threshold-Free Cluster Enhancement (TFCE) to perform cluster-level thresholding (Smith & Nichols, 2009). A 5mm Gaussian kernel was used to perform variance smoothing. The visualization program mri3dX was used to identify activation foci, to automatically label areas based on the Talairach Daemon database (Lancaster, et al., 2000), and to render group results onto the MNI averaged brain. Three groups of contrasts (i.e., Independent dual vs. Chained dual, Independent single vs. Chained single and Independent-stick vs. Independent-switch) were performed.

3 Results

3.1 Behavioural analysis

Effect of chaining operations on RT & accuracy. Since the comparison between the Chained and Stick conditions is the best way to evaluate the effect of chaining (i.e., neither condition requires a switch of spatial location), we carried out six paired-sample ttests (Chained vs. Independent-stick) for RT1, RT2, T1 Accuracy, T2 Accuracy, standard deviation (*SD*) of RT1, and SD of RT2. RT1 for the Chained
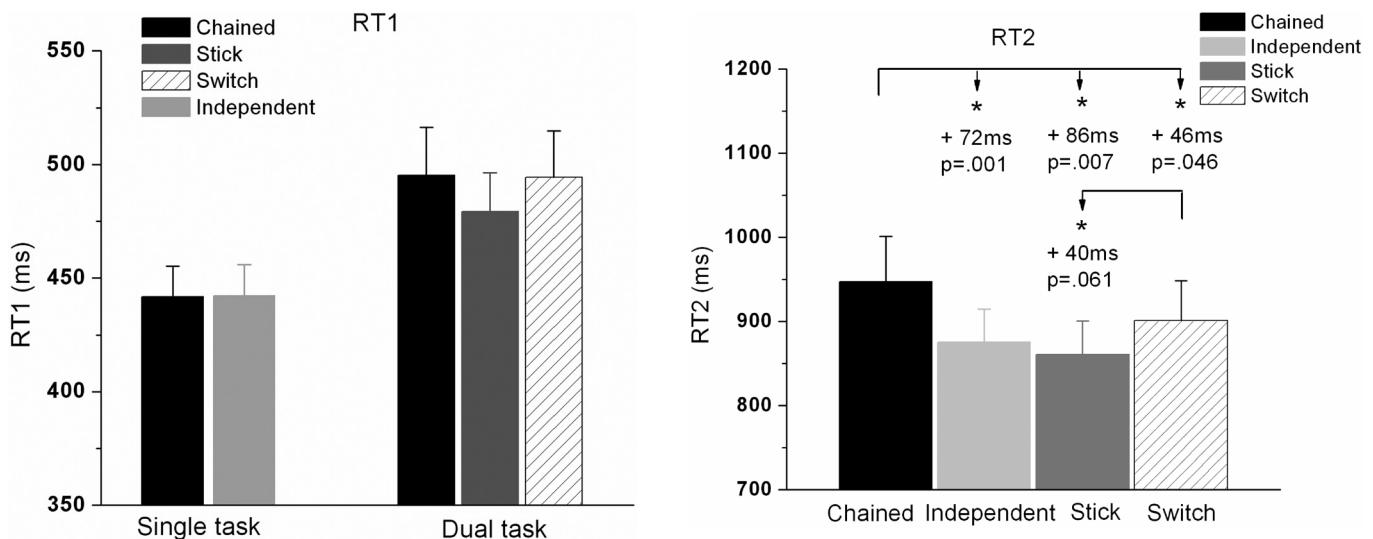
Figure. 2. Grand means of RT1 (left panel) / RT2 (right panel) of Chained and Independent conditions. On the X axis (left panel), the first column group shows RT1 from single task trials (no T2); the second column group shows RT1 from the Dual−task condition. The right panel shows RT2 from the four conditions. Vertical bars represent one standard error.

condition was slower than RT1 for the Independent−stick condition (Difference=16ms, SE=8.5ms, t=1.863, df=20, p=0.038, one−tailed). RT2 for the Chained condition was significantly slower than RT2 for the Independent−stick condition (Difference=86ms, SE=29ms, t=2.988, df=20, p=0.007) (see Figure2). T1 Accuracy for Independent−stick was not significantly different from T1 Accuracy for the Chained condition (p=0.477). However, T2 Accuracy for the Chained condition was significantly worse than T2 Accuracy for the Independent−stick condition (Difference =−6%, SE = 1.8%, t=−3.349, df= 20, p=0.003). SD of RT1 for the Chained condition was significantly larger than SD for RT1 of the Independent−stick condition (Difference= 20ms, SE= 6ms, t=3.253, p=0.005). The SD of RT2 for the Chained was also significantly larger than the SD of RT2 for the Independent−stick condition (Difference=88ms, SE=41ms, t=2.988, p=0.045). Given that RT and accuracy effects are in the same direction we can conclude there was no speed−accuracy trade−off.

Effect of chaining on task correlations. We first computed the within−trial Pearson's correlation between RT1 and RT2 for each participant and then submitted all individual mean correlation coefficients to paired samples

ttests, which revealed that the correlation between RT1 and RT2 was smaller (Difference=−.08, SE=0.039, t=−1.934, df=20, p=0.033, one−tailed) in the Chained condition (Mean=0.36, SD=0.20) relative to that in the Independent− stick condition (Mean=0.44, SD=0.23). The correlation difference between the Independent−stick (Mean=0.44, SD=0.23) and Independent−switch conditions (Mean=0.37, SD=0.25) did not reach significance (p=0.13), neither did the correlation difference between the Chained and the Independent−switch conditions (p= .899).

Effect of switching attention. We also explored the effect of switching attention by comparing the Independent−stick to the Independent−switch condition. Both mean RT1 (p=0.035) and RT2 (p=0.061, marginally) were significantly slower in the Independent−switch than their counterparts in the Independent−stick condition. However, unlike the effect of serial chaining, the SD of RT1 (p=0.130) and RT2 (p=0.527) in the Independent− stick condition were not significantly different from those in the Independent−switch condition.

Effect of T2 on T1: dual versus single tasks. There were only two types of single task trials (e.g. in Chained blocks and in Independent blocks) with no
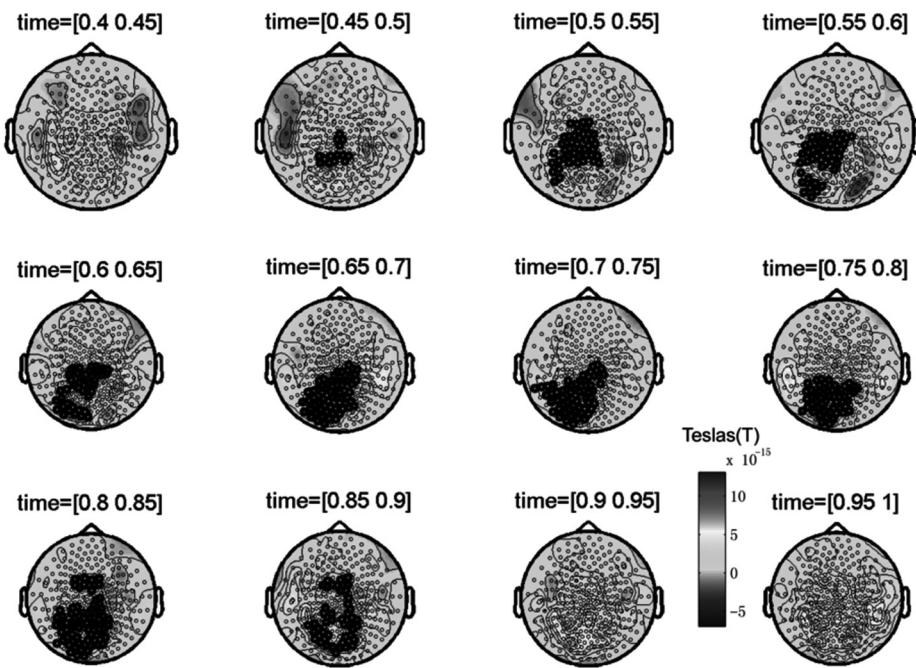
Figure 3. Results of the multi–sensor ERF analysis. Topography of the ERF difference between Chained (dual) and Independent–stick (dual) is overlaid with the topography of the significant clusters (p<.05, two–tailed, corrected by a nonparametric permutation test). Sensors that belong to significant positive clusters, i.e., higher amplitude in Chained (dual) ERF relative to Stick (dual) ERF, are indicated by black circles. Each subplot represents the topography during a time window of 50ms. Data were submitted to ERF analysis after axial to planar transformation.

significant difference between them on both RT (p=0.928) and percentage correct (p=0.926). Additional analyses were performed to explore how T2 affected T1. RT1 in the single task was significantly faster than RT1 in the dual task, F (1, 20) =24.408, p= .000; Difference =−49ms, SE=9.8ms, p=0.000. T1 Accuracy in the single task was significantly greater than T1 Accuracy in the dual task, F(1, 20)=8.626, p=0.008; Difference=3%, SE=0.8%, p= .008. Both RT and Accuracy data suggest that T2 produced an extra cost to the central processing stage of T1 when the two tasks were temporally adjacent (100ms SOA).

3.2 Multisensor ERF analysis

Three pairs of whole–head multisensor ERF contrasts were performed with a cluster–based permutation test at the grouplevel to correct for multiple comparisons along a time window from 200 ms before T1 onset (300ms before T2 onset) to 1200ms after T1 onset (1100ms after T2 onset). No cluster was significant before the onset of T1 for each of the three contrasts, which suggests

baseline levels were the same for all conditions. For the contrast of Chained dual vs. Independent–stick dual, only positive (higher amplitude in Chained condition) but not negative (higher amplitude in Stick condition) clusters were significant. For the other two contrasts (i.e., Chained vs. Independent Single and Independent–stick ERF vs. Independent–switch ERF), neither positive nor negative clusters were significant.

Figure 3 shows the group results of multisensor ERF analysis for the contrast of (dual) Chained versus Independent Stick over time. The amplitude difference between these two conditions reaches maximum in the posterior part of the brain and reveals a sustained effect. This sustained posterior activity distinguishing serial chaining from traditional PRP (independent) dual–task processing emerges from central posterior sensors approximately 450ms after T1 onset and spreads gradually to left parietal/temporal/occipital sensors. The sustained activity lasts for several hundred milliseconds and is not

fully attenuated until 900ms after T1 onset (800ms after T2 onset), near the end of T2 execution (average RT2 was around 900ms). There are no other significant clusters outside the time window of this activity.

3.3 Sensor–Group ERF analysis

The previously described multi–sensor ERF analysis reveals this sustained activity lasting from 450 – 900 ms after T1 onset, which appears to be linked with serial chaining. Inspection of this analysis suggests that, whereas 116 out of 272 sensors were statistically relevant in the contrast between (dual) Chained versus Independent–stick conditions (Figure 4), not all of the 116 sensors are unambiguously linked to the sustained activity we observed. For example, some sensors above motor areas likely represent the effect of the RT difference between these conditions. For this reason, we used three criteria to restrict our subsequent analyses to sensors that were directly linked with the sustained activity. The three criteria included were: First, based on the result of the multi–sensor ERF analysis, the time window used in this sensor selection was 450–900 ms after T1 onset. Second, only sensors in which the sustained effect lasted continuously longer than half of the maximum effect were selected, resulting in a cut–off value of 185ms since the maximum sustained effect of the input sensors was 370ms at sensor MLP31. Third, any individually selected sensor in the final sensor group had to have at least two neighbouring sensors that were also included. Only 19 of the 116 sensors meeting these criteria were used in the analysis.

By applying such criteria, 19 sensors (see Figure 4) met the above criteria and were used for the sensor–group analysis. These 19 sensors were derived from left and central parietal, left occipital, and left temporal clusters. A standard classification approach to the CTF system was used to define 14 sensor clusters, including central frontal, central central, central parietal, central occipital,

left frontal, left central, left parietal, left occipital, left temporal, right frontal, right central, right parietal, right occipital and right temporal.

For each individual participant, 5 ERF curves were calculated based on the averaged ERFs of all sensors in the selected sensor group (N=19). These five synthetic ERFs were calculated for five conditions (i.e., Chained single, Independent single, Chained dual, Independent–stick dual and Independent–switch dual). Three contrasts of paired ERFs(Chained single vs. Independent single, Chained dual vs. Independent–stick dual and Independent–stick dual vs. Independent–switch dual) were also performed using a paired–samples ttest at each sample point of each paired ERFs (i.e., moving ttests)

The results of the sensor–group ERF analysis revealed clear sustained ERF activity from left occipital, left temporal and left parietal sensors, starting from 450ms and lasting until 900ms after T1 onset, congruent with the multi–sensor ERF analysis (see Figure 5). Importantly, this difference is manifest only in the contrast between the Chained and Independent–stick conditions, suggesting this activity is directly linked with serial chaining and not modulated by other factors such as switching the locus of spatial attention. Moving t–tests in the contrast Stick vs. Switch suggest that the field amplitude for the Independent–switch condition was significantly higher in a time window from 900–960 ms after T1 onset, likely reflecting the switching of attention required in the latter condition.

3.4 SAMerf analysis

Similar to the ERF analysis in sensor space, we performed the same three contrasts (Chained dual vs. Stick dual; Stick dual vs. Switch dual, and Chained single vs. Independent single) in source space to localize event–related and phased–locked source activity. The results of the SAMerf approach reveal that only one out of the three contrasts (Chained dual vs. Stick dual, showed a significant
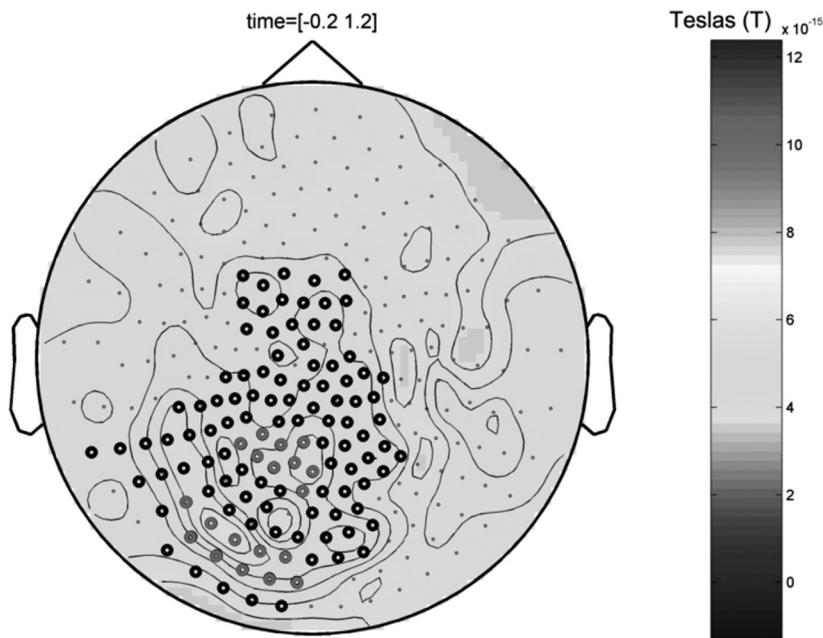
Figure. 4. The selected 19 sensors (labelled in red), out of a cluster of 116 sensors (labelled with black), that were used in the sensor–group ERF analysis. The 116 sensors (out of total 272 sensors) were those belonging to a significant cluster based on multi–sensor ERF analysis of Chained dual vs. Stick dual. Background topography was the ERF of Chained (dual) minus the ERF of Independent–stick (dual). Data were submitted to ERF analysis after axial to planar transformation.

effect, consistent with the ERF analysis. Figure 6 shows the group results of this analysis for the time window of 450ms to 850ms from T1 onset. The results suggest that source power for the Chained condition was significantly larger than that for the Independent–stick condition between 450ms to 850ms in the left superior temporal gyrus (Talairach coordinates at peak value = –45.2 –63.2 19.0). This result was consistent with the ERF analysis, which demonstrated sustained higher ERF amplitude for posterior sensors in the same time window. Importantly, this analysis points to a left lateralized source consistent with the finding in the sensor–group ERF analysis, that is, 18 out of 19 sensors in the selected sensor group were located on the left side.

Based on the source localization group results, we reconstructed the time course of source power activity at the left superior temporal gyrus (Talairach coordinates –45.2 –63.2 19.0), for each individual participant for five conditions. The T1–onset locked timecourse of source power activity was calculated based on the average of the

source power profile for each condition. The top panel of Figure 7 shows the T1–locked activity (normalized onto the baseline period from 200 ms pre–T1 to T1 onset) for each of the different conditions for the localized source shown in Figure 6 (i.e., a virtual sensor). Global covariance matrices and beamformer weights were used to produce the virtual sensor. The moving t–test profiles of the three contrasts of interest are presented in the lower panel. The data suggest that the left superior temporal source indicates a sustained higher level of activity in the contrast Chained dual vs. Independent–stick dual, but not in the other two contrasts (Stick dual vs. Switch dual and Chained single vs. Independent single). As in the ERF analysis, an equal number of T1–left and T1–right epochs were entered into the SAMerf analysis for each condition and for each participant in order to avoid a potential bias from the lateralized stimulus distribution.

3.5 Discussion

The behavioural data from the present experiment reveal the dramatic performance cost of serial chaining
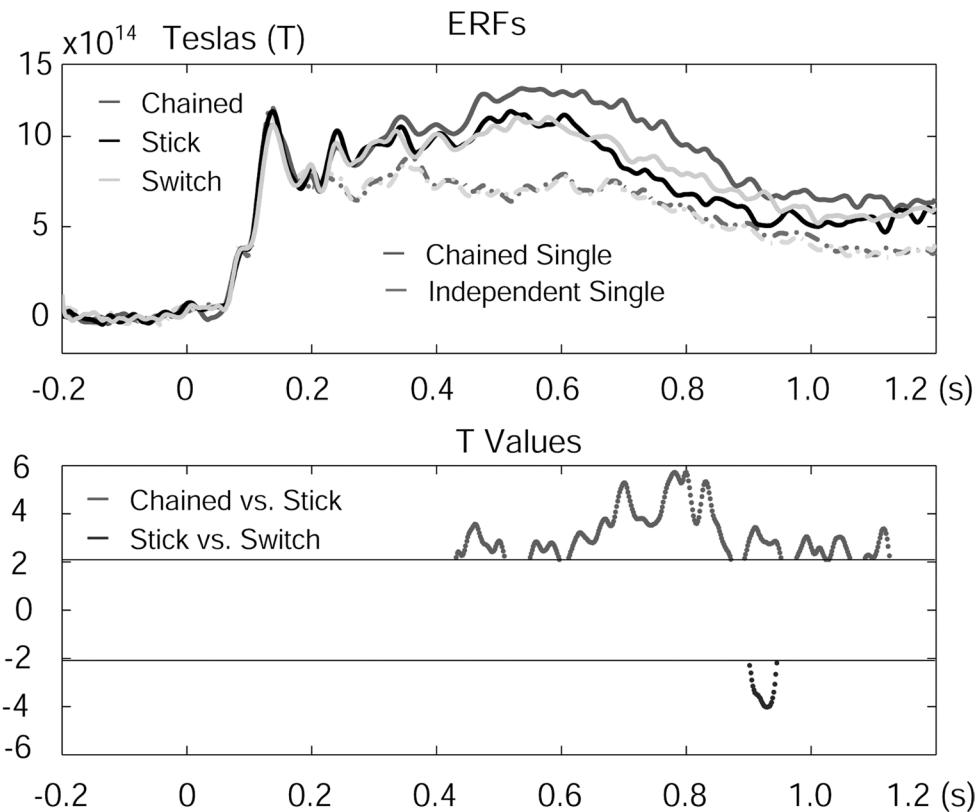
Figure. 5. Sensor–group ERF analysis for 19 selected sensors. The upper panel shows five synthetic ERFs for the Chained single condition (magenta dashed), Independent single condition (blue dashed), Chained dual (red solid), Independent–stick dual (black solid) and Independent–switch dual (green solid). The bottom panel shows the results of a moving t–test between the Chained dual vs. Independent–stick dual (red) and Independent–stick dual vs. Independent–switch dual (blue). There was no significant effect for the contrast of Chained single vs. Independent single. T–values between the two horizontal lines (t(20)= 2.086, e.g. two–tailed alpha value at 0.05), were masked and not shown. A 30ms time window was used to threshold the moving t–test values along the temporal dimension. Waveforms shown are the result of axial to planar transformation, i.e. synthetic planar gradient data.

in two important aspects: First, a slowed RT and corresponding increase in variability; and second, a decreased temporal coupling between the two sequential tasks. Importantly, this behavioural cost is not revealed when the chaining operation is not required, that is in the independent task with and without the requirement to switch the location of attention, even though location switching has an effect on RT. Our behavioural results are consistent with a recent report (Sackur & Dehaene, 2009), revealing that serial chaining consumes central processing resources and replicates our previous behavioural results (Fan, et al., 2011).

Importantly, the present report reveals the neural correlates of the large behavioural cost associated with chaining cognitive operations. Sensor and source space MEG analyses converge to reveal sustained posterior activity occurring from 450ms to 900ms after T1 onset (350ms to 800ms after T2 onset) that is linked to the serial chaining requirement. This activity appears only in the contrast between Chained vs. Independent–stick; not in the other two contrasts (Chained single vs. Independent single; Independent–switch vs. Independent–stick), mirroring the effects found in both the behavioural and subsequent SAMerf (source space) analyses. Localisation in source space using an event–related beamformer approach suggests the source of this activity is distributed along the boundary of left temporal, left parietal and left occipital brain areas. The event–related beamformer (Robinson,
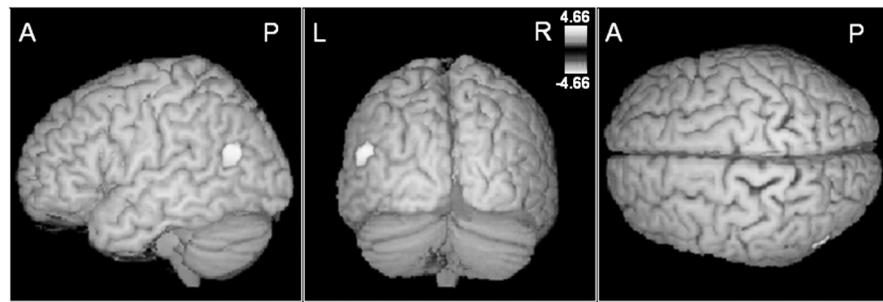
Figure. 6. Group–level SAMerf analysis showing the results of the contrast between Chained dual vs. Independent–stick dual conditions for ERFs during the time window 450ms to 850ms after T1 onset. A TFCE approach of thresholding was used to correct for multiple comparisons. Labelled voxels are significant at the p<0.05 level (two–tailed) after correction. Units are pseudo–*t* values. Yellow/orange areas indicate relatively larger source power in the Chained dual condition.
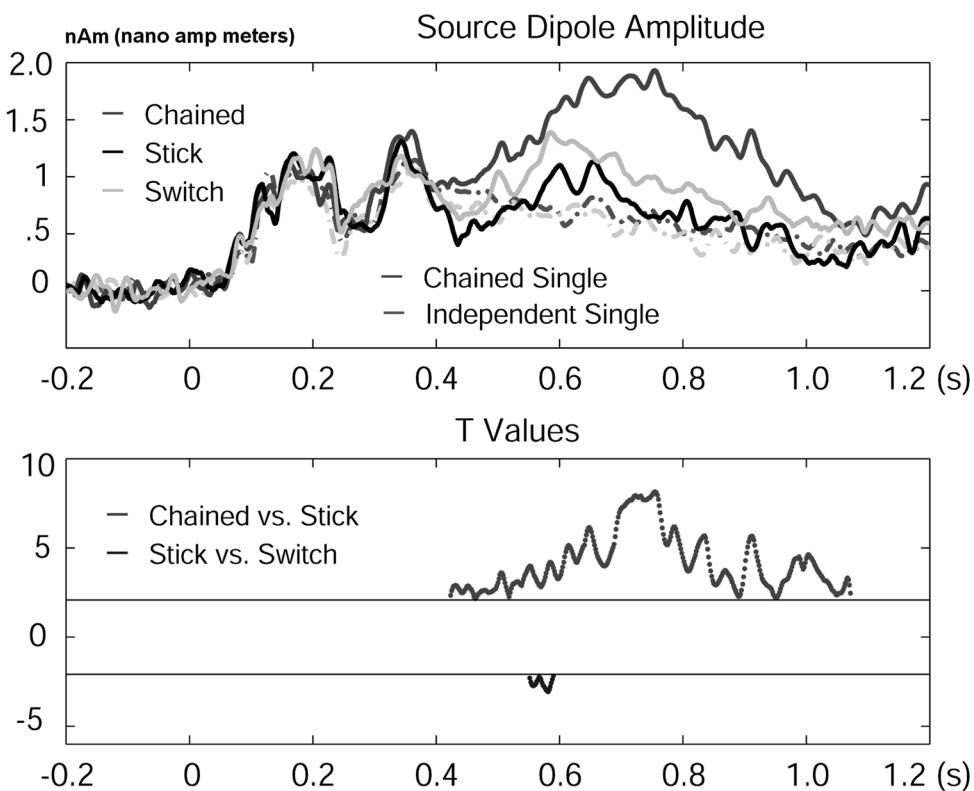
Figure. 7. Group–level time series of source dipole amplitude for different conditions and their contrasts. The upper panel shows the time series of activation for the Chained single condition (magenta dashed), Independent single condition (blue dashed), Chained dual (red solid), Independent–stick dual (black solid) and Independent–switch dual (green solid). Bottom panel: *t* values of a moving *t*– test analysis, i.e., Chained dual vs. Independent–stick dual (red) and Independent–stick dual vs. Independent–switch dual (blue). There was no significant effect for the contrast of Chained single vs. Independent single. A 30ms time window was used to perform extent thresholding on *t*–test values along the temporal dimension. *T*values between two horizontal lines ,*t*(20)=±2.086, e.g. two–tailed alpha value at 0.05, were masked and are not shown.

2004) is a powerful, relatively recent method for localising ERFs with very few assumptions required (Cheyne, et al., 2006). One shortcoming of the beamformer approach is that it is insensitive to sources that are highly correlated, although simulations have shown that beamformers can detect correlated sources depending on the degree of correlation and SNR (Gross, et al., 2001). It is therefore theoretically possible that the present source localisation analysis has missed some source activity which is both temporally correlated and spatially distributed that other approaches such as minimum norm–based solutions might detect. However, the consistency both spatially, temporally

and experimentally of the event-related beamformer results with the left lateralisation pattern observed in the sensor space analysis suggests this not a major issue with the present analyses.

It could be argued that the sustained activity observed in the Chained, relative to the Independent-stick, condition occurred as a result of the longer display duration for Task 2 stimuli in the former. The longer duration is a consequence of the fact that RT2 for the Chained condition is significantly slower than RT2 for the Independent-stick condition. The temporal characteristics of this activity revealed by the multichannel and SAMerf analyses, however, argue against this account. The larger magnitude in the Chained condition emerges as early as 350ms after T2 onset, well before the stimulus terminates in either condition. Thus, the increased activity in the Chained condition cannot be explained by a low-level, sensory-based account.

The characteristics of the sustained activity we observed, which distinguish chained from independent dual-task processing, are consistent with previous reports from human ERP studies (Brisson & Jolicoeur, 2007a, 2007b; Dell'Acqua, Sessa, Jolicoeur, & Robitaille, 2006; Jolicoeur, Sessa, Dell'Acqua, & Robitaille, 2006a, 2006b; Klaver, Talsa, Wijers, Heinze, & Mulder, 1999; McCollough, Machizawa, & Vogel, 2007; Vogel, Luck, & Shapiro, 1998; Vogel & Machizawa, 2004) and MEG (Robitaille, et al., 2009; Robitaille, et al., 2010) where a temporally and topographically similar component, the SPCN (see also Note 1) or SPCM, respectively, have strongly been linked with maintenance in VSTM (or 'working' memory). The SPCN typically begins at about 300ms after target onset and is sustained for at least several hundred milliseconds into the period during which the stimulus generating this component has to be retained (McCollough, et al., 2007).

Although the sustained activity witnessed in the Chained condition likely relates to the SPCM and its ERP counterpart (i.e., SPCN or CDA), we speculate it may represent a marker for information buffering over and above the reflection of VSTM activity normally ascribed to it. First, we note that the SPCM is typically believed to index increases in working memory (WM) capacity as it reflects the number of stimuli stored in posterior brain areas. As we are comparing the Chained to Independent conditions, where the number of stimuli is not being manipulated, the differences observed may reflect other cognitive machinery. Second, although the WM requirement for the Chained relative to the Independent conditions requires the content to be changed more frequently; even in the latter condition participants must actively maintain the contents of WM in order to correctly make a (T2) response. Yet, behavioural and electrophysiological indices suggest the former exacts an increased cost that we suggest reflects the chaining operation, per se. Third, Mitchell and Cusack (2011) note that the effects of memory load using MEG are predominantly bilateral (see also Grimault, et al., 2009). This suggests that the present result, which did not reveal bilateral activity, may not simply index VSTM load but is sensitive to the more complex operation required by chaining. Finally, we note that using MEG to assess VSTM usually reveals generators in the superior intraparietal sulcus (IPS; Mitchell & Cusack, 2011; Robitaille, et al., 2010). We, on the other hand, localised MEG activity on the boundary of the temporal, occipital, and parietal lobes, further supporting our contention we are not merely revealing differences due to increased memory load.

An interesting observation from the present study is the left lateralization of the serial chaining-related neural activity in both sensor and source spaces. Since these analyses contained exactly the same number of T1-left and T1-right trials, it is unlikely that this left lateralization effect is due to an uneven distribution of T1 stimuli or other artefact. Instead, the leftlateralization of serial chaining

is consistent with a well-known phenomenon in clinical neurology known as the left hemispheric lateralization for praxis planning. In the early 20th century, Hugo Liepmann observed that stroke patients with a lefthemispheric lesion are more likely to develop apraxia than those with a right-hemispheric lesion (reviewed in Goldenberg, 2003). Clinical neurological studies (Alexander, Baker, Naeser, Kaplan, & Palumbo, 1992; Geschwind, 1965) suggest that praxis planning, particularly the programming of a motor sequence, is largely dominated by the left hemisphere. In the past decade, this suggestion has been confirmed repeatedly using functional imaging (Bohlhalter, et al., 2009; Buccino, et al., 2004; Fridman, et al., 2006; Hermsdorfer, Terlinden, Muhlau, Goldenberg, & Wohlschlager, 2007; Johnson-Frey, Newman-Norlund, & Grafton, 2005). In particular, a recent fMRI study (Bohlhalter, et al., 2009) revealed a left hemispheric lateralization in posterior parietal cortex (PPC) and premotor cortex (PMC) association areas when participants were planning to perform a pantomime task (such as 'wave good-bye'), regardless of whether the planning occurred for the right or left hand.

We suggest our results provide evidence of a supervisory control system, over and above that revealed by typical dual-task (e.g., PRP) demands, by the existence of sustained activity in a left posterior brain area. This area, likely in combination with prefrontally based executive control centres, subserves planning and execution demands by organising hierarchies of cognitive subunits into complex goal-directed behaviour. This cognitive organisation, though essential to complex human behaviour, comes at a behaviourally measurable cost, as evidenced by slower response times and disrupted response coupling between relevant tasks. The present findings lend new understanding to the brain processes that underlie the moment-by-moment need for humans to chain sequential cognitive operations as they go about their daily behaviour.

References

Alexander, M. P., Baker, E., Naeser, M. A., Kaplan, E., & Palumbo, C. (1992). Neuropsychological and neuroanatomical dimensions of ideomotor apraxia. *Brain*, 115, 87–107. doi: 10.1093/brain/115.1.87

Bohlhalter, S., Hattori, N., Wheaton, L., Fridman, E., Shamim, E. A., Garraux, G., ... Hallett, M. (2009). Gesture subtype-dependent left lateralization of praxis planning: An event-related fMRI Study. *cerebral Cortex*, 19, 1256–1262. doi: 10.1093/cercor/bhn168

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433–436. doi: 10.1163/156856897X00357

Brisson, B., & Jolicoeur, P. (2007a). Cross-modal multitasking processing deficits prior to the central bottleneck revealed by event-related potentials. *Neuropsychologia*, *45*(13), 3038–3053. doi: 10.1016/j.neuropsychologia.2007.05.022

Brisson, B., & Jolicoeur, P. (2007b). A psychological refractory period in access to visual short term memory and the deployment of visual-spatial attention: Multitasking processing deficits revealed by event-related potentials. *Psychophysiology, 44*, 323–333. doi: 10.1111/j.1469-8986.2007.00503.x

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H. J., ... Rizzolatti, G. (2004). Neural circuits underlying imitation learning of hand actions: An event-related fMRI study. *Neuron*, *42*, 323–334. doi: 10.1016/S0896-6273(04)00181-3

Cheyne, D., Bakhtazad, L., & Gaetz, W. (2006). Spatiotemporal mapping of cortical activity accompanying voluntary movements using an event-related beamforming approach. *Human. Brain Mapp, 27*(213–219). doi: 10.1002/hbm.20178

Dell'Acqua, R., Sessa, P., Jolicoeur, P., & Robitaille, N.

(2006). Spatial attention freezes during the attentional blink. *Psychophysiology*, *43*, 394–400. doi: 10.1111/ j.1469–8986.2006.00411.x

Fan, Z., Singh, K. D., Muthukumaraswamy, S. D., Sigman, M., Dehaene, S., & Shapiro, K. (2011). The cost of serially chaining two cognitive operations. *Psychological Research*. doi: 10.1007/s00426–011– 0375–y

Fridman, E. A., Immisch, I., Hanakawa, T., Bohlhalter, S., Waldvogel, D., Kansaku, K., … Hallett, M. (2006). The role of the dorsal stream for gesture production. *NeuroImage*, *29*, 417–428. doi: 10.1016/ j.neuroimage.2005.07.026

Geschwind, N. (1965). Disconnexion syndromes in animals and man. *I. Brain, 88*, 237–294. doi: 10.1093/ brain/88.2.237, II doi: 10.1093/brain/88.3.585

Goldenberg, G. (2003). Apraxia and beyond: life and work of Hugo Liepmann. *Cortex*, *39*, 509–524. doi: 10.1016/ S0010–9452(08)70261–2

Grimault, S., Robitaille, N., Grova, C., Lina, J. M., Dubarry, A. S., & Jolicoeur, P. (2009). Oscillatory activity in parietal and dorsolateral prefrontal cortex during retention in visual short–term memory: Additive effects of spatial attention and memory load. *Human Brain Mapping, 30*, 3378–3392. doi: 10.1002/ hbm.20759

Gross, J., Kujala, J., Hamalainen, M., Timmermann, L., Schnitzler, A., & Salmelin, R. (2001). Dynamic imaging of coherent sources: studying neural interactions in the human brain. *Proceeding of the National Academy of Science*, *98*, 694–699. doi: 10.1073/pnas.98.2.694

Hermsdorfer, J., Terlinden, G., Muhlau, M., Goldenberg, G., & Wohlschlager, A. M. (2007). Neural representations of pantomimed and actual tool use: Evidence from an event–related fMRI study. [Supplement 2].*NeuroImage*, *36,* T109–T118. doi: 10.1016/j.neuroimage.2007.03.037

Huang, M. X., Mosher, J. C., & Leahy, R. M. (1999). A sensor–weighted overlapping–sphere head model and exhaustive head model comparison for MEG. *Physics in Medicine and Biology. 44,*, 423–440. doi: 10.1088/0031–9155/44/2/010

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis, 5*, 143–156. doi: 10.1016/ S1361–8415(01)00036–6

Johnson–Frey, S. H., Newman–Norlund, R., & Grafton, S. T. (2005). A distributed left hemisphere network active during planning of everyday tool use skills. *Cerebral Cortex*, *15*, 681–695. doi: 10.1093/cercor/bhh169

Jolicoeur, P., Sessa, P., Dell'Acqua, R., & Robitaille, N. (2006a). Attentional control and capture in the attentional blink paradigm: Evidence from human electrophysiology. *European Journal of Cognitive Psychology*, *18*, 560–578. doi: 10.1080/09541440500423210

Jolicoeur, P., Sessa, P., Dell'Acqua, R., & Robitaille, N. (2006b). On the control of visual spatial attention: Evidence from human electrophysiology. *Psychological Research, 70*, 414–424. doi: 10.1007/s00426–005– 0008–4

Klaver, P., Talsa, D., Wijers, A. A., Heinze, H. J., & Mulder, G. (1999). An eventrelated brain potential correlate of visual short–term memory. *Neuroreport*, *10*, 2001–2005. doi: 10.1097/00001756–199907130– 00002

Lancaster, J. L., Woldorff, M. G., Parsons, L. M., Liotti, M., Freitas, E. S., Rainey, L., … Fox, P. T. (2000). Automated Talairach Atlas labels for functional brain mapping. *Human Brain Mapping*, *10*, 120–131. doi: 10.1002/1097–0193(200007)10:3<120::AID– HBM30>3.0.CO;2–8

Lien, M.–C., Croswaite, K., & Ruthruff, E. (2011). Controlling spatial attention without central

attentional resources: Evidence from event−related potentials. *Visual Cognition, 19*, 37−78. doi: 10.1080/13506285.2010.491643

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG− and MEG−data. *Journal of Neuroscience Methods, 164*(1), 177−190. doi: 10.1016/j.jneumeth.2007.03.024

McCollough, A. W., Machizawa, M. G., & Vogel, E. K. (2007). Electrophysiological measures of maintaining representations in visual working memory. *Cortex, 43*, 77−94. doi: 10.1016/S0010−9452(08)70447−7

Mitchell, D. J., & Cusack, R. (2011). The temporal evolution of electromagnetic markers sensitive to the capacity limits of visual short−term memory. *Frontiers in Human Neuroscience*, 5, 1−20. doi: 10.3389/fnhum.2011.00018

Oberauer, K., & Kliegl, R. (2004). Simultaneous cognitive operations in working memory after dual−task practice. *Journal of Experimental Psychology: Human Perception & Performance*, 30, 689−707. doi: 10.1037/0096−1523.30.4.689

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception & Performance*, 10, 358−377. doi: 10.1037/0096−1523.10.3.358

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, 10, 437−442. doi: 10.1163/156856897X00366

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology: Human Perception & Performance*, 18, 849−860. doi: 10.1037/0096−1523.18.3.849

Robinson, S. E. (2004). Localization of event−related activity by SAM(erf). *Neurolagy & Clinical Neurophysiology, 109*, 109.

Robinson, S. E., & Vrba, J. (1999). Functional neuroimaging by synthetic aperture magnetometry (SAM). In T. Yoshimoto, M. Kotani, S. Kuriki, H. Karibe & N. Nakasato (Eds.), *Recent Advances in Biomagnetism* (pp. 302−305). Sendai: Tohoku University Press.

Robitaille, N., Grimault, S., & Jolicoeur, P. (2009). Bilateral parietal and contralateral responses during maintenance of unilaterally encoded objects in visual short−term memory: Evidence from magnetoencephalography. *Psychophysiology, 46*, 1090−1099. doi: 10.1111/j.1469−8986.2009.00837.x

Robitaille, N., Marois, R., Todd, J., Grimault, S., Cheyne, D., & Jolicoeur, P. (2010). Distinguishing between lateralized and nonlateralized brain activity associated with visual short−term memory: fMRI, MEG, and EEG evidence from the same observers. *NeuroImage, 53*, 1334−1345. doi: 10.1016/j.neuroimage.2010.07.027

Sackur, J., &Dehaene, S. (2009). The cognitive architecture for chaining of two mental operations. *Cognition, 111*, 187−211. doi: 10.1016/j.cognition.2009.01.010

Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: A characterization of the mind's bottleneck. *PLoS Biology*, 3, e37. doi: 10.1371/journal.pbio.0030037

Sigman, M., & Dehaene, S. (2008). Brain mechanisms of serial and parallel processing during dual−task performance. *Journal of Neuroscience*, 28, 7585−7598. doi: 10.1523/JNEUROSCI.0948−08.2008

Singh, K. D., Barnes, G. R., & Hillebrand, A. (2003). Group imaging of task−related changes in cortical synchronisation using nonparametric permutation testing. *NeuroImage*, 19, 1589−1601.

Smith, S., & Nichols, T. E. (2009). Threshold−free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, 44, 83−98. doi: 10.1016/

j.neuroimage.2008.03.061

Tombu, M. N., Asplund, C. L., Dux, P. E., Godwin, D., Martin, J. W., & Marois, R. (2011). A Unified attentional bottleneck in the human brain. *Proceeding of the National Academy of Science*, *108*, 13426–13431. doi: 10.1073/pnas.1103583108

Turing, A. M. (1936). On computable numbers, with an application to the Entscheidungs problem. *Proceedings of the London Mathematical Society* ,*42*. 230–265.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology*, *47*, 631–650. doi: 10.1080/14640749408401131

Vogel, E. K., Luck, S. J., & Shapiro, K. L. (1998). Electrophysiological evidence for a postperceptual locus of suppression during the attentional blink. *Journal of Experimental Psychology: Human Perception & Performance*, *24*, 1656–1674. doi: 10.1037/0096–1523.24.6.1656

Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751. doi: 10.1038/ nature02447

Vrba, J., & Robinson, S. E. (2001). Signal processing in magnetoencephalography. *Methods*, *25*, 249–271. doi: 10.1006/meth.2001.1238

Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short–term memory for objects. *Nature*, *440*(7080), 91–95. doi: 10.1038/ nature04262

The cost of serially chaining two cognitive operations

Zhao Fan[1,2], Krish Singh[3], Suresh Muthukumaraswamy[3], Mariano Sigman[4], Stanislas Dehaene[5], Kimron Shapiro[6]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior ,Central China Normal University, Ministry of Education, Wuhan 430079, China)

([2] School of Psychology, Central China Normal University, Wuhan 430079, China)

([3] School of Psychology, Cardiff University, CUBRIC, CF10 3AT Cardiff, UK)

([4] Physics Department, University of Buenos Aires, 1428 Buenos Aires, Argentina)

([5] Collège de France, 75005 Paris, France)

([6] School of Psychology, Bangor University, Bangor LL57 2AS , Wales,UK)

Abstract As Turing (1936, *Proceedings of the London Mathematical Society)* noted, a fundamental process in human cognition is to effect chained sequential operations in which the second operation requires an input from the preceding one. Although a great deal is known about the costs associated with 'independent' (unrelated) operations, e.g., from the classic psychological refractory period paradigm, far less is known about those operations to which Turing referred. We present the results of two behavioural experiments, where participants were required to perform two speeded sequential tasks that were either chained or independent. Both experiments reveal the reaction time cost of chaining, over and above classical dual–task serial costs. Moreover, the chaining operation significantly altered the distribution of reaction times relative to the Independent condition in terms of an increased mean and variance. These results are discussed in terms of the cognitive architecture underlying the serial chaining of cognitive operations.

Keywords dual task; psychological refractory period; chaining; sequential operations

1 Introduction

During the past few decades, cognitive psychologists have gained considerable insight into the organization behind a *single* cognitive operation including how an elementary decision is reached by the accumulation of sensory evidence. Yet to date, we still have surprisingly little knowledge of the cognitive mechanisms by which *multiple* elementary operations are sequentially assembled into mental 'routines', especially when a cognitive operation partially or completely depends on the input from a previous cognitive operation–a situation called 'chaining'. Such chaining operations are vital for human decision making

in everyday life; a behavioural understanding of which constitutes the motivation for the present report.

Temporal proximity between two cognitive operations, even if they are not chained, reveals a resource conflict. Indeed, several behavioral observations such as the phenomena of the psychological refractory period (PRP; cf. Pashler, 1984; Sigman & Dehaene, 2005, 2006) and the attentional blink (AB; Raymond, Shapiro, & Arnell, 1992) demonstrate our inability to process multiple independent stimuli when presented closely in time. Researchers have offered various accounts of the above-mentioned resource conflict. In the case of PRP, a structural central bottleneck (Pashler, 1984, 1994), a central capacity sharing mechanism (Tombu & Jolicoeur, 2003), and an adaptive executive control mechanism (Meyer & Kieras, 1997a,b), have each been proposed to account for the hallmark effect of PRP, i.e., RT to the second of two independent tasks (both require speeded responses) increases rapidly with the decrease of SOA between them. Turning to the AB, the interference model (Shapiro, Raymond, & Arnell, 1994), the two-stage model (Chun & Potter, 1995), and the goal-directed control model (Kawahara, 2003), have been proposed to explain the AB phenomenon, e.g. the impaired ability to identify the second of two independent targets (speeded responses are *not* required) presented in close succession. At the heart of all models for both the PRP and AB phenomena lies an information processing 'bottleneck' in which the processing of one task temporarily prevents the execution of crucial steps of a temporally overlapping task.

In contrast to numerous studies investigating of the cost associated with performing two independent tasks, few experiments have been conducted to examine the cost associated with tasks where performance of the second depends on the outcome of the first. A recent report by Sackur and Dehaene (2009) describes a detailed chronometric analysis of such a 'chaining' task, which required participants to add or subtract an input number to a fixed reference digit (arithmetic operation), then compare the intermediate result with another fixed reference digit (comparison operation) to determine which was greater. The two operations of this chaining task share a single external stimulus, e.g. the input number. Using this significant departure from traditional PRP approaches, these investigators revealed a partially serial mechanism with cross-talk between the two successive tasks, i.e., the comparison operation was able to be launched prior to completion of the arithmetic operation. Importantly, however, those authors did not compare chained versus non-chained tasks and thus could not isolate the effect of chaining from the effect due to temporal proximity. This vital gap is what the present study seeks to fill (see "General Discussion" for a more complete treatment of this issue).

To accomplish our goal a novel variation of the PRP paradigm was derived, which required two successive tasks to be performed: In one, Independent condition, successful performance of the second task was unrelated to the first, as is the case in a typical PRP experiment. In the second, Chained condition, successful performance of the second task was dependent on successful performance of the first. Our goal was to characterise the performance difference between these two conditions, i.e., to isolate a chaining effect over and above the traditional PRP effect.

Previous evidence from PRP literature suggests that the information processing bottleneck to which we refer above arises from a central decision stage, which consumes cognitive resources and suffers from seriality. For example, Pashler (1984, 1994) argued that, whereas the perceptual (P) and response (M) stages of information processing can operate independently on each task, the bottleneck occurs when the central (C) stage must connect P to M stages. A recent model proposed by Sigman & Dehaene (2005, 2006) provides a promising *computational implementation* of the so-called 'central bottleneck' account. In their recent studies, the authors suggest that the central bottleneck

occurs as a result of stochastic evidence accumulation, required prior to executing a decision. In this model, the decision−making stage is viewed as a noisy integrator that accumulates perceptual evidence from the sensory system via a 'random walk'. Sigman and Dehaene's model relates the 'psychological' bottleneck (Pashler, 1984) to accumulation−based decision mechanisms (Gold & Shadlen, 2001, 2002; Ratcliff, 1988; Usher & McClelland, 2001). Specifically, Sigman and Dehaene (2005, 2006) suggest that when two sequential operations are performed as part of a larger "routine", their central stages conflict and are forced to execute serially. When this occurs, response speed is determined by a tight succession of multiple stochastic accumulation stages (one for each task).

What prediction does the stochastic evidence accumulation model make for the present experiment? As with Pashler's (1984, 1994) central bottleneck model, this new model must account for the additional resources required to handle the increased complexity of the cognitive operations required to transfer information from the first to the second task in addition to dealing with the dual−task requirements. The stochastic accumulation model suggests the Chained condition will require additional resources during the 'central decision stage'; a stage already characterised above as 'noisy'. Accordingly, the extra demand arising from the chaining requirement will introduce further variability, in turn yielding slower RTs and, importantly, greater RT *variability*. Thus, according to the prediction of the 'stochastic evidence accumulation' model, we would expect to observe an altered distribution of reaction times, in terms of an increased mean and variance / standard deviation, of the chained condition relative to the independent condition (See "Discussion" for a full treatment of this issue).

2　Experiment 1

A novel PRP paradigm composed of two spatial arrow−
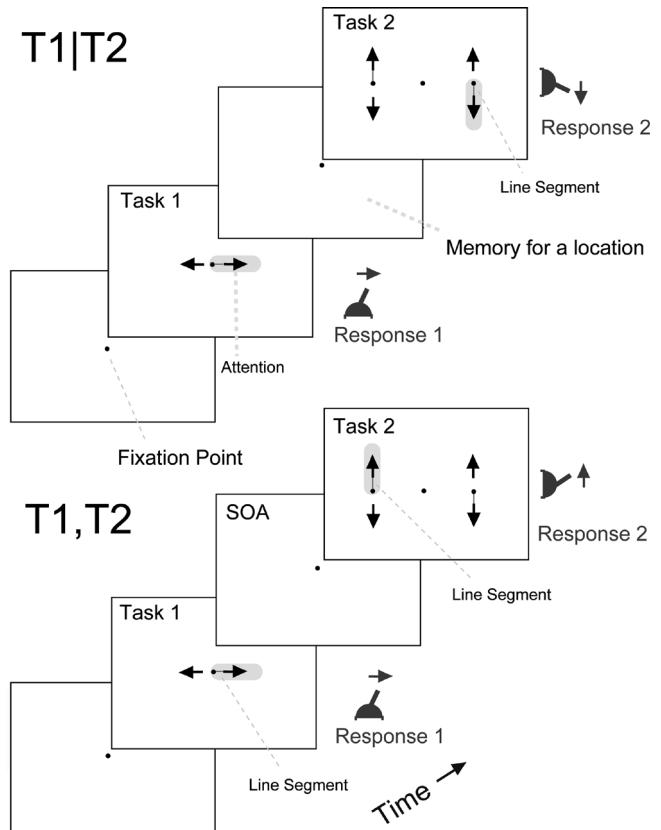
Fig. 1 Schematic representation of the tasks used to study serial processing. *Upper*, in Task 1 of the Chained task (T1|T2) the participant sees two *horizontal arrows*; one of these arrows is connected by a *dim line segment* to the fixation point. The participant is required to push a joystick as soon as possible with left hand to the corresponding side of the line segment and remembers this location. For Task 2, *two vertical arrows* and one line segment appear on both sides simultaneously and the participant indicates the direction of the arrow connected to the T2 line segment of the remembered side by pushing a joystick with his/her right hand as quickly as possible. The grey region indicates the anticipated distribution of attention as well as the location that has to be remembered but was not present in the display. *Lower* in the independent task (T1,T2), Task 1 is identical. Task 2 is now independent from Task 1 because the participant reports about either the left or right side T2 line segment for a block of trials.

tracking tasks was used in this experiment (see Fig.1). Participants were instructed to respond as quickly as possible to each task using two joysticks, one in each hand. There were two conditions in this experiment. In the Chained condition, participants were instructed to remember the result of Task 1 and use it to perform Task 2. In the Independent condition, the tasks were identical but had no dependency. Four SOAs enabled us to assess the effect of this variable.

2.1 Method

2.1.1 Participants

A total of 26 participants, 22 females and 4 males, aged from 18 to 35 (mean=21; STD=4), took part in this experiment. Half of the participants were randomly assigned to the Chained group, the other half to the Independent group. All participants were University students or employees and were paid by either course credits or cash for their participation. All participants had normal or corrected to normal vision with no history of visual or neurological disorders. This experiment was approved by the Bangor University Research Ethics Committee, and participants gave their informed consent before the experiment.

2.1.2 Apparatus

The participants were seated in a room which was dark except for the display. All stimuli were produced with MATLAB (Mathworks Inc.) and Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) and were displayed on a Dell P1130 colour monitor, driven by a NVIDIA FX 5200 Graphics adapter in a Pentium 4 host computer. The refresh rate of the monitor was 100Hz with a display area 41^0 in the horizontal axis by 30.5^0 in the vertical axis. Participants' responses were recorded by two Competition Pro 5000 Joysticks connected to the PC via USB cables. The viewing distance between the centre of the screen and the mid point of the participant's eyes was 57cm.

2.1.3 Stimuli

Two tasks were presented sequentially on each trial (see Fig.1) In Task 1, two horizontal arrows and a single line segment connecting one of the two arrows to the fixation point were presented. Each arrow contained an equal−side triangle head (0.91^0 length for the side and 0.82^0 length for the base) and a horizontal arrow line (1.1^0 length). The luminance of the background and that of the arrows was 80 and 1.2 cd/m^2, respectively (measured with a Minolta CS−100 Chromameter photometer). The vertex of each arrow head was 2.91^0 away from the central fixation point horizontally. For a particular trial, Target 1 was to be judged "left" or "right" by virtue of on which side of the display the arrow line was connected to the central fixation point by a horizontal line segment (1^0 length). Participants were instructed to push the left−hand joystick horizontally (left or right) according to the location of the target as quickly as possible after the target appeared. The stimuli in the second task (Task 2) were composed of four vertically oriented arrows and two vertical line segments/targets. The size of the arrows and line segments were the same as those in Task 1, however the orientation and presented locations were different. The vertex of each arrow head was 2.91^0 away from the central fixation point both horizontally and vertically. The locations of the two targets were either within top−left and bottom−right quadrants or top−right and bottom−left quadrants of the visual field. By this arrangement one target set was always on the left and the other on the right side of the display. For Task 2, participants were instructed to use the right−hand joystick and push it vertically up or down, respectively, depending on whether the upper or lower arrow was connected to the centre point. Unlike Task 1, Task 2 contained a target on both the right and left, so participants required additional information to perform Task 2 correctly. In the Chained group (T1|T2; upper display in Figure 1), this extra pre−cuing information was the input from the result of Task 1. For example, if the single target in Task 1 was located on the left side of the display then the left side of the display in Task 2 had to be evaluated for a response, and vice versa. In the Independent group, (T1,T2; lower display in Figure 1), this pre−cuing information (left or right target display side), was provided at the beginning of each (Independent) block.

2.1.4 Design

A two−factor mixed design was used in this experiment. The first factor (Chained vs. Independent) was a between−

participant factor with thirteen participants in each experimental group. The second factor (SOA, Stimulus Onset Synchrony) was a within−participant variable with four levels (100, 200, 350 and 650 ms) between Task 1 and 2. The location(s) of the target(s) in both tasks was counterbalanced across trials. Every combination of SOA was repeated sixty−four times, resulting in a total of 256 trials in each experimental session.

The experimental session was divided into 8 blocks with 32 trials in each block. For the Chained group, all eight blocks contained Chained task (T1|T2) trials. For the Independent group, four blocks contained Independent task trials (T1,T2) in which the *left* target of Task 2 was pre−defined by presenting a verbal instruction in the display at the beginning of each block. For the other four blocks, the *right* target of Task 2 was pre−defined at the beginning of each block. The order of 'left pre−defined' and 'right pre−defined' blocks in the Independent condition was counterbalanced across different participants. Unlike in the Chained condition where Tasks 1 and 2 by definition occurred in the same spatial location, in the Independent task (T1,T2) the two tasks occurred in the same location on half trials and in a different location on the other half. Thus the Independent condition had two sub−conditions, IND−stick (without a location switch) and IND−switch (with a location switch), both of which were randomly mixed within each Independent block of trials. Since the Chained condition involves no switch of spatial attention, the IND−stick trials are a more appropriate comparison to the Chained task.

2.1.5 Procedure

Participants were instructed to respond as quickly and as accurately as possible. Response order was also emphasized in that a Task 1 response should be always followed by a Task 2 response. A verbal instruction was presented on the display in the beginning of each block to inform participants in which block they were (Chained,

Independent Left, or Independent Right). At the beginning of each trial, participants were prompted to press the 'space' key when ready. After a 750−ms blank interval, a black fixation point (radius 0.16^{0}) was presented in the centre of the display and lasted until the presence of feedback provided at the end of each trial. Participants were instructed to perform the tasks while maintaining central fixation. Task 1 stimuli appeared at 1000 ms after the onset of the fixation point and lasted for 100 ms, after which time only the fixation point remained. The onset of Task 2 varied according to the SOA value selected for that particular trial (for the shortest SOA values the offset of Task 1 coincided with the onset of Task 2). Task 2 stimuli remained on the display until participants made their response. Feedback was provided immediately after the occurrence of the response to Task 2 with the fixation point replaced by the feedback. The feedback took the form of two coloured dots appearing to the left and right side of the location previously occupied by the fixation dot. If the participant's response to Task 1 was correct, the left dot was green (red if it was incorrect). The same rule applied for the right dot, which represented Task 2 accuracy. The left dot was yellow if participants made no response to Task 1. The feedback dots remained on the display for 1000 ms before initiation of the next trial.

Response grouping[1] between Tasks 1 and 2 was discouraged by presenting a yellow warning point in the location of the fixation point if Task 2 response (RT2 + SOA) was made in less than 125% of the response to Task 1 measured from Task 1 onset. Two practice blocks (with minimum 32 trials in each block) were given before the formal experiment. The first practice block used only a 1000 ms SOA, though all other parameters remained the same as

---

[1] Response grouping is the tendency for participants to wait for the occurrence of Task 2 before initiating their response to Task 1. Procedure to discourage grouping was adapted from Van Selst and Jolicoeur (1994).

Table 1: Summary of statistical results by using RT, Standard deviation of RT and accuracy as dependent variables in Experiment 1

Contrast	Chained versus Independent/IND–stick						IND–stick versus IND–switch					
	RT1	STD1	Ac1	RT2	STD2	Ac2	RT1	STD1	Ac1	RT2	STD2	Ac2
Condition	*	*		*	*		*		*			*
SOA	*		*	*	*	***	*		*	*	*	*
Condition × SOA						***						

RT1 Reaction time to Task 1; *RT2* Reaction time to Task 2; *Ac1* Accuracy to Task 1; *Ac2* Accuracy to Task 2. *STD1* standard deviation of RT1. STD2 standard deviation of RT2.
Significant main effects or interactions ($p < 0.05$) are indicated by asterisks. *** Effect only significant ($p < 0.05$) in the contrast of *Chained versus. IND–stick* but not in *Chained versus. Independent.*

in the formal experiment, whereas the second practice block was identical to the experimental block. Participants were required to complete both practice blocks with a minimum 85% accuracy level in each before proceeding to the formal experiment. To reduce fatigue, self–controlled breaks between two continuous blocks were provided. The entire experimental session took about 50 min.

2.2 Results

2.2.1 Road Map

Our analysis focuses primarily on the effects of the chaining operation using six dependent variables, i.e., reaction time of Task1[2](RT1) / Task2 (RT2), accuracy of Task 1(Accuracy1) / Task2 (Accuracy2) and standard deviation of RT1 (STD1) / RT2 (STD2) in six two–way ANOVAs (Condition × SOA).

Reaction time and accuracy are standard measures of assessing performance in a PRP paradigm thus constitute our primary dependent variables. For the purpose of simplifying presentation of the results, we have noted in each section that we found no speed versus accuracy

---

[2] For RT analysis, only trials with correct Task 1 and Task 2 responses and correct response order were entered into the analysis. An outlier screening procedure (Van Selst & Jolicoeur, 1994) was used to exclude outlier RTs in each cell for each participant. Less than 3.6% of trials were labelled as outliers in the RT analysis using this approach. Post hoc analyses in each ANOVA of each experiment were conducted using the Bonferroni correction for multiple comparisons.

Fig. 2 Grand means of RT1 and RT2 for Chained, IND–stick and IND–switch conditions with different levels of SOAs in Experiment 1. *Vertical bars* represent standard errors. Marker coding: *Empty squares* for Chained condition; *downward triangles* for IND–stick condition; *upward triangles* for IND–switch condition. *Dotted curves* are for RT1 and *solid curves* for RT2.

trade–off. Our analysis of the standard deviation of RT was to determine if the change in mean RT was paralleled by a change in the variance. This is essential to evaluate the specific prediction of the stochastic evidence accumulation model on RT variability.

With regard to the main independent variable, i.e., Condition, follow–up ANOVAs were performed to compare *Chained* vsrsus *Independent* (between–subject contrast), *Chained* vsrsus *IND–stick* (between–subject), and *IND–stick* vsrsus *IND–switch* (within–subject contrast). We note that the IND–stick condition refers to a sub–condition where, on Independent trials, Task 2 occurred in the same spatial location as Task 1. In contrast, in the IND–switch

condition, Task 2 occurred on the opposite side as that on which Task 1 occurred. Thus the first of these comparisons averages over the two Independent sub-conditions (*switch* and *stick*) to provide an overall assessment of the effects of chaining. The second evaluates the Independent sub-condition (*IND-stick*) most comparable to the Chained condition, as there is no switch of spatial location in either. The third comparison (*IND-stick* vs. *IND-switch*) has been moved to Appendix 1 to facilitate the readability of the results. This comparison is less important as it evaluates only the effect of switching spatial location and is presented in Appendix1.

Table 1 summarizes these results showing the significant main effects and interactions. (Note that main effects and interactions not listed in this and following tables, or not mentioned in the main text, were nonsignificant and are not reported.) Figure 2 shows the grand means of RT1 and RT2 for Chained, IND-stick and IND-switch conditions at different levels of SOAs.

Effect of Condition and SOA on accuracy, RT and standard deviation

2.2.2 Task 1: accuracy

In terms of Task 1 accuracy, there was no significant main effect of Condition between Chained and Independent conditions ($p=0.59$) or between Chained and IND-stick conditions ($p=0.838$). There was, however, a significant main effect of SOA ($F (3, 72) =13.585$, $p<.001$). Task 1 accuracy at 100ms SOA was significantly lower than that at 650ms SOA condition (difference $=-3.0\%$, SE=0.9% ms, $p<.012$). As stated previously, there were no speed-accuracy trade-offs in this or any subsequent analysis.

2.2.3 Task 1: RT

The main effect of Condition was significant in the two-way ANOVA ($F(1, 24) =4.929$, $p<.037$) with RT1 as dependent variable. RT1 in the Chained condition was significantly slower than RT1 in the Independent condition (Difference $=47$ ms, SE=21 ms, $p<.037$) and RT1 in the

IND-stick condition ($F (1, 24) = 6.667$, Difference $=54$ ms, SE=21 ms, $p<.017$). A main effect of SOA on RT1 (F (3, 72) $=7.739$, $p<.001$) was also found. This main effect came from the linear decrease of RT1 from short levels of SOA (100ms, $p<.002$; 200ms, $p<.008$) to larger levels of SOA (350ms). However, there was no significant quadratic trend between SOA and RT1 ($p=0.40$).

2.2.4 Task 1: STD

Using the standard deviation of RT1 as the dependent variable, we found most of the effects with RT to be mirrored in the effects with STD1 in the contrast of Chained versus Independent conditions, as can be observed in Table 1. Particularly, STD1 in the Chained condition was significantly larger than STD1 in both the Independent (Difference $=14$ms, SE=6ms, $p<.039$) and IND-stick conditions (Difference $= 19$ms, SE=6ms, $p<.006$).

2.2.5 Task 2: accuracy

Accuracy analyses on Task 2 revealed no main effect of Condition between Chained and Independent conditions ($p=0.64$). However, T2 accuracy at 100ms SOA was lower compared with other SOA levels, i.e., 200ms (Difference $=-4.9\%$, SE=1.2%, $p<.011$), 350ms (Difference $=-5.3\%$, SE=1.4%, $p<.019$) and 650ms conditions (Difference$= -6.9\%$, SE=1.1%, $p<.001$). As with Task 1, there was no speed-accuracy trade-off.

2.2.6 Task 2: RT

The main effect of Condition was significant ($F (1, 24) = 4.980$, $p<.036$) with RT2 as the dependent variable. RT2 in the Chained condition was significantly slower than both RT2 in the Independent condition (Difference $=64$ ms, SE=29 ms, $p<.036$) and RT2 in the IND-stick condition (Difference$=74$ ms, SE=30ms, $p<.021$). At the same time the two-way interaction of Condition (Chained vs. IND-stick) × SOA was significant ($F(3, 72)=2.791$, $p<.048$). This interaction was derived from the difference in RT2 in the Chained condition being significantly slower than RT2

in the IND–stick condition at 3 levels of SOA, i.e., 100ms (Difference =91ms, SE=36 ms, $p<.021$), 200ms (Difference =81ms, SE=32 ms, $p<.019$) and 350ms (Difference=81ms, SE=32 ms, p<.018), but no significant difference at 650ms SOA (p=0.108, Difference=43 ms).

The main effect of SOA ($F(1.603, 38.478)=167.683$, Greenhouse–Geisser correction, $p<.001$) was also significant in the contrast of Chained versus Independent. The linear trend of SOA was significant ($p<.001$). Further post hoc comparisons demonstrated that RT2 of each smaller SOA level was significantly slower than RT2 of each longer SOA level. From 650ms SOA downwards, the mean differences of RT2 were 44ms (for 350ms SOA, $p<.001$), 125ms (for 200ms SOA, $p<.001$) and 193ms (for 100ms SOA, $p<.001$) respectively. At the same time, highly significant quadratic ($p<.009$) and cubic ($p<.008$) trends between SOA and RT2 were found. Thus, there was a clear rapid increase in RT2 with a decrease in SOA, which is a hallmark pattern of PRP results.

2.2.7 Task 2: STD

Using the standard deviation of RT2 as the dependent variable, we found most of the effects with RT to be mirrored in the effects with STD2 in the contrast of Chained vs. Independent conditions, as can be observed in Table 1. Particularly, STD2 in the Chained condition was significantly larger than STD2 in both the Independent (Difference= 34 ms, SE=14 ms, $p<.023$) and IND–stick conditions ((Difference=47 ms, SE=17ms, p=0.011).

2.3 Discussion

Experiment 1 revealed a PRP effect for both Chained and Independent conditions, i.e., a nonlinear increase in RT2 with the decrease of SOA between the first and second tasks. Confirming predictions of most PRP models, the interaction of Condition × SOA in Experiment 1 reveals that the cost of chaining disappears at longer SOA values, suggesting a disengagement of Task 1 from the central bottleneck. Beyond this well–established finding of

classical dual–task serial costs, the design of Experiment 1 allowed us to further explore the cost of chaining, i.e., piping the results of Task 1 to Task 2. Using a variant of the PRP paradigm, we revealed two aspects of the cognitive cost associated with chaining in Experiment 1. The chaining process is revealed by RT costs in both tasks, taking the form of increased mean and variability. Such outcomes support the stochastic accumulation model, as elaborated later in the "General Discussion".

3 Experiment 2

Experiment 1 was designed to examine the differences between processing two sequentially presented tasks when the second task is either dependent (Chained) on the first or not (Independent). To accomplish this, in the first experiment we employed a standard PRP paradigm, which typically requires a broad range of SOAs to assess the time course of the PRP effect. We employed a between–subject design because we were concerned about carry–over effects between the Chained and Independent conditions. In Experiment 2 our goal was to (1) replicate the outcome of Experiment 1, (2) explore more subtle differences between Chained and Independent conditions, and (3) examine the difference between dual– and single– target (baseline) conditions. Accordingly we chose to repeat the same experimental design from Experiment 1 but to use the same participants in both conditions, to use fewer SOAs, and to implement a single–target baseline control.

In the second experiment we made a few other minor changes as well. As we used a pre–set accuracy criterion (85%) during the two practice blocks of Experiment 1, we were concerned that participants might be over–practiced, in turn rendering accuracy a less sensitive index for measuring task difficulty. Indeed this feature, together with the between–subject design, could explain why we found no significant accuracy difference between the Chained

and IND−stick conditions in Experiment 1. In Experiment 2, we omitted this pre−set accuracy criterion and used only one practice block prior to the experimental block. In Experiment 2 we used only 100ms and 200ms SOA values to gain power. Finally, in Experiment 2 we employed a single/dual task paradigm where half of the trials contained only a single task to evaluate how Task 2 influenced performance in Task 1.

3.1 Method

3.1.1 Participants

Nineteen participants, 12 females and 7 males, aged 18 to 34 (mean=20; STD=3.5) participated in Experiment 2 and were paid by either course credit or cash.

3.1.2 Apparatus and stimuli

The apparatus was the same as in Experiment 1. Stimuli had the same layout as in Experiment 1, except half the trials were dual−task (containing both Task 1 and Task 2 stimuli) while the other half were single−task trials, containing only Task 1.

3.1.3 Design

A three−factor within−participant design was used in this experiment. The first factor, *Trial*, contained two levels: single−task (Task 1 only) and dual−task (Task 1 and 2). This factor was only used when we analysed the performance of Task 1 but not for analysing the performance of Task 2. The second factor, *Condition*, employed blocks of Chained and Independent trials, as in Experiment 1. The third factor, SOA, had 2 levels,

i.e., 100ms and 200ms. The location(s) of the target(s) in both tasks was/were counterbalanced across trials. Every combination of trial type (single−task or dual−task), dual−task relationship (Independent or Chained) and SOA (100ms or 200ms) was repeated forty−eight times, resulting in a total of 384 trials in each experimental session.

The experimental session was divided into 8 blocks with 48 trials for each. Half the trials in each block were single−task and the other half dual−task trials with both trial types randomly mixed within a particular block. Four out of 8 blocks were 'Chained', each containing 24 dual− (T1|T2) and 24 single− task trials. The other four blocks were 'Independent', each containing 24 dual− (T1,T2) and 24 single− task trials. Two of these 4 Independent blocks were 'Left' pre−defined and the other two 'Right' pre−defined. All the other aspects of the design were the same as in Experiment 1.

3.1.4 Procedure

All aspects of the procedure were the same as in Experiment 1 with the following exceptions. First, for a single−task trial, since there was no second target, the fixation point remained for another 1800ms after Task 1 stimulus offset but terminated before the onset of the feedback. Second, the right feedback point was green by default and only became red if participants mistakenly pushed the right−hand joystick either up or down in a single−task trial before the onset of the feedback. Finally, only one practice block (containing 40 trials) was given to participants before the formal experiment. All aspects of the practice block were exactly the same as

Table 2: Summary of statistical results using RT, standard deviation of RT and accuracy as dependent variables in experiment 2

Contrast	Single versus. Dual			Chained versus. IND−stick						IND−stick versus. IND−switch					
	RT1	STD1	Ac1	RT1	STD1	Ac1	RT2	STD2	Ac2	RT1	STD1	Ac1	RT2	STD2	Acc2
Condition	*	*	*	*	*	*	*	*	*	*			*		
SOAᵃ				*	*	*	*		*	*	*	*	*		*
Condition × SOAᵃ				*	*					*					

RT1 Reaction time to Task 1; *RT2* Reaction time to Task 2; *Ac1* Accuracy to Task 1; *Ac2* Accuracy to Task 2. *STD1*: standard deviation of RT1. *STD2*: standard deviation of RT2.
Significant main effects or interactions ($p < 0.05$) are indicated by *asterisks*.
ᵃN/A for ′Single versus. Dual′.
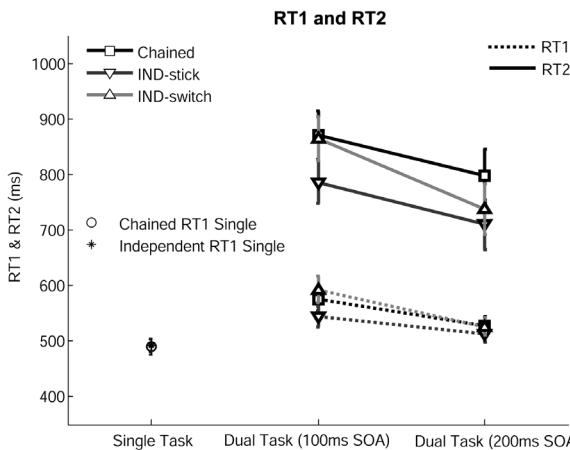
Fig. 3 Grand means of RT1 (*thin curves*) and RT2 (*thick curves)* for different conditions with different levels of SOAs in Experiment 2. *Empty squares* for Chained condition; *downward triangles* for IND–stick condition; *upward triangles* for IND–switch condition; *empty circle* for RT1 of Chained single task condition and *star* for RT1 of Independent single task condition. *Dotted curves* are for RT1 and *solid curves* for RT2. Along *x* axis, the first column is RT1 of single task trials (no T2), the second column is RT1 and RT2 of 100ms SOA condition, the third column is RT1 and RT2 of 200ms SOA condition. *Vertical bars* represent standard errors.

the following experimental blocks. The experimental session lasted approximately 60 min.

3.2 Results

3.2.1 Road map

As in Experiment 1 we explored the effects of chaining by six two–way ANOVAs (Condition × SOA) with the dependent variables (RT1, RT2, Task 1/ Task 2 Accuracy, and the STD of RT1 and RT2). As before, separate ANOVAs were performed using the contrast Chained versus IND–stick and IND–stick versus. IND–switch. A third analysis was done to explore how Task 2 affected Task 1. Table 2 shows the significant main effects and interactions for each analysis. Figure 3 shows the grand means of RT1 (thin curves) and RT2 (thick curves) for each conditions at different levels of SOA.

Effect of Condition and SOA on accuracy, RT and standard deviation

3.2.2 Task 1 performance

For Task 1 accuracy, there was no significant two–way interaction. There were, however, significant main effects of Condition (F (1, 18) = 4.597, $p<.047$) and SOA (F (1,

18) =32. 844, $p<.001$). The chained condition had lower Task1 accuracy (difference = 4%, SE=1.7%, $p<.047$) relative to the IND–stick condition. The Task 1 accuracy in 200ms SOA condition was significantly higher than that in 100ms SOA condition (difference=12%, SE=2%, $p<.001$).

For RT1, the two–way interaction between Condition and SOA was significant (F(1, 18) =5.139, $p<.037$). At 100ms SOA level RT1 of the Chained condition is significantly slower than RT1 of IND–stick condition (Difference= −36ms, SE=7ms, $p<.001$). However, this effect was only marginally significant at 200ms SOA level (Difference= −13ms, SE=7ms, $p=0.078$). Given that RT and accuracy effects are in the same direction, we can conclude that there was no speed–accuracy trade–off.

Turning to the STD of RT1, there was a significant 2–way interaction of Condition and SOA (F (1, 18) = 6.254, $p<.023$). At 100ms SOA, the STD of RT1 in the Chained condition was significantly larger than that of the IND–stick condition (difference =27ms, SE=9.4 ms, $p<.011$). At 200ms SOA, this effect was not significant ($p=0.902$).

3.2.3 Task 2 performance

For Task 2 accuracy, there was no significant 2–way interaction. The main effects of both Chained versus. IND–stick (F(1, 18) =12.251, $p<.004$) and SOA (F(1, 18) = 15.442, $p<.002$) were significant. Task 2 accuracy in the Chained was significantly lower than in the IND–stick condition (difference =−7%, SE=1.9%, $p<.004$). The 200ms SOA condition had higher Task2 accuracy relative to that of 100ms SOA (Difference=−7%, SE=1.8%, $p<.002$). As before there was no speed–accuracy trade–off.

For RT2 as the dependent variable, there was no significant two–way interaction. The main effect of the Chained vs. IND–stick variable was significant (F (1, 18) =11.793, $p<.004$). The RT2 in the chained condition was significantly slower than RT2 in the IND–stick condition (difference =−78 ms, SE=23 ms, $p<.004$). Further analysis

by subtracting individual mean RT1 and mean RT2 of the IND−Stick condition from those in the Chained condition revealed a significantly larger RT cost of chaining in RT2 relative to RT1 (F (1, 18) =6.703, $p<.020$, RT1 cost versus RT2 cost; Difference =53 ms, SE=21 ms, $p<.020$). The main effect of SOA was also significant (F(1, 18) = 23.261, $p<.001$). RT2 in the 200ms SOA condition was significantly faster than RT2 in the 100ms SOA condition (difference =−69 ms, SE=14 ms, $p<.001$).

Examining the STD of RT2, there was no significant two−way interaction between Condition and SOA. The main effect of Chained vs. IND−stick conditions was significant (F(1, 18) = 8.752, $p<.009$). The STD of RT2 in the chained condition was significantly larger than the STD of RT2 in the IND−stick condition (Difference = 60 ms, SE=20 ms, $p<.009$). The main effect of SOA was not significant ($p=0.831$). As with Experiment 1, the results of the analysis examining the spatial switch of attention (IND−stick vs. IND−switch conditions) has been moved to Appendix 1.

3.2.4 Effect of task 2 on task 1: dual versus single tasks

In addition to the dual−target trials required to assess the PRP effect, Experiment 2 introduced single−task trials, using an approach similar to Brisson & Jolicoeur (2007), which made it possible for us to explore how the second task might differentially affect performance of the first between the Chained and Independent conditions. Since there were two types of single−task trials (i.e., Chained and Independent), our first analysis was to evaluate any performance difference between these two types of single−task trials. Paired−samples ttests suggested that there was no significant difference in terms of mean RT1 ($p=0.953$), accuracy of Task 1 ($p=0.758$) and STD of RT1 ($p=0.117$) whether this single−task trial was part of a Chained block or Independent block.

The second analysis examined whether T2 influenced

T1. In the Chained blocks, RT1 of single−task trials were significantly faster than that of dual−task trials at 100ms SOA condition (difference=87 ms, SE=10 ms, $p<.001$) and 200ms SOA condition (difference=36 ms, SE=7ms, $p<.001$). In the Independent blocks, the advantage of RT1 of single−task trials were 51ms in 100ms SOA condition (SE=9 ms, $p<.001$) and 29ms in 200ms SOA condition (SE=8 ms, $p<.004$) relative to dual IND−stick trials.

With regard to the STD of RT1, single−task trials had a significantly smaller STD in Chained blocks at the 100ms SOA level only (difference=56 ms, SE=9ms, $p<.001$) but not at the 200ms SOA level ($p=0.124$). In the Independent blocks, a smaller STD of RT1 in single−task trials was found in both 100ms SOA condition (difference =34ms, SE=7ms, $p<.001$) and 200ms SOA condition (difference= 17ms, SE=7ms, $p<.034$) relative to dual IND−stick trials.

3.3 Discussion

Experiment 2 successfully replicated two important effects of the chaining operation observed in Experiment 1. We observed an increase in both the mean and variance of RT due to chaining. These results provide support for the stochastic accumulation model as discussed below. Given that Experiments 1 and 2 employed different experimental designs and SOA values, taken together, these two experiments suggest the robustness and generality of the effects induced by chaining. By comparing the performance for Task 1 between single− and dual− task trials, Experiment 2 also revealed that all single−task trials had similar RT1 values whether they arose from Chained or Independent blocks. However, the arrival of the second task significantly changed the distribution of RT1, i.e., revealing a larger mean and STD on dual− relative to single− task trials. At the same time, the accuracy of Task 1 was influenced by Task 2 when the SOA was at the shortest (100 ms) duration.

4 General discussion

The present experiments examined the behavioural consequences of chaining, where successful performance requires the output of Task 1 to be piped to Task 2, using an innovative variant of the PRP paradigm. The chaining operation was compared to a task-equivalent 'independent' operation, similar to the standard PRP paradigm, where the output of Task 1 was not required for Task 2. Both experiments, though different in design affording specific conclusions to be drawn, reveal the chaining operation to yield consistently slower and more variable[3] responding. Although it is self evident that two tasks requiring information to be transferred one to the other requires more cognitive resources than the same two tasks not requiring transfer, it is nevertheless important to examine the precise nature of the costs ensued by such a demand.

The present experiments fill an important gap in our understanding of such cognitive demands when compared to a recent study by Sackur and Dehaene (2009). First, these authors did not use an 'independent' condition as used in the present experiment, thus were unable to separate the effect of chaining from temporal proximity. Second, our paradigm used two visuospatial tasks requiring separate responses to each, similar to most PRP studies. Sackur and Dehaene (2009), in contrast employed two numerical tasks that were performed on a single external stimulus, possibly triggering concurrent automatic

---

[3] Most mental operations are Poissonian, increasing the variance linearly with RT; there are, however a few important exceptions (cf. Wagenmakers & Brown, 2007). For example, Sigman and Dehaene (2005) showed that there are manipulations that affect mean RT without affecting the variance, e.g., changing the notation of a number from Arab digits to words in a number comparison task. It was not clear a priori whether the chaining task we inserted between T1 and T2 (which contributes to increased RT) also provided a significant contribution to the variance.

processes. Traditional PRP tasks reveal 'cross-talk' between processing stages in a more controlled way and thus are more sensitive to differences between Chained and Independent conditions. Finally, the Sackur and Dehaene task was subject to stimulus-response (SR) mapping, potentially preventing participants from interacting directly with the underlying cognitive operations. The issue of access to underlying processes is important, as exemplified in the distinction between stimulus-based control versus plan-based control (cf. Tubau, Hommel, & Lóppez-Moliner, 2007) in the linking of actions involved in sequential learning.

We turn now to examine the implications of our results for cognitive models of dual-task processing. We do so in the context of the stochastic accumulation model (Sigman & Dehaene, 2005, 2006) for which the results of the present experiments provide support. Sigman and Dehaene proposed two cognitive operations that serve to explain the chaining effects witnessed in the present experiments. The first operation is 'task setting', i.e., planning the appropriate sequence of actions, which entails a cost to both tasks prior to execution of the first. The second arises from the execution of the chaining operation itself. This operation includes two sub-components, a buffering component to hold information from the first task, and a result-passing component to pipe previously held information to the second task. In theory, the buffering component can be initiated before the completion of Task 1, yielding RT costs to both tasks. Similar imperfect executive control has been observed in a recent study (Sackur & Dehaene, 2009) where they showed that the second operation of a chaining task can start before completion of the first. In contrast, the result-passing component exclusively affects Task 2. The timing of the initiation of the task setting and buffering components is not rigid but subject to factors such as task requirement and SOA. When task setting is initiated earlier, the chaining operation takes longer as

a result of the larger effect on Task 1 due to the earlier initiation of the buffering component. It is worth noting that both sources are intrinsic components of the chaining operation, per se, in contrast to routines generated by two independent operations.

Several findings from the chaining condition of the present study support the existence of these two cognitive operations. First, 'task setting' and 'buffering' affected both tasks, whereas 'result−passing' affected only the second, accounting for the observed relatively small Task 1 but larger Task 2 cost. Second, the model's prediction that task setting and buffering components are subject to task requirements is consistent with the different outcomes observed between Experiments 1 and 2. In Experiment 1 participants knew in advance that all trials were dual−task chained or independent. Thus a rigid task−setting routine could be established prior to the start of each trial in the Chained condition. This is confirmed by the nearly constant RT1 modulation for the Chaining condition across different levels of SOA in Experiment 1. However, in Experiment 2, as participants did not know prior to Task 2 onset whether a given trial was a dual− or single−task trial, initiation of the task setting occurred rather late compared with Experiment 1; consistent with the stochastic accumulation model's prediction that the RT1 chaining cost should be larger in Experiment 1 than Experiment 2. The difference in RT between the Chained vs. Independent conditions, and between Experiments 1 and 2 thus provide support for the stochastic accumulation model.

In addition to changes in RT as discussed above, the stochastic accumulation model predicts greater RT variability arising from the nature of the stochastic process by which evidence is accumulated prior to a response. Here again evidence from the present experiments support the assumptions of this model. Whereas there was a significant change in RT distribution between the Chained and Independent conditions, neither RT1 nor RT2 showed

a difference in variability in either experiment when only a switch in the spatial location of attention distinguished the two conditions (IND−switch v. IND−stick; see Appendix 1). This is consistent with the model's prediction that chaining two cognitive operations inserts an extra processing stage requiring stochastic evidence accumulation, in turn introducing variability, whereas switching the spatial location of attention does not involve such an additional central stage.

Importantly, the existence of a Task 1 *location* switch cost in both experiments supports the notion that the central processing bottleneck is not strictly serial. Rather, the results suggest that at least some resources are allocated to evaluate aspects of T2 (e.g., its spatial location in the independent condition) during the central processing stage of T1. As the classic PRP model predicts that T1 is not affected by aspects of T2, this finding is of considerable interest. Alternative proposals of central "capacity sharing" (Tombu & Jolicoeur, 2003), multiple bottlenecks (Arnell & Duncan, 2002; De Jong, 1993), or coordination by additional executive processes (Logan & Gordon, 2001; Meyer & Kieras, 1997a; Meyer & Kieras, 1997b; Meyer & Kieras, 1999) have received considerable support in their attempts to explain the limitation of processing temporally overlapping tasks. Indeed the chaining cost observed on the first task suggests that the existence of a supervisory control system, likely linked to the functions of planning and execution, which serves to organise hierarchies of cognitive subunits into a complex goal−directed behaviour (Botvinick, 2008). Consistent with this idea, a recent study (Schumacher & Schwarb, 2009) demonstrated that implicit sequence learning of a visual−motor task is impaired by a concurrent auditory discrimination task, and that this effect arises from conflict in the central processing of the two tasks.

To conclude, both experiments consistently demonstrate that the chaining cost on the second task is relatively constant, as long as the central bottleneck is

exerting an effect, i.e., at short SOAs. This implies that serial chaining can operate in parallel with PRP. This conclusion is consistent with a recent study (Miller, Ulrich, & Rolke, 2009), revealing that, in the context of the PRP paradigm, different components of multiple tasks do not always follow a strict serial mode. The authors suggest a balance between a parallel and a serial mode to optimise performance. Moreover, under certain circumstances such as over–practiced second tasks (Maquestiaux, Lagu?– Beauvais, Ruthruff, & Bherer, 2008), the activation of number categories (Fischer, Miller, & Schubert, 2007; Oriet, Tombu, & Jolicoeur, 2005) and valence processing (Fischer & Schubert, 2008), the central bottleneck can even be bypassed as revealed by a reduced PRP effect. This is consistent with the idea proposed by the adaptive executive control (AEC) model (Meyer & Kieras, 1997a,b,1999) that flexible control over secondary–task processing can be exerted and that the 'central decision stages' for two concurrent tasks may temporally overlap rather than follow a strict serial order (as suggested by an immutable structural central bottleneck). The results of the present experiments add a new and important level of understanding to dual– task interference, suggesting that some part of the chaining process, most likely the buffering component, is initiated before the completion of Task 1 and proceeds in a partially parallel way with Task 1. These results are consistent with recent literature (Sackur & Dehaene, 2009), supporting the claim that serial chaining is a relatively slow and effortful process that consumes central processing resource and requires persistent conscious control.

Appendix 1

Spatial location switching effects: Experiment 1

With regard to Task 1 RT, a significant main effect of Condition [$F(1, 12)=4.774$, $p<0.05$] was found comparing

the IND–stick to the IND–switch condition. RT1 of the IND–stick condition was 12 ms (SE=5 ms) faster than RT1 of IND–switch condition. This effect was also mirrored on the accuracy of Task 1 with significantly higher accuracy of Task 1 [$F(1, 12)=9.452$, $p<0.011$; difference = 1.6%, SE=0.5%] in the IND–stick condition relative to the IND– switch condition. However, this effect was not significant with regard to the standard deviation of RT1 ($p=0.972$).

Turning to Task 2 RT, there was no significant main effect of Condition ($p=0.191$). However, switching the spatial location of attention had an effect on Task 2 accuracy [$F(1, 12)=9.452$, $p<0.011$]. Accuracy of Task 2 was significantly higher (difference=1.6%, SE=0.5%, $p<0.011$) for the IND– stick relative to the IND–switch condition. However, this effect of Condition difference was not significant with regard to the standard deviation of RT2 ($p=0.313$).

Spatial location switching effects: Experiment 2

For RT1, there was a significant two–way interaction of Condition and SOA [$F(1, 18)=10.585$, $p<0.005$] when comparing the IND–stick to IND–switch conditions. Further post hoc comparisons revealed that this two–way interaction arose when RT 1 in the 100 ms SOA IND– stick condition was significantly faster than RT1 in the IND–switch condition (difference=–44 ms, SE =11ms, p<0.002); however, there was no significant differences at 200 ms SOA ($p=0.199$). The main effect of Condition was significant [$F(1, 18)=11.409$, $p<0.004$]; RT1 of IND–stick condition was significantly faster than RT1 of IND–switch condition (difference=–27 ms, SE=8 ms, $p<0.004$). The main effect of SOA was also significant [$F(1, 18) = 21.517$, $p<0.001$] with RT1 in 200 ms SOA condition significantly faster than RT1 in 100 ms SOA condition (difference=–46 ms, SE=10 ms, $p<0.001$). Turning to Task 1 accuracy, there was no significant two–way interaction. The only significant main effect was SOA [$F(1, 18)= 18.237$, $p<0.001$] indicating

Task 1 accuracy in 200 ms SOA condition was significantly better than that in 100 ms SOA condition (difference = 10%, SE=2.4%, $p<0.001$). As before, there was no evidence of a speed－accuracy trade－off.

Looking at the STD of RT1, there was no significant twoway interaction. The main effect of the IND－stick versus INDswitch comparison was not significant (p=0.754). The main effect of SOA was significant [F(1, 18)=17.026, $p<0.002$]. The STD of RT1 in 200 ms SOA condition was significantly smaller than that in 100 ms SOA condition (difference=−26 ms, SE=6 ms, $p<0.002$).

Turning to RT2, there was no significant two－way interaction. The main effect of the IND－stick versus INDswitch comparison was significant [F(1, 18)=9.881, $p<0.007$]. RT2 in IND－stick condition was significantly faster than RT2 in the IND－switch condition (difference=−47 ms, SE=15 ms, $p<0.007$). The main effect of SOA was also significant [F(1, 18) =23.689, p < 0.001]. RT2 in the 200 ms SOA condition was significantly faster than RT2 in the 100 ms SOA condition (difference=−96 ms, SE=20 ms, p < 0.001). For Task 2 accuracy, there was no significant two－way interaction. The only significant main effect was SOA [F(1, 18)=15.829, $p<0.002$]. Task 2 accuracy at 200 ms SOA condition was significantly higher than Task 2 accuracy in the 100 ms SOA condition (difference=6%, SE=1.5%, $p<0.002$). As before, accuracy results do not provide any support for a speed－accuracy trade－off.

With regard to the STD of RT2, there was no significant two－way interaction (p=0.270). The main effect of the IND-stick versus IND－switch comparison was not significant (p=0.093). The main effect of SOA was not significant (p=0.276).

References

Arnell, K. M., & Duncan, J. (2002). Separate and shared sources of dual－task cost in stimulus identification and response selection. *Cognitive Psychology*, *44*, 105–147.

Botvinick, M. M. (2008). Hierarchical models of behavior and prefrontal function. *Trends in Cognitive Sciences*, *12*(5), 201–208.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10,* 433–436.

Brisson, B., & Jolicoeur, P. (2007). Cross－modal multitasking processing deficits prior to the central bottleneck revealed by event－related potentials. *Neuropsychologia*, *45*(13), 3038–3053.

Chun, M. M., & Potter, M. C. (1995). A two－stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception & Performance*, *21*, 109–127.

De Jong, R. (1993). Multiple bottlenecks in overlapping task performance. *Journal of Experimental Psychology*: *Human Perception and Performance*, *19*, 965–980.

Fischer, R., Miller, J., & Schubert, T. (2007). Evidence for parallel semantic memory retrieval in dual tasks. *Memory & Cognition*, *35*(7), 1685–1699.

Fischer, R., & Schubert, T. (2008). Valence processing bypassing the response selection bottleneck? Evidence from the psychological refractory period paradigm. *Experimental Psychology*, *55*(3), 203–211.

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Sciences*, *5*, 10–16.

Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, *36*, 299–308.

Kawahara, J. (2003). The effect of observer's set on the processing of temporally distributed items. *Japanese Psychological Research*, *45*, 109–114.

Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual－task situations. *Psychological*

Review, *108*, 393–434.

Maquestiaux, F., Laguë–Beauvais, M., Ruthruff, E., & Bherer, L. (2008). Bypassing the central bottleneck after single–task practice in the psychological refractory period paradigm: Evidence for task automatization and greedy resource recruitment. *Memory & Cognition*, *36*, 1262–1282.

Meyer, D. E., & Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple–task performance: Part 1. *Basic mechanisms*. *Psychological Review*, *104*, 3–65.

Meyer, D. E., & Kieras, D. E. (1997b). A computational theory of executive cognitive processes and multipletask performance: Part 2. Accounts of psychological refractory–period phenomena. *Psychological Review*, *104*, 749–791.

Meyer, D. E., & Kieras, D. E. (1999). Precis to a practical unified theory of cognition and action: Some lessons from computational modeling of human multiple–task performance. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII* (pp. 15–88). Cambridge, MIT Press.

Miller, J., Ulrich, R., & Rolke, B. (2009). On the optimality of serial and parallel processing in the psychological refractory period paradigm: Effects of the distribution of stimulus onset asynchronies. *Cognitive Psychology*, *58*, 273–310.

Oriet, C., Tombu, M., & Jolicoeur, P. (2005). Symbolic distance affects two processing loci in the number comparison task. *Memory & Cognition*, *33*(5), 913–926.

Pashler, H. (1984). Processing stages in overlapping tasks: Evidence for a central bottleneck. *Journal of Experimental Psychology: Human Perception & Performance*, *10*, 358–377.

Pashler, H. (1994). Dual–task interference in simple tasks: Data and theory. *Psychological Bulletin*, *116*, 220–244.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*, 437–442.

Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review*, *95*, 238–255.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: an attentional blink? *Journal of Experimental Psychology: Human Perception & Performance*, *18*, 849–860.

Sackur, J., & Dehaene, S. (2009). The cognitive architecture for chaining of two mental operations. *Cognition*, *111*(2), 187–211.

Schumacher, E. H., & Schwarb, H. (2009). Parallel Response Selection Disrupts Sequence Learning Under Dual–Task Conditions. *Journal of Experimental Psychology: General, 138*(2), 270–290.

Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1994). Attention to visual pattern information produces the attentional blink in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception & Performance*, *20*, 357–371.

Sigman, M., & Dehaene, S. (2005). Parsing a cognitive task: a characterization of the mind's bottleneck. *PLoS Biology*, *3*(2), e37.

Sigman, M., & Dehaene, S. (2006). Dynamics of the central bottleneck: Dual–task and task uncertainty. *PLoS Biology*, *4*(7), e220 doi:210.1371/journal. Pbio. 0040220.

Tombu, M., & Jolicoeur, P. (2003). A central capacity sharing model of dual–task performance. *Journal of Experimental Psychology: Human Perception & Performance*, *29*(1), 3–18.

Tubau, E., Hommel, B., & López–Moliner, J. (2007). Modes of executive control in sequence learning: From

stimulus−based to plan−based control. *Journal of Experimental Psychology: General*, *136*, 43−63.

Turing, A. M. (1936) On computable numbers, with an application to the Entscheidungs problem.*Proceedings of the London Mathematical Society, 42*.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: the leaky, competing accumulator model. *Psychological Review, 108*, 550−592.

Van Selst, M., & Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *The Quarterly Journal of Experimental Psychology*, *47*(3), 631−650.

Wagenmakers, E.−J., & Brown, S. (2007). On the linear relation between the mean and the standard deviation of a response time distribution. *Psychological Review*, *114*, 830−841.

Decreased resting-state connections within the visuospatial attention-related network in advanced aging

Yujie Li[1,2,5], Chunlin Li[3], Qiong Wu[5], Zhihan Xu[5], Tomoko Kurata[6], Seiichiro Ohno[7], Susumu Kanazawa[8], Koji Abe[6], Jinglong Wu[4, 5,*]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan 430079, China)

([2] School of Psychology, Central China Normal University, Wuhan 430079, China)

([3] School of Biomedical Engineering, Capital Medical University, You An Men, Beijing, China)

([4] Intelligent Robotics Institute，School of Mechatronics Engineering，Beijing Institute of Technology，Beijing，China)

([5] Biomedical Engineering Laboratory, Graduate School of Natural Science and Technology, Okayama University, Japan)

([6] Department of Neurology, Graduate School of Medicine, Dentistry and Pharmaceutical Science, Okayama University, Okayama, Japan)

([7] Central Division of Radiology, Okayama University Hospital, Okayama, Japan)

([8] Department of Radiology, Graduate School of Medicine, Dentistry and Pharmaceutical Sciences, Okayama University, Okayama, Japan)

Abstract Advanced aging is accompanied by a decline in visuospatial attention. Previous neuroimaging and electrophysiological studies have demonstrated dysfunction in specific brain areas related to visuospatial attention. However, it is still unclear how the functional connectivity between brain regions causes the decline of visuospatial attention. Here, we combined task and rest functional magnetic resonance imaging (fMRI) to investigate the age−dependent alterations of resting−state functional connectivity within the task−related network. Twenty−three young subjects and nineteen elderly subjects participated in this study, and a modified Posner paradigm was used to define the region of interest (ROI). Our results showed that a marked reduction in the number of connections occurred with age, but this effect was not uniform throughout the brain: while there was a significant loss of communication in the anterior portion of the brain and between the anterior and posterior cerebral cortices, communication in the posterior portion of the brain was preserved. Moreover, the older adults exhibited weakened resting−state functional connectivity between the supplementary motor area and left anterior insular cortex. These findings suggest that the disrupted functional connectivity of the brain network for visuospatial attention that occurs during normal aging may underlie the decline in cognitive performance.

Keywords posner task; visual−spatial attention; aging effect; functional magnetic resonance imaging (fMRI); resting−state

1 Introduction

Visuospatial attention is often studied within the context of the Posner paradigm. In this paradigm, a spatially informative cue is presented to indicate the location of an upcoming peripheral target. Targets appearing in the predicted location (valid) are detected more rapidly and accurately than those that are not (invalid) Over the years, a considerable number of functional magnetic resonance imaging (fMRI) studies have revealed a frontoparietal network involved in controlling visuospatial attention using the Posner paradigm. This network consists of the anterior insular cortex (AIC), supplementary motor area (SMA) and dorsolateral prefrontal cortex (DLPFC) as well as the intraparietal sulcus (IPS), frontal eye field (FEF) and middle temporal area (MT+).

As a result of normal aging, a decline in various cognitive domains has been commonly observed. Using the Posner paradigm, previous studies have shown that elderly subjects responded to the target more slowly than young subjects. One possible reason for this reduction in visuospatial attention is the dysfunction of specific gray matter areas. Electrophysiological studies have demonstrated that aging has an impact on the neural activity in those attention-related brain regions. In addition to changes in task-relevant regions, alterations in communication between the different nodes of brain networks have also been found to be correlated with normal aging For example, older adults' poorer performance on working memory tasks and emotional memory tasks have been shown to be associated with increases and decreases in connectivity within and/or between task-related networks. Therefore, an investigation on the interactions between brain areas provides insight into the neural mechanisms underlying age-related changes.

Recently, resting-state fMRI (rs-fMRI) has become a powerful tool for understanding the functional organization of the human brain. Based on the synchrony of spontaneous fluctuations in the blood oxygenation level dependent (BOLD) signal from functionally related brain regions, several large-scale coherent spatial patterns, namely resting-state networks (RSNs), have been identified. An effect of aging on resting-state functional connectivity has been found within and/or between RSNs, such as the default mode network (DMN) and motor network. In addition, age-related differences in functional connectivity have also been revealed between brain areas pertaining to the visuospatial attention-related network. Specifically, the authors identified ROIs in the IPS, FEF and MT+, which comprised corresponding areas in both hemispheres, and observed age-related reductions in the resting-state functional connectivity between the IPS and MT+. Since the functional connectivity between the SMA and AIC play an important role in initiating and maintaining task-level control and in suppressing irrelevant distracting information, it has a great influence on behavioural performance during a visuospatial attention task. Therefore, including the SMA and AIC in a functional connectivity analysis is necessary to understand the effect of aging on the resting-state functional connectivity within the entire system that participates in visuospatial attention.

In the current study, we investigated age-related changes in resting-state functional connectivity between regions that are associated with visuospatial attention by combining task and rest fMRI. A modified version of Posner paradigm was used to determine visuospatial attention-evoked brain activation based on which eleven spherical ROIs (seeds) were defined. These ROIs were centered in the SMA, bilateral AIC, bilateral DLPFC, bilateral FEF, bilateral IPS and bilateral MT+. Then, for each subject, the mean time series within each ROI was extracted from the resting-state fMRI data and Pearson correlation coefficients were calculated between

every possible pair of ROIs. Finally, age-dependent differences in resting-state functional connectivity were examined using two-sample t-tests, corrected for multiple comparisons. The results showed a marked reduction in either the number of connections or the strength of interactions in older adults.

2 Methods

2.1 Subjects

Twenty-three healthy young volunteers (ages 21-32; mean 22.7; male/female 23/0) and 19 healthy older volunteers (ages 60-78; mean 66.5; male/female 16/3; MMSE score 29.5 ± 0.1) took part in the fMRI experiment. All subjects had normal or corrected-to-normal vision and reported that they were all right-handed. None of the subjects had a history of neurological or psychiatric dysfunction or a previous experience in a neuropsychological experiment. The study was approved by the ethics committee of Okayama University, and written informed consent was obtained before the study. Three young subjects with excessive head movements and one older subject for whom fMRI data acquisition failed were excluded. In addition, although we failed to record behavioural data for three other older subjects, their imaging data were available and analysed in our study.

2.2 Experimental design

Resting-state fMRI data were first recorded with one scan when subjects were instructed to keep their eyes closed, not to fall asleep and not to think of anything in particular. This was followed by one scan during a simple visual spatial attention task consisting of 120 trials (Fig. 1A). Each trial began with the fixation display followed by an arrow appearing at the centre of the visual field. This arrow was presented for 200 ms and served as a cue, instructing the subjects to pay attention to the left or right visual field. After an inter-stimulus interval (ISI) of 200,

400, or 800 ms, the target appeared for 100 ms on the side indicated by the arrow 90% of the time (valid trial) and on the opposite side 10% of the time (invalid trial). Subjects were instructed to indicate whether a target appeared in the left or right visual field by pressing the left or right key with the forefinger or middle finger of their right hand, respectively. The duration of each trial was 3000 ms, and there were equal numbers of left and right directional cues and targets on each side. Subjects were asked to hold their gaze on the central fixation cross throughout the trial and to press the key as quickly and accurately as possible. Stimuli were presented through a projector onto a paper screen located in front of the subjects' feet. Subjects viewed the screen via a 45 degree angled mirror attached to the head-coil of the MRI setup.

2.3 Data acquisition

All subjects were imaged using a 1.5 T Philips scanner vision whole-body MRI system (Okayama University Hospital, Okayama, Japan), which was equipped with a head coil. The imaging area consisted of 32 functional gradient-echo planar imaging (EPI) axial slices (TR=3000 ms, TE=50 ms, FA=90°, acquisition matrix=80×79, FOV=240 mm^2, slice thickness=4mm, gap=0.5 mm) that were used to obtain T2*-weighted fMRI images in the axial plane. We obtained 176 functional volumes for the resting-state session and 124 functional volumes for the task run. The first 4 images of each functional scan were discarded to allow for the equilibration of the magnetic field. After the EPI scans, a T1-weighted 3D magnetization-prepared rapid acquisition gradient echo (MP-RAGE) sequence was acquired (TR=9.4 ms, TE=4.6 ms, FA=10°, acquisition matrix=240×240, voxel size=1×1×1mm^3, 200 contiguous axial slices).

2.4 fMRI data analysis

2.4.1 *Preprocessing*

The imaging data were analysed using statistical parametric mapping software (SPM8; Wellcome

Department of Cognitive Neurology, London, UK; http://www.fil.ion.ucl.ac.uk/spm) running under the MATLAB environment (Version 7.8; The MathWorks, Inc., Natick, MA). To correct for differences in slice acquisition time, all images were synchronised to the middle slice. Subsequently, images were spatially realigned to the first volume due to head motion. Subjects whose head movements exceeded 3 mm on any axis or rotations greater than 3° were excluded. After the correction, the imaging data were normalised to the Montreal Neurological Institute (MNI) EPI template supplied with SPM8 (resampled to $3 \times 3 \times 3 \text{ mm}^3$ voxels). Finally, the normalised images were smoothed with a Gaussian kernel of 8-mm full-width half-maximum (FWHM).

2.4.2 Task activation and ROIs

For the task run, the time series at each voxel for each

subject were high-pass filtered at 128 s to remove low-frequency artifacts, and were temporally corrected for autocorrelations using the AR(1) model in SPM8. Then a statistical analysis based on a general linear model (GLM) was conducted. When only the canonical hemodynamic response function (HRF) was included in the GLM, the delay-induced amplitude bias is observed. Hence, in the current study, a "derivative boost" was utilised to counteract this effect, as was suggested previously. First, regressors were normally generated by convolving the stimulus-function corresponding to each experimental condition (valid cue or invalid cue), which was a sum of the delta functions with the SPM8 canonical HRF and its temporal derivative. After orthogonalization and normalization of the regressors, the model was fit to the data and beta images were obtained. Then, the "derivative

Seed region	MNI coordinates		
	x	y	z
Supplementary motor area (SMA)	3	12	51
Left anterior insular cortex (L-AIC)	-42	3	0
Right anterior insular cortex (R-AIC)	42	18	-6
Left dorsolateral prefrontal cortex (L-DLPFC)	-42	24	18
Right dorsolateral prefrontal cortex (R-DLPFC)	42	30	18
Left frontal eye field (L-FEF)	-39	0	57
Right frontal eye field (R-FEF)	42	3	54
Left intraparietal sulcus (L-IPS)	-36	-51	45
Right intraparietal sulcus (R-IPS)	33	-51	45
Left middle temporal area (L-MT+)	-48	-63	-3
Right middle temporal area (R-MT+)	51	-60	0

Fig. 1. Diagrammatic representation of the visual display and trial design (A). (B) shows the locations of region of interests (ROIs). Activation maps for young (C) and older (D) subjects during the visuospatial attention task (FDR corrected $p < 0.05$). Eleven spherical ROIs were defined based on the activation map for young group (labeled). For abbreviations, please refer to Fig. 1B. L: left; R: right.

boost" was calculated for each voxel as follow:

$$H=\text{sign}(\hat{\beta}_1)\sqrt{\hat{\beta}_{1+}^2\hat{\beta}_2^2}$$

where   and   are the least−squares estimates of the activation amplitude for the canonical HRF and the temporal derivative term, respectively. These resulting "combined" images from each group were entered into second−level one−sample t−tests to yield group−level activations. A p<0.05 false discovery rate (FDR) corrected for multiple comparisons was considered significant. Based on a statistical parametric map for the young group, eleven spherical regions of interest (ROIs) were created with a radius of 8 mm centered at the voxels with the maxima local T values in the attention−related regions.

2.4.3 *Functional connectivity analysis*

For the resting−state run, using the DPARSF software (V2.3; http://rfmri.org/DPARSF) and REST toolkit (V1.8; http://restfmri.net/forum/REST), the preprocessed imaging data were removed from the linear trend of the time series and were temporally band−pass filtered (0.01−0.08 Hz) to reduce the effects of low−frequency drifts and high−frequency physiological noises. Then, several sources of spurious variance, including the six estimated head motion parameters and the average time series in the cerebrospinal fluid and white matter regions, were removed from the data through a linear regression. For each subject, the mean time series of each ROI, which was defined based on their attention−related activation maps, was obtained by simply averaging the time series of all voxels within that region. To measure the functional connectivity among regions, we calculated the Pearson correlation coefficients between any possible pair of regional time series, and then obtained a temporal correlation matrix (11 × 11) for each subject. We applied Fisher's r−to−z transformation to improve the normality of the correlation matrix. Then, two−tailed one−sample t−tests were performed for all the 55 possible [i.e.,

(11 × 10)/2] pairwise correlations across subjects according to each group to examine whether each inter−regional correlation significantly differed from zero. Moreover, to explore age−related changes in functional connectivity, two−tailed two−sample t−tests were conducted to compare pairwise correlations between the young and elderly groups. A p<0.05, corrected for multiple comparisons by the Bonferroni−Holm method, was considered significant.

2.4.4 *Voxel−based morphometry*

We did voxel−based morphometry analysis on the T1 images using SPM8 and conducted a two−sample t−test to assess between−group (young versus older) gray matter differences. The result shows aging−associated cortical atrophy in some specific regions within normal older adults, most prominently in sensorimotor and prefrontal areas, which is consistent with previous study (see supplementary material for more details).

3 Results

The behavioural data were collected from 20 young subjects and 15 elderly subjects. Reaction times (RTs) faster than 100 ms or slower than 1000 ms were excluded from the analysis. The mean RT and mean accuracy for the young group were 374.22 ± 61.80 ms and 98.89 ± 1.30%, respectively, while for the elderly group, they were 418.59 ± 63.73 ms and 92.78 ± 11.93%, respectively. Two−tailed two−sample t−tests were conducted to compare the mean RTs and mean accuracies between the healthy young and elderly adults (SPSS 16.0). The results showed that elderly subjects responded to the target significantly more slowly than young subjects did (t(33)=−2.075, p=0.046), whereas no significant difference in accuracy was found between the groups (t(33)=1.976, p=0.068, corrected for heterogeneity of variance).

Consistent with previous studies, the attention task elicited activation across a distributed frontoparietal
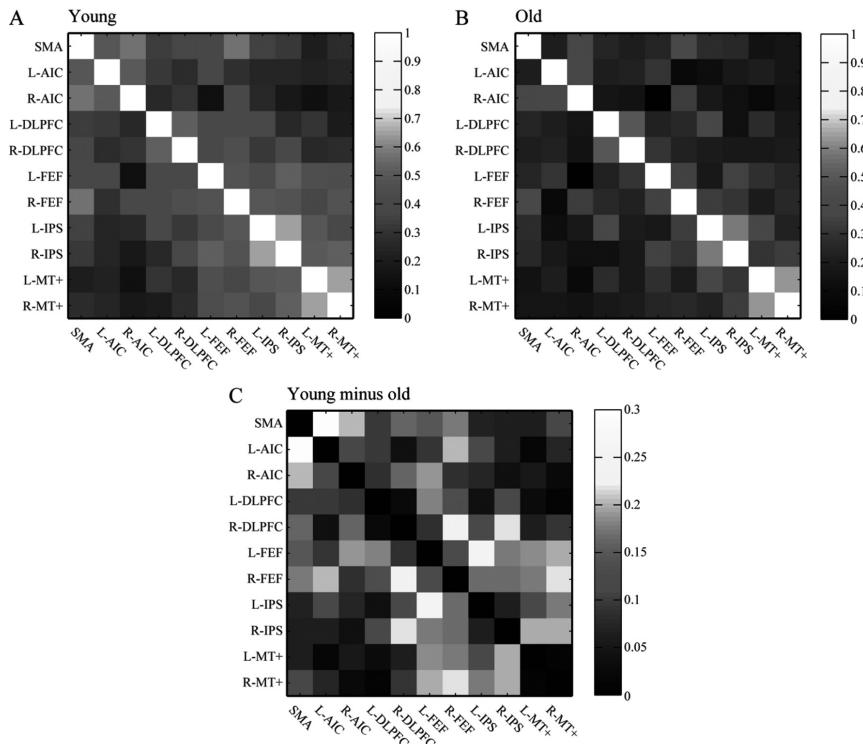
Fig. 2. Correlation matrices representing resting–state functional connectivity within visuospatial attention–related network for young (A) and older (B) subjects. Color bar indicates Pearson correlation coefficients for each region pair, stronger for lighter cells and weaker for darker cells. The comparison (C) of both groups shows a significant reduction in functional connectivity between left AIC and SMA (nearly white cell). For abbreviations please refer to Fig. 1B.

network in both young and elderly adults (Fig. 1C and 1D). Unfortunately, no significant aging effect on task–related activation pattern was found. Next, eleven spherical ROIs were defined based on the activation map for the young group (see Fig. 1B). The centre of each ROI was placed on the brain area that well known to be involved in attention, such as the bilateral FEF, bilateral IPS, bilateral MT+, bilateral AIC, bilateral DLPFC and SMA.

A seed–based functional connectivity analysis was conducted on the resting–state fMRI data by calculating the inter–regional correlation between all possible pairs of ROIs. The results are shown in Fig. 2A and 2B for both the young and the elderly groups. A significant positive correlation could be observed in 52 of the 55 pairs of ROIs in the young group (p<0.05, corrected for multiple comparisons). By contrast, only 22 of the 55 pairs of ROIs were positively correlated in the elderly group (p<0.05, corrected for multiple comparisons). Specifically, in the older adults, most of the connections between the seven ROIs located in the anterior portion of the brain decreased (8/21), whereas the connections between the four ROIs defined in the posterior cerebral cortex were entirely preserved (6/6). In addition, a marked loss of connections between the anterior ROIs and posterior ROIs also occurred (8/28). In addition to a decrease in the number of connections, the correlation matrix map for the elderly subjects exhibited a darker colour than that for the young subjects, implying overall reduction in the strength of interactions with age (Fig. 2C). Furthermore, the age–related changes in functional connectivity could be directly revealed by comparing pairwise correlations between the groups. One pair of ROIs, the SMA with the left AIC, had reduced functional connectivity with age, whereas no age–related increase in functional connectivity was found (p<0.05, corrected for multiple comparisons).

4 Discussion

The goal of this research is to fully explore the effects of normal aging on the resting-state functional connectivity between the visuospatial attention-related brain regions. By combining task and rest fMRI, we found prominent age-related changes in the intrinsic functional organization of the brain, which were a marked reduction in either the number of connections or the strength of interactions. Moreover, the elderly subjects exhibited significantly decreased functional connectivity between the SMA and left AIC. Our results indicate that the disruption of the functional brain network for visuospatial attention that occurs as a result of normal aging may underlie the decline in cognitive performance.

Consistent with previous studies, elderly adults had slower RTs and a tendency to be less accurate in the task, which reflects a deterioration of visuospatial attention with age. This cognitive decline in normal aging has been considered to arise from changes in the communications between different brain areas, as well as functional alterations in task-related activation patterns . The rs-fMRI findings provide evidence for the former "disconnection" hypothesis, since the older adults showed a decreased number of connections and reduced strength of the interactions within the entire visuospatial attention-related network. More precisely, the loss of communications mainly occurred between the seven anterior ROIs as well as between the anterior and posterior ROIs, while correlations between the four posterior ROIs were preserved. In other words, the age-related deterioration observed in resting-state functional connectivity was not uniform throughout the brain. These phenomena are consistent with the previous report of the disruption of structural connectivity in the aging brain, given that resting-state functional connectivity can be predicted from structural connectivity. In that study, diffusion tensor imaging (DTI) was used to assess diffusion anisotropy in white matter tracts in young and elderly subjects, and a similar change in the structural integrity of the brain was observed: the largest proportional loss of anisotropy associated with age was found in the anterior regions, there was intermediate but significant loss in the middle regions, and no significant loss in the posterior regions. Therefore, our results for the reduction in resting-state functional connectivity in the older adults may reflect the deterioration in structural connectivity and provide a plausible neural basis for the decline in visuospatial attention. Moreover, the difference from previous study in which an age-related decrease in resting-state functional connectivity between the IPS and MT+ was observed, may result from the different approach for defining ROIs that was taken in this study.

In addition to a decrease in the number of connections, the elderly subjects exhibited an overall reduction in the strength of their interactions. A further analysis revealed that significantly attenuated functional connectivity occurred between the SMA and left AIC in the aging brain. While the SMA has been found to play a role in integrating the neural information relevant for the internal generation of movements, the AIC has been shown to be involved in emotional and interoceptive processing. Moreover, a "salience network", which consists of the pre-SMA/SMA and bilateral AIC, has been proposed to integrate highly processed sensory data with visceral, autonomic, and hedonic "markers," so that the organism can decide what to do (or not to do) next. In addition, in an influential study, Dosenbach et al.suggested a "core system" which similarly contains the pre-SMA/SMA and bilateral AIC and has been thought to be associated with the stable maintenance of task sets. It is noteworthy that the present ROIs centered in the SMA and bilateral AIC were defined in a very similar position as the corresponding regions

belonging to either the "salience network" or the "core system". Moreover, in the current study, a slower RT and tendency toward lower accuracy during the task were observed for elderly subjects, which are consistent with the previous study. Hence, the weakened resting-state functional connectivity between the SMA and left AIC may imply that elderly subjects have difficulties in stably maintaining task sets and effectively suppressing task-irrelevant distracting information during a visuospatial attention task.

In conclusion, the current study showed that aging has prominent effects on the resting-state functional connectivity within the entire visuospatial attention-related network. Our results suggest that the changes in the connections between brain regions while at rest may provide valuable information for understanding the neural mechanisms underlying cognitive aging.

References

B. Biswal, F. Zerrin Yetkin, V.M. (1995). Haughton, J.S. Hyde, Functional connectivityin the motor cortex of resting human brain using echo-planar MRI, Magn. Reson. *Med, 34,* 537‐541.

C. Honey, O. Sporns, L. Cammoun, X. (2009). Gigandet, J.-P. Thiran, R. Meuli, P.Hagmann, Predicting human resting-state functional connectivity fromstructural connectivity. *Proc. Nat. Acad. Sci. U. S. A. 106,* 2035‐2040.

C. Roski, S. Caspers, R. Langner, A.R. Laird, P.T. Fox, K. Zilles, K. Amunts, S.B.Eickhoff. (2013). Adult age-dependent differences in resting-state connectivity withinand between visual-attention and sensorimotor networks. *Front. AgingNeurosci. 5.*

D.A. Robin, M. Rizzo.(1992). Orienting attention in audition and between audition andvision: Young and elderly subjects. *J. Speech Hearing Res. 35,* 701‐707.

D.J. Madden. (1990). Adult age differences in the time course of visual attention. *J. Gerontol. 45,* 9‐16.

D.R. Gitelman, A.C. Nobre, T.B. Parrish, K.S. LaBar, Y.-H. Kim, J.R. Meyer, M.-M.Mesulam. (1999). A large-scale distributed network for covert spatial attention Furtherattention Further anatomical delineation based on stringent behavioral and cognitivecontrols. *Brain 122,* 1093‐1106.

F. Doricchi, E. Macci, M. Silvetti, E. Macaluso. (2010). Neural correlates of the spatialand expectancy components of endogenous and stimulus-driven orienting of attention in the Posner task. *Cereb. Cortex, 20,* 1574‐1585.

F. Kurth, K. Zilles, P.T. Fox, A.R. Laird, S.B. Eickhoff. (2010). A link between thesystems: functional differentiation and integration within the human insularevealed by meta-analysis. *Brain Struct. Funct. 214,* 519‐534.

J. Damoiseaux, C. Beckmann, E.S. Arigita, F. Barkhof, P. Scheltens, C. Stam, S.Smith, S. Rombouts. (2008). Reduced resting-state brain activity in the defaultnetwork in normal aging. *Cereb. Cortex, 18,* 1856‐1864.

J. Damoiseaux, S. Rombouts, F. Barkhof, P. Scheltens, C. Stam, S.M. Smith, C.Beckmann. (2006). Consistent resting-state networks across healthy subjects. *Proc. Nat. Acad. Sci. U. S. A. 103,* 13848‐13853.

J.R. Andrews-Hanna, A.Z. Snyder, J.L. Vincent, C. Lustig, D. Head, M.E. Raichle,R.L. Buckner. (2007). Disruption of large-scale brain systems in advanced aging. *Neuron, 56,* 924‐935.

J. Steffener, C.G. Habeck, Y. Stern. (2012). Age-related changes in task relatedfunctional network connectivity. *PLoS One, 7,* e44421.

J. Steffener, M. Tabert, A. Reuben, Y. Stern. (2010). Investigating hemodynamicresponse variability at the

group level using basis functions, *Neuroimage, 49*, 2113－2122.

J.T. Coull, A.C. Nobre. (1998). Where and when to pay attention: the neural systemsfor directing attention to spatial locations and to time intervals as revealed byboth PET and fMRI. *J. Neurosci, 18*, 7426－7435.

K.R. Van Dijk, T. Hedden, A. Venkataraman, K.C. Evans, S.W. Lazar, R.L.Buckner. (2010). Intrinsic functional connectivity as a tool for human connectomics:theory, properties, and optimization. *J. Neurophysiol, 103*, 297－321.

L. Geerligs, N.M. Maurits, R.J. Renken, M.M. Lorist. (2014). Reduced specificity offunctional connectivity in the aging brain during task performance. *Hum.Brain Mapp, 35*, 319－330.

L. Geerligs, R.J. Renken, E. Saliasi, N.M. Maurits, M.M. Lorist. (2014). A brain－widestudy of age－related changes in functional connectivity. Cereb. *Cortex*. bhu012.

L. Lorenzo－López, S. Doallo, C. Vizoso, E. Amenedo, S.R. Holguín, F. (2002). Cadaveira,Covert orienting of visuospatial attention in the early stages of aging. *NeuroRep, 13*, 1459－1462.

M.A. Lindquist, J. Meng Loh, L.Y. Atlas, T.D. Wager. (2009). Modeling thehemodynamic response function in fMRI: efficiency, bias and mis－modeling. *Neuroimage, 45*, S187－S198.

M.D. Fox, M.E. Raichle. (2007). Spontaneous fluctuations in brain activity observedwith functional magnetic resonance imaging, *Nat. Rev. Neurosci. 8,* 700－711.

M.D. Greicius, K. Supekar, V. Menon, R.F. Dougherty. (2009). Resting－state functionalconnectivity reflects structural connectivity in the default mode network, Cereb. *Cortex, 19,* 72－78.

M. O'sullivan, D. Jones, P. Summers, R. Morris, S. Williams, H. Markus. (2001). Evidence for cortical disconnection as a mechanism of age－related

cognitivedecline. *Neurology, 57*, 632－638.

M.I. Posner. (1980). Orienting of attention. *Q. J. Exp. Psychol, 32*, 3－25.

N. Picard, P.L. Strick. (1996). Motor areas of the medial wall: a review of their locationand functional activation. Cereb. *Cortex, 6,* 342－353.

N.U. Dosenbach, K.M. Visscher, E.D. Palmer, F.M. Miezin, K.K. Wenger, H.C.Kang, E.D. Burgund, A.L. Grimes, B.L. Schlaggar, S.E. Petersen. (2006). A core systemfor the implementation of task sets. *Neuron, 50*, 799－812.

P.L.S. Jacques, F. Dolcos, R. Cabeza. (2009). Effects of aging on functional connectivityof the amygdala for subsequent memory of negative pictures a networkanalysis of functional magnetic resonance imaging data. *Psychol. Sci. 20*, 74－84.

P.M. Greenwood, R. Parasuraman, J.V. Haxby. (1993). Changes in visuospatialattention over the adult lifespan. *Neuropsychologia, 31*, 471－485.

R. Cunnington, C. Windischberger, L. Deecke, E. Moser. (2002). The preparation andexecution of self－initiated and externally－triggered movement: a study ofevent－related fMRI. *Neuroimage, 15*, 373－385.

T. Wu, Y. Zang, L. Wang, X. Long, M. Hallett, Y. Chen, K. Li, P. Chan.(2007). Aginginfluence on functional connectivity of the motor network in the resting state. *Neurosci. Lett. 422*, 164－168.

V. Calhoun, M. Stevens, G. Pearlson, K. Kiehl. (2004). fMRI analysis with the generallinear model: removal of latency－induced amplitude bias by incorporation ofhemodynamic derivative terms. *Neuroimage, 22*, 252－257.

V. Della－Maggiore, W. Chau, P.R. Peres－Neto, A.R. McIntosh. (2002). An empiricalcomparison of SPM preprocessing parameters to the analysis of fMRI data. *Neuroimage, 17*, 19－28.

V. Menon, L.Q. Uddin. (2010). Saliency, switching,

attention and control: a networkmodel of insula function, Brain Struct. *Funct. 214*, 655 – 667.

W.W. Seeley, V. Menon, A.F. Schatzberg, J. Keller, G.H. Glover, H. Kenna, A.L.Reiss, M.D. Greicius. (2007). Dissociable intrinsic connectivity networks for salienceprocessing and executive control. *J. Neurosci. 27*, 2349 – 2356.

Dynamic neural network of insight：A functional magnetic resonance imaging study on solving Chinese 'chengyu' riddles

Qingbai Zhao[1,2], Zhijin Zhou[1,2,*], Haibo Xu[3,*], Shi Chen[1,2], Fang Xu[1,2], Wenliang Fan[3], Lei Han[4]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan, People's Republic of China)

([2] Key Laboratory of Human Development and Mental Health of Hubei Province, School of Psychology, Central China Normal University, Wuhan, People's Republic of China)

([3] MRI Center of Union Hospital, Huazhong University of Science and Technology, Wuhan, People's Republic of China)

([4] School of Psychology, Shandong Normal University, Jinan , People's Republic of China)

Abstract The key components of insight include breaking mental sets and forming the novel, task−related associations. The majority of researchers have agreed that the anterior cingulate cortex may mediate processes of breaking one's mental set, while the exact neural correlates of forming novel associations are still debatable. In the present study, we used a paradigm of answer selection to explore brain activations of insight by using event−related functional magnetic resonance imaging during solving Chinese 'chengyu' (in Chinese pinyin) riddles. Based on the participant's choice, the trials were classified into the insight and non−insight conditions. Both stimulus−locked and response−locked analyses are conducted to detect the neural activity corresponding to the early and late periods of insight solution, respectively. Our data indicate that the early period of insight solution shows more activation in the middle temporal gyrus, the middle frontal gyrus and the anterior cingulate cortex. These activities might be associated to the extensive semantic processing, as well as detecting and resolving cognitive conflicts. In contrast, the late period of insight solution produced increased activities in the hippocampus and the amygdala, possibly reflecting the forming of novel association and the concomitant "Aha" feeling. Our study supports the key role of hippocampus in forming novel associations, and indicates a dynamic neural network during insight solution.

Keywords insight; fMRI; Chinese 'chengyu' riddle; novel association; hippocampus; anterior cingulate cortex; neural network

1 Introduction

Since Köhler observed that chimpanzees could resolve problems suddenly rather than by an approach of trial and error , the processing of insight has attracted attention of many researchers. Following the early Gestalt psychologists, who thought that insightful problem−solving was based on a reconstruction to the whole problem, some cognitive psychologists proposed that it is due to the

*E−mail: zhouzj@mail.ccnu.edu.cn(zz); xuhaibo1120@hotmail.com (HX)

unsuitable representations of problem that led to the failure of effective problem solving in many situations. They further suggested that the representation change such as constraint relaxation and chunk decomposition should be the crucial process of insight. This theory was successful in interpreting visual−representation−based insightful problem solving, such as nine−dot problem.

With the development of neuroimaging technique, especially from 1990s onwards, the investigations on the neural correlates of insight flourished. However, since a number of homogenous mental events which can be repeatedly observed are required for the neuroimaging approach, the classic insightful paradigms such as nine−dot problem and six−matchstick problem are no longer suitable. Thus, a variety of insightful paradigms by using semantic representation have been applied, such as riddles and compound remote associates (CRA) problems. The generation−selection model has successfully interpreted the cognitive processing of these problems. According to this theory, the first step of insightful problem solving is to activate a variety of information and to form the temporary connection between them. The second step is to choose the task−related connection to solve the problem successfully. The key components of this model include processes to break the mental sets and to form novel, task−related associations among the old nodes of concepts or cognitive skills.

Luo and Niki used fMRI to record neural activities during insight by providing a trigger (the solution) to catalyze insightful solving processes to Japanese riddles. They found that the hippocampus played an important role in both breaking mental set and forming novel associations. Other studies demonstrated the availability of the paradigm to catalyze insight in studying breaking the mental sets, and showed that the anterior cingulate cortex (ACC) was the key brain region associated to this process.

However, providing a trigger made the problem solving more an apperception rather than an insight in the strict sense. Therefore, a learning−testing paradigm was introduced into Chinese logographs by Qiu et al. to explore the brain mechanism of insightful problem solving. The event−related potential (ERP) studies indicated that the ACC mediated the cognitive conflicts in both stages of mental preparatory and problem solving. And other fMRI results showed that the insight activated more in the precuneus, the left inferior/middle frontal gyrus, the inferior occipital gyrus and the cerebellum. These activities might be linked with successful prototype events retrieval, forming novel association and breaking mental sets, and re−arrangement of visual stimulus.

The CRA problems were also applied to investigate insight. In such a paradigm, insight and noninsight conditions were distinguished based on subsequent responses of participants' subjective feeling. Both fMRI and electroencephalogram (EEG) found an increased activity at the right anterior superior temporal gyrus during insight. This brain area was thought to be associated with making connections across distantly related information during comprehension. These studies supported the right−brain dominance theory of insight.

Besides, Aziz−Zadeh et al. used the anagrams to study insight, and found verbal insight solutions activated a distributed neural network, including the bilateral insula, the right prefrontal cortex and the ACC. However, in "Chinese anagrams", the fusiform and the right superior temporal gyri were found to be involved in insight solution.

As reviewed above, the majority of studies agreed that the ACC mediates processes of breaking one's mental set, while the exact neural correlates of forming novel associations are still debatable. In our opinion, the different tasks/experimental paradigms, as well as different definitions of timing insight event, may be the main cause leading to the disagreement.

In the learning−testing paradigm, the problem solving

should be similar to the analogy thinking in some sense. The learning of prototype events reduces the difficulty of problem solving so that participants can find a solution on their own initiative. However, the learning of prototype events also weakens the insight effect. At the same time, the forming of novel associations should occur at the end of insightful problem solving, while the insight events defined by Qiu were time-locked at the beginning of the presentations of target logographs. This stimulus-locked analysis would catch the process of breaking mental sets, but not forming novel associations. Therefore, it seems that the left lateral prefrontal cortex should not be the key brain region for forming novel associations.

In the CRA problems, the mean reaction time of producing insight solutions was about 10 seconds. Jung-Beeman et al. marked the insight events at the point about 2 seconds prior to each solution button pressing. This response-locked analyis could catch the process of forming associations. However, since the CRA problems are hybrid insight problems which can be solved through either insight or noninsight processes, the associations formed during CRAs might not be always novel. Moreover, as a well-known brain region linked with language ability, the right superior temporal gyrus was activated in processing the literal meaning of idioms. In that sense, the right superior temporal gyrus might be associated with the activation of extensive semantic network.

In the paradigm of providing a trigger to catalyze insight, the mean reaction time of insight solving was less than the repetition time of fMRI scanning (about 2 seconds). Thus, the definition of timing insight events (as the beginning of the presentations of the trigger) might catch both the breaking of mental sets and the forming of novel associations. However, the short reaction times led to the fact that the different periods of insight solving could not be separated by fMRI data analysis. Additionally, the activation of the hippocampus was not found in another

research by using similar paradigm, though it might be associated with forming novel associations due to its function in path reorientation and relational memory.

In the present study, we adopted Chinese 'chengyu' (in Chinese pinyin) riddles to explore the underlying neural mechanism of insight. 'Chengyu' is a type of traditional Chinese idiomatic expressions. Most of them consist of four characters. As most 'chengyu' are often intimately linked with the myth or historical event from which they were derived, their meaning usually surpasses the simple combination of the four characters. Different from the English idioms, which have the literal and figurative meanings, these 'chengyu' only have the figurative meaning and their literal meanings are unintelligible. It is worth to mention that there are some other 'chengyu' that are not born of a well-known fable. This type of 'chengyu' is just the shortened expression of its literal meaning, and they have no figurative meanings.

As each 'chengyu' only has one meaning, the meanings of its four component characters are constrained by the chunk of the 'chengyu'. This produces a mental set, which prevents the successful riddle solving because the riddles aim at an unconstrained meaning of the key character rather than the meaning of the 'chengyu' as a whole. To solve the riddle, the chunk of 'chengyu' must be decomposed, and extensive meanings of individual characters must be explored and retrieved. For example, the answer of the riddle 'shan zhan er duo mou' (善 战 而 多 谋 , means adept at fighting and planning) is the chengyu 'jing da xi suan' (精打细算 , means being very careful in reckoning). The key character in this riddle is 'da' (打 , one of its meanings is to hit), corresponding to 'zhan' (战 , with the meaning to fight). However, inside the 'chengyu', the 'da' (打) is bound with 'suan' (算). And the meaning of their combination 'da suan' (打 算) is to plan or to reckon. Obviously, the successful riddle solving is relied on the successful constraint relaxation to the key

character or the successful chunk decomposition. This is theoretically similar with the visual chunk decomposition of Chinese characters. Once extensive meanings of key character were retrieved, a number of temporary connections between the riddle and the 'chengyu' would be formed, and then the riddle would be solved by the selection process of task-related connections.

To explore the neural correlates of forming novel associations, another 'chengyu' that is similar in meaning to the riddle is used as the control condition (see in Methods). This 'chengyu' is normally associated with the riddle. For example, the 'chengyu' normally associated with the riddle 'shan zhan er duo mou' (善战而多谋, means adept at fighting and planning) is "zu zhi duo mou"（足智多谋, means being able and adept at planning). Thus, by directly manipulating the novelty of associations between the riddle and answer, we can accurately detect the neural correlates of forming novel associations.

Additionally, since different analyze methods, such as stimulus-locked or response-locked approaches, may be a factor resulting in the inconsistent findings in previous studies, both approaches were adopted in this study to detect the neural activity of early and late periods of insight solution respectively. Although fMRI is usually linked with poor temporal resolution, the timing of different periods of insight can be resolved by using the temporal analysis, in which the brain regions related to different periods have different peaks of the time course. Note that the separation of different periods needs a longer solution time, maybe double of repetition time of scanning or even more.

2　Methods

2.1　Participants

As paid volunteers, 20 undergraduates or graduates (13 women, 7 men), aged 21–35 years (mean age, 23.6 years) from Central China Normal University (CCNU),

participated in the experiment, and gave their informed consent according to the requirements of Institutional Review Board of CCNU. All participants were healthy, right-handed, and had normal or corrected to normal vision. Two participants were excluded from analysis due to their experiencing of less than 15% normal associations during the experiment. Another participant was excluded due to the excessive head motion during fMRI scanning.

2.2　Stimuli and task

A Chinese 'chengyu' riddle may be a phrase, or a saying, and its answer is a four-character 'chengyu'. Since there is a process of representation change when the participants tried to associate the riddle with the original answer, it is considered as the answer with novel association. In the current work, a control with normal association was produced in a pretest. A group of subjects were asked to report the four-character 'chengyu' that came to mind first when they saw the riddle in the pretest. Mostly, they could not find the novel answer and gave some different answers. The 'chengyu' with the highest frequency was chosen as the control. Thus, there are two answers, one of which is novel, and the other is normal. Taking the example used in the Introduction, the novel answer to the riddle 'shan zhan er duo mou' (善战而多谋, means adept at fighting and planning) is the 'chengyu' of 'jing da xi suan' (精打细算, means being very careful in reckoning), while its normal answer is 'zu zhi duo mou' (足智多谋, means being able and adept at planning).

In order to determine the difference between the answers with novel and normal associations, we had another group of subjects (totally 32) to rate their understanding of the Reasonability (matching with the answers to riddles) and Novelty on a scale of 1 to 5 for each of the 120 riddles. In the end, 84 riddles whose answers (both novel and normal ones) were evaluated as reasonable (mean scores > 3.5) were selected as the test riddles. Results showed that there was a significant difference in

novelty [paired t–test, t (83) =16.84, p<0.001] between the answers with novel (mean score =3.6) and normal association (mean score=2.6).

To familiarize the participants with the procedure and pace of this task, participants were trained with another set of 10 similar materials before they were put into the scanner. In the formal experiment, 84 test riddles were presented one by one with an event–related design. There was not any repetition of stimuli in the test. The Chinese characters, appearing in both the riddles and answers, had a font size of 28 (Song Ti font). The experimental paradigm was illustrated in Figure 1. The trial began with an 8–second black plus, a sign for rest, and a star sign for 1 second, followed by a warning of the presentation of riddle. After the riddle was displayed for 4 seconds, the novel association answer, normal association answer and two answers with no associations were presented. Participants were asked to select a novel and reasonable answer among these options within a limited 8–second period. Then was a 1–second blank followed by the next trial. The spatial positions were balanced among the different answers.

Since it is difficult to produce the novel answer on subjects' own initiative , while it seems an apperception if only providing the novel answer, the answer selection paradigm used here is an inevitable compromise. According to the selections of participants, the trials are classified into insight and noninsight solutions, respectively. Because including a process of representation change, the selection of novel answer is indeed an insight–based solution; however, as the selection of normal answer means the representation is not changed successfully, it can be considered as

noninsight–based solution. Although it might increase subjects' suppression, the simultaneous presentation of the novel and normal answers is still necessary in the answer selection paradigm. If only the novel or normal answer was presented with unrelated ones, subjects would get the target easily. That would be similar with the paradigm of providing the trigger. Additionally, subjects' suppression exists in both insight and noninsight solutions, and it can be eliminated by the comparison in the stage of data analysis.

2.3 fMRI acquirement

During MRI scanning, whole brain T2*–weighted echo planar imaging, based on blood oxygenation level–dependent contrast (EPI–BOLD) fMRI data, was acquired with a Siemens Trio 3.0–T MR–scanner using a standard head coil at the MRI Center of Wuhan Union Hospital. 32 interleaved slices, covering the entire brain, were acquired using a gradient–echo echo–planar pulse sequence. The slice thickness was 3.75 mm and the voxel size was 3 mm × 3 mm (TR=2s, TE =30 ms, FA=78°, FOV=192 × 192mm，Matrix size=64 × 64). Head motion was restricted with plastic braces and foam padding. The whole scanning sequence was divided into 2 runs, each consisted of 42 trials.

2.4 fMRI data analysis

The statistical parametric mapping (SPM5, http://www. fil.ion.ucl.ac.uk/spm/) was used for image preprocessing and voxel–based statistical analysis. Scans were first slice–time corrected, realigned, normalized (using the functional EPI template provided in SPM2), and smoothed (a Gaussian kernel with a full width at the full width at half maximum–FWHM of 8 mm). The resultant images had
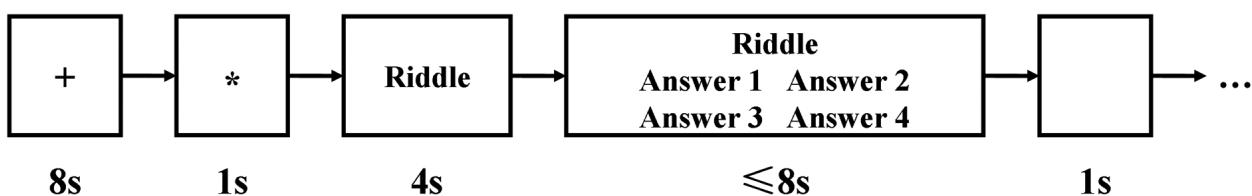
Fig. 1. The flow map of the formal experiment.
doi:10.1371/journal.pone.0059351.g001

cubic voxels of $2 \times 2 \times 2$ mm.

Two types of events, the insight solution (IS) and the noninsight solution (NS), were defined according to participants' selections to the answers. For each type, both stimulus-locked and response-locked analyses were conducted. In the stimulus-locked analysis, the event was defined as the presentation of the answers which was time-locked to the onset of the answer's presentation with zero duration. This event was considered as the early period activity of problem solving (EIS or ENS). Then, the time vector of the event was convoluted by the classic haemodynamic response function (HRF). Finally, by the general linear model, the activated brain regions associated with the EIS and ENS, as well as the differences between the two conditions, were obtained for each participant, and then combined in a random effect analysis to identify differences consistent across all participants. The thresholds were at p<0.001 (uncorrected for multiple comparisons) and 50 or more contiguous voxels (more than 10 original-sized voxels) for insight versus noninsight. The response-locked analysis was similar to the stimulus-locked analysis, except that the event was time-locked to 2 seconds prior to solution button pressing with zero duration. Similarly, this event was considered as the late period activity of problem solving (LIS or LNS).

3　Results

3.1　Behavioral performance

On average, in 61.7% of trials participants selected the answers with novel associations (average RT was 3.70s with a standard deviation of 0.78s), and in 26.3% of trials they selected the answers with normal associations (average RT was 4.06s with a standard deviation of 0.94s). There was a significant difference between the reaction times of the two conditions [paired t-test, t (16) = −5.63, p<0.001]. The larger trial percentage and short reaction time of

insight solutions might result from the instruction before the experiment that asked participants to select a novel and reasonable answer.

3.2　fMRI　results

Insight versus noninsight in the early period of solution. This contrast (EIS−ENS) examined brain areas that were more activated in the early period of insight solutions relative to the noninsight solutions. Several peaks of activation were found, including the left superior temporal pole, inferior temporal gyrus, anterior cingulate cortex and middle frontal gyrus, the right inferior frontal gyrus and parahippocampal gyrus, and the bilateral angular gyrus and middle temporal gyrus (see in Table Ⅰ). There were no voxels that were significantly more active for the early processing of noninsight solutions than the insight solutions.

Insight versus noninsight in the late period of solution. This contrast (LIS−LNS) examined brain areas that were more activated in the late period of insight solutions relative to the noninsight solutions. Several peaks of activation were found, including the left olfactory, middle frontal gyrus, anterior cingulate cortex, medial frontal gyrus and inferior parietal gyrus, the right putamen and amygdala, and the bilateral middle temporal gyrus, hippocampus and angular gyrus (see in Table Ⅰ). There were no voxels that were significantly more active for the late processing of noninsight solutions than the insight solutions.

Timecourse activity of regions associated with insight. To clarify the differences between the early and late period activity of insight, the timecourse activity of the brain regions associated with insight was visualized. According to the HRF by SPM, the haemodynamic response to the EIS would delay for six seconds from the onset of the answer presentation, and that to the LIS would delay for ten seconds (the delay time plus mean solution time). As shown in Figure 2, the haemodynamic response of the left and right middle temporal gyri, left anterior cingulate cortex and middle frontal gyrus started at the

Table 1. Brain areas more activated in insight than in noninsight solution

Area	BA	Voxels	x	y	z	T	Z
Insight > Noninsight in the early period of solution							
Left superior temporal pole	38	752	−28	10	−28	6.31	4.41
Left inferior temporal gyrus	21		−42	0	−38	5.95	4.26
Left anterior cingulate cortex	32	58	−10	42	8	6.06	4.31
Left middle frontal gyrus	8	148	−44	20	48	5.96	4.27
Right angular gyrus	40	326	50	−56	36	5.81	4.20
Right middle temporal gyrus	21	126	54	2	−30	5.50	4.07
Left angular gyrus	39	1070	−50	−56	36	5.81	4.20
Left middle temporal gyrus	21		−52	−30	−2	5.04	3.84
Right middle temporal gyrus	21	248	60	−48	8	4.94	3.79
Right inferior frontal gyrus	47	81	24	14	−20	4.74	3.69
Right parahippocampal gyrus	28		24	4	−32	4.49	3.56
Insight > Noninsight in the late period of solution							
Right middle temporal gyrus	21	834	68	−44	−6	7.46	4.83
Right middle temporal gyrus	37		56	−56	8	5.52	4.07
Left olfactory	34	1720	−24	6	−18	7.02	4.68
Left middle temporal gyrus	20		−54	−8	−26	6.73	4.57
Left hippocampus	20		−26	−10	−20	6.20	4.36
Right angular gyrus	39	158	46	−56	36	6.88	4.63
Right putamen	48	349	22	6	−12	6.50	4.48
Right amygdala	34		28	−4	−10	5.96	4.27
Right hippocampus	20		30	−6	−20	4.37	3.49
Left middle frontal gyrus	46	124	−38	22	42	6.38	4.44
Left anterior cingulate cortex	10	1032	−4	52	2	5.83	4.21
Left medial frontal gyrus	10		−8	54	18	5.79	4.19
Left middle temporal gyrus	21	333	−52	−32	−2	5.49	4.06
Left angular gyrus	39	642	−52	−64	40	4.80	3.73
Left inferior parietal gyrus	40		−60	−48	44	4.78	3.71

BA, Brodmann area. Coordinates (x,y,z) were the MNI (Montreal Neurological Institute) coordinates. The thresholds were set at p<0.001 (uncorrected for multiple comparisons) and 50 or more contiguous voxels. T− and Z−scores of the activations were also shown.

beginning of the answers presentation and continued until the riddle was solved (specifically, the timecourse of left middle temporal gyrus has a bi−peak); while that of the left hippocampus and right amygdala started just before the riddle was solved (about four seconds after the onset of answer presentation), with a peak at ten seconds. These were consistent with the statistical results above.

4 Discussion

According to the generation−selection model, the insightful riddle solving meant to activate a variety of information covering the riddle and answers, to decompose the chunk of target 'chengyu', and then to form the novel, task−related associations. Accordingly, the process of riddle solving was divided into the early and late periods of solution, which can be separated by the stimulus− locked and response−locked anslyses of fMRI data. In the early period, the insight activated more in broad lateral temporal areas, left middle frontal gyrus and ACC; while in the late period, a broader network of brain areas was activated, including the lateral temporal areas, left middle frontal gyrus, ACC, hippocampus, amygdala. These areas suggested a dynamic neural network of insightful problem solving. Their roles were discussed below.

4.1 Middle temporal gyrus in extensive semantic processing

The bilateral middle temporal gyri were activated in
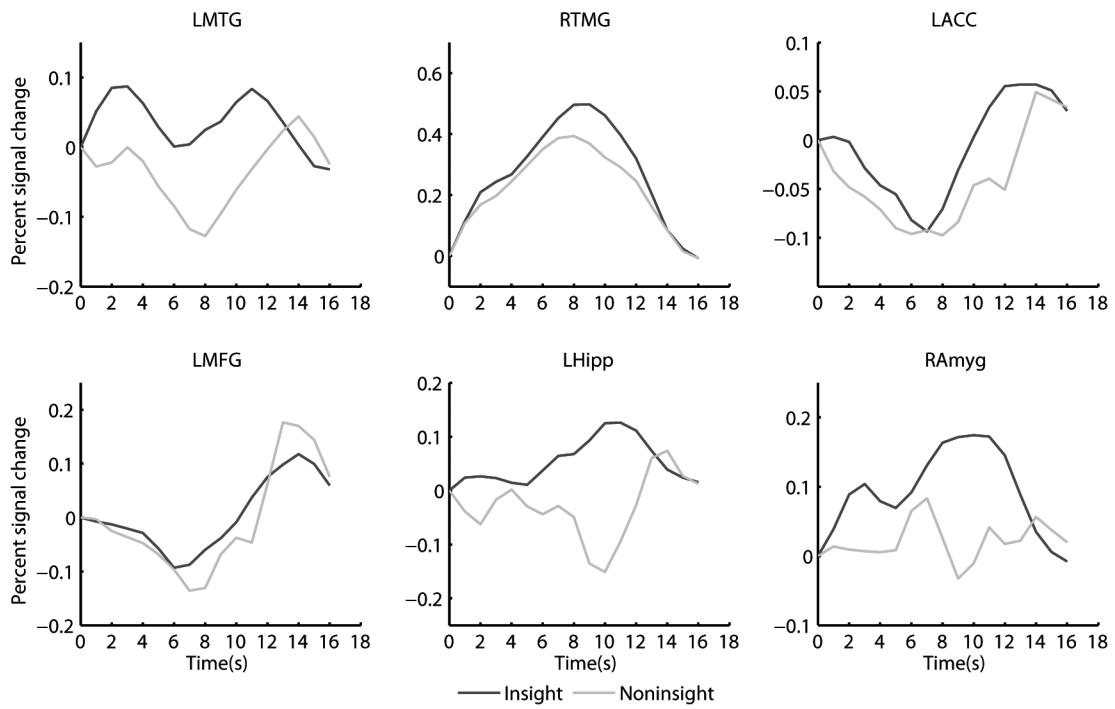
Fig. 2. Average signal change of the interested brain regions, following the onset (time=0) of answer presentation for insight solution (red line) and noninsight solution (blue line). LTMG means left middle temporal gyrus with the peak at (−52,−30,−2) in MNI space, RTMG means right middle temporal gyrus with the peak at (60,−48,8), LACC means left anterior cingulate cortex with the peak at (60,−48,8), LMFG means left middle frontal gyrus with the peak at (−44,20,48), Lhipp means left hippocampus with the peak at (−26,−10,−20) and Ramyg means right amygdala with the peak at (28,−4,−10). Note that the timecourse data were linearly interpolated from the original 2−s resolution down to 1−s resolution to obtain the smoothed curves.

both early and late periods of insight solution relative to the noninsight solution. These brain areas are well−known in semantic processing.

In the present study, the normal association was formed just when the whole meaning of the normal answer was understood, while the forming of novel association needed to decompose the target 'chengyu' and to activate the extensive (normal and novel) meanings of the key character. It was found that the processing of the novel meaning of idioms was associated with increased activity in the right middle and superior temporal gyri. It was proposed that the right superior temporal gyrus is the key brain area in forming connections across distantly related information. The right superior temporal gyrus was not emphasized in other insight studies, probably due to its well−established function in language processing. Additionally, it was found by others that the bilateral superior temporal gyrus , rather than only right side one,

were activated in insight problem solving. In the present study, it was the bilateral middle temporal gyrus rather than the right superior temporal gyrus that was more activated in insight solution. Consistent with some others' work, the current results provide no support to the right−brain dominance theory in insight.

Although the insightful problem solving comes suddenly, it still needs activations of extensive and detailed semantic information in order to form novel associations. Thus, the increased activity of temporal areas might be associated with the activation of extensive semantic network rather than the novel associations. This was supported by the fact that there was no significant difference in the activation of middle temporal gyrus between the early and late insight effect.

4.2 ACC and middle frontal gyrus in breaking mental sets

The left middle frontal gyrus and ACC were activated

in both early and late insight solution relative to the noninsight solution.

Insight means to break the mental sets and to form novel, task-related associations. To break a mental set, subjects need to detect and resolve cognitive conflict. Studies indicated that the anterior cingulate cortex (ACC) was more activated in insight problem solving. Due to its well-known role in cognitive conflict monitoring, the ACC was highlighted and proposed to mediate processes of breaking one's mental set. In addition, it was found that increased activity in ACC, prior to problem presentation, might reflect increased readiness for monitoring of competing responses and for applying of cognitive control mechanisms. The increased ACC activity could also predict subsequent successful solution in term of a sudden insight.

After detecting the cognitive conflicts, one should shift its mental sets to solve the conflicts in insight. It was found that the lateral prefrontal cortex, including the inferior and middle frontal gyri, was activated in chunk decomposition of Chinese characters, set-shift problems and insightful riddle solving. Due to its role in establishing and shifting the attentional sets, the lateral prefrontal cortex was thought to be associated with conflict resolution.

Note that the ACC and middle frontal gyrus were deactivated relative to the resting stage. This was different from some findings, but consistent with others, in which the activity of the ACC reached the peak during the prepared period but decreased in the middle of the task. It was a reasonable result since the ACC is one part of the default mode network which shows greater activity during resting state than during cognitive tasks.

4.3 Hippocampus in forming novel associations

The bilateral hippocampi were activated in the late (but not the early) period of insight solution compared with the noninsight solution.

In the first neuroimaging study on insight, the right hippocampus was found activated in the insight event. Due to its function in path reorientation and relational memory, the hippocampus was proposed to play an important role in both of breaking mental set and forming novel associations. As discussed above, the ACC and lateral prefrontal cortex should be the key brain area responsible for processes of breaking one's mental set. Thus, the hippocampus might be mainly associated with the forming of novel associations. This was supported by two facts. First, an increased activity of hippocampus was detected during encoding of novel visual stimuli, and hippocampal damage altered the ERPs in response to novel stimuli. Second, in task to judge the relationship of semantically related or unrelated word pairs, the hippocampus was responsive to the retrieval of semantic associations.

Our result supported the role of hippocampus in forming novel associations. Firstly, we manipulated the novelty of association between the riddle and answer, and found an increased activity of hippocampus in the selections to the novel answers. Secondly, the hippocampus showed more activation than the noninsight solution only in the late period of insight solution. It was the time of forming novel associations.

4.4 Amygdala in "Aha" feeling

The amygdala was activated in the late period of insight solution relative to the noninsight solution.

Studies demonstrated that the amygdala was associated with emotional learning and expression. Although the majority of studies reported the amygdala was sensitive to negative stimuli, recent studies revealed that it was also involved in positive emotions.

Both the cognitive processing of forming novel associations and the corresponding "Aha" feeling should be included in the current paradigm of insight solution. Due to its role in emotion, the amygdala was proposed to reflect the subjective experience during insight. As it is approaching to the completion of insightful

solution, the insight feeling is strengthened, and the activity of amygdala is also increased.

5 Conclusion

Our data demonstrated that the middle temporal gyrus, the middle frontal gyrus, the ACC, the hippocampus and the amygdala, are more activated during the insight solution than during the noninsight solution. These brain areas consist of a dynamic neural network, associated with the cognitive and affective functions during insight of Chinese 'chengyu' riddles solving. The increased activity of middle temporal gyrus may reflect the activation of extensive semantic network; the activation of lateral prefrontal cortex and ACC may be responsible for detecting and resolving cognitive conflict; the hippocampus seems to be the key brain areas in forming novel associations; and the activation of amygdala might reflect the subjective experience during insight. Although the superior temporal gyrus was proposed previously to be the key brain area in forming connections across distantly related information, this region is not found activated during insight in the present study, and the right−brain dominance theory on insight is thus not supported by our results.

References

Ahrens, K., Liu, H. L., Lee, C. Y., Gong, S. P., Fang, S. Y., et al. (2007). Functional MRI of conventional and anomalous metaphors in Mandarin Chinese. *Brain Lang, 100,* 163−171.

Aziz−Zadeh, L., Kaplan, J. T., & Iacoboni, M. (2009) "Aha!": The neural correlates of verbal insight solutions. *Hum Brain Mapp, 30,* 908−916.

Balleine, B. W., & Killcross, S. (2006) Parallel incentive processing: an integrated view of amygdala function. *Trends Neurosci, 29,* 272−279.

Barch, D. M., Braver, T. S., Sabb, F. W., & Noll, D. C. (2000). Anterior cingulate and the monitoriing of response conflict: evidence from an fMRI study of overt verb generation. *J Cogn Neurosci, 12,* 298−309.

Bink, M. L., & Marsh, L. M. (2000). Cognitive Regularities in Creative Activity. *Review of General Psychology, 4,* 59−78.

Bowden, E. M., & Jung−Beeman, M. (2003). Aha! Insight experience correlates with solution activation in the right hemisphere. *Psychon Bull Rev, 10,* 730−737.

Bowden, E. M., & Jung−Beeman, M. (2007). Methods for investigating the neural components of insight. *Methods, 42,* 87−99.

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neurosci Biobehav Rev, 26,* 321−352.

Cohen, N. J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., et al. (1999). Hippocampal system and declarative (relational) memory: summarizing the data from functional neuroimaging studies. *Hippocampus, 9*, 83−98.

Cranford, E. A., & Moss, J. (2012) Is Insight Always the Same? A Protocol Analysis of Insight in Compound Remote Associate Problems. *The Journal of Problem Solving,* 4, 128−153.

Dietrich, & A., Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychol Bull, 136*, 822−848.

Fanselow, M. S., & Gale, G. D. (2003). The amygdala, fear, and memory. *Ann N Y Acad Sci, 985,* 125−134.

Giora, R. (1997). Understanding figurative and literal language: the graded salience hypothesis. *Cognitive Linguistics, 7*, 183−206.

Goel, V., & Vartanian, O. (2005) Dissociating the roles of right ventral lateral and dorsal lateral prefrontal cortex in generation and maintenance of hypotheses in set−

shift problems. *Cereb Cortex, 15,* 1170–1177.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proc Natl Acad Sci USA, 100*, 253–258.

Huang, F. R., Zhou, Z. J., & Zhao, Q. B. (2013). An eye movement study of associate competition in Chinese idiom riddles solving. *Acta Psychologica Sinica, 45,* 36–46.

Jamal, N. I,, Piche, A. W., Napoliello, E. M., Perfetti, C. A., & Eden, G. F. (2012). Neural basis of single–word reading in Spanish–English bilinguals. *Hum Brain Mapp, 33*, 235–245.

Johnson, J. D., Muftuler, L. T., & Rugg, M. D. (2008). Multiple repetitions reveal functionally and anatomically distinct patterns of hippocampal activity during continuous recognition memory. *Hippocampus, 18*, 975–980.

Jung–Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel–Liu, S., et al. (2004) Neural activity when people solve verbal problems with insight. *PLoS Biol, 2*, E97.

Kershaw, T. C., & Ohlsson, S. (2004). Multiple causes of difficulty in insight: the case of the nine–dot problem. *J Exp Psychol Learn Mem Cogn, 30,* 3–13.

Knight, R. (1996). Contribution of human hippocampal region to novelty detection. *Nature, 383,* 256–259.

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Mem Cognit, 29,* 1000–1009.

Kounios, J., Fleck, J. I., Green, D. L., Payne, L., Stevenson, J. L., et al. (2008). The origins of insight in resting–state brain activity. *Neuropsychologia, 46,* 281–291.

Kounios, J., Frymiare, J. L., Bowden, E. M., Fleck, J. I., Subramaniam, K., et al. (2006), The prepared mind: neural activity prior to problem presentation predicts subsequent solution by sudden insight. *Psychol Sci, 17,* 882–890.

Köhler, W. (1925). *The mentality of apes*. London: Routledge & Kegan Paul.

Luks. T. L., Simpson, G. V., Feiwell, R. J., & Miller, W. L. (2002), Evidence for anterior cingulate cortex involvement in monitoring preparatory attentional set. *Neuroimage, 17,* 792–802.

Luo, J. (2004). Neural correlates of insight. *Acta Psychologica Sinica, 36*, 219–234.

Luo, J., & Niki, K. (2002). Role of medial temporal lobe in extensive retrieval of task–related knowledge. *Hippocampus, 12*, 487–494.

Luo, J., & Niki, K. (2003). Function of hippocampus in "insight" of problem solving. *Hippocampus, 13,* 316–323.

Luo, J., Niki, K., & Knoblich, G. (2006). Perceptual contributions to problem solving: Chunk decomposition of Chinese characters. *Brain Res Bull, 70,* 430–443.

Luo, J., Niki, K., & Phillips, S. (2004). Neural correlates of the 'Aha! reaction'. *Neuroreport, 15,* 2013–2017.

MacDonald, A. W., 3rd, Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288*, 1835–1838.

Mai, X. Q., Luo, J., Wu, J. H., & Luo, Y. J. (2004). "Aha!" effects in a guessing riddle task: an event–related potential study. *Hum Brain Mapp, 22*, 261–270.

Mashal, N., Faust, M., Hendler, T., & Jung–Beeman, M. (2008). Hemispheric differences in processing the literal interpretation of idioms: converging evidence from behavioral and fMRI studies. *Cortex, 44,* 848–860.

Paton, J. J., Belova, M. A., Morrison, S. E., & Salzman, C. D. (2006). The primate amygdala represents the positive and negative value of visual stimuli during learning. *Nature, 439,* 865–870.

Qiu, J., Li, H., Jou, J., Liu, J., Luo, Y., et al. (2010) Neural

correlates of the "Aha" experiences: evidence from an fMRI study of insight problem solving. *Cortex, 46,* 397–403.

Qiu, J., Li, H., Jou, J., Wu, Z., & Zhang, Q. (2008). Spatiotemporal cortical activation underlies mental preparation for successful riddle solving: an event–related potential study. *Exp Brain Res, 186*, 629–634.

Qiu, J., Li, H., Luo, Y., Chen, A., Zhang, F., et al. (2006). Brain mechanism of cognitive conflict in a guessing Chinese logogriph task. *Neuroreport, 17,* 679–682.

Qiu, J., Li, H., Yang, D., Luo, Y., Li, Y., et al. (2008) The neural basis of insight problem solving: an event–related potential study. *Brain Cogn, 68,* 100–106.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., et al. (2001). A default mode of brain function. *Proc Natl Acad Sci USA, 98*, 676–682.

Raichle, M. E., & Snyder, A. Z. (2007). A default mode of brain function: a brief history of an evolving idea. *Neuroimage, 37,* 1083–1090; discussion 1097–1089.

Redish, A. D. (2001). The hippocampal debate: are we asking the right questions? *Behav Brain Res, 127,* 81–98.

Sandkuhler, S., & Bhattacharya, J. (2008) . Deconstructing insight: EEG correlates of insightful problem solving. *PLoS One, 3*, e1459.

Vandenberghe, R., Price, C., Wise, R., Josephs, O., & Frackowiak, R. S. (1996). Functional anatomy of a common semantic system for words and pictures. *Nature, 383*, 254–256.

Van Veen, V., Cohen, J. D., Botvinick, M. M., Stenger , V. A., & Carter, C. S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. *Neuroimage, 14,* 1302–1308.

Weisberg, R. W. (1996). P*rolegomena to Theories of Insight in Problem sovling: A Taxonomy of Problem*s. Chapt5 in "The Nature of Insight" by Robert J Sternberg and Janet E Davidson, MIT Press.

Zhang, M., Tian, F., Wu, X., Liao, S., & Qiu, J. (2011). The neural correlates of insight in Chinese verbal problems: an event related–potential study. *Brain Res Bull, 84*, 210–214.

Neural pathway in the right hemisphere underlies verbal insight problem solving

Qingbai Zhao[1,2], Zhijin Zhou[1,2,*], Haibo Xu[3,*], Wenliang Fan[3], Lei Han[4]

([1] Key Laboratory of Adolescent CyberPsychology and Behavior (CCNU), Ministry of Education, Central China Normal University,Wuhan 430079, China)

([2] Key Laboratory of Human Development and Mental Health of Hubei Province, School of Psychology, Central China Normal University, Wuhan 430079, China)

([3] MRI Center of Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China)

([4] School of Psychology, Shandong Normal University, Jinan 250014, China)

Abstract Verbal insight problem solving means to break mental sets, to select the novel semantic information and to form novel, task–related associations. Although previous studies have identified the brain regions associated with these key processes, the interaction among these regions during insight is still unclear. In the present study, we explored the functional connectivity between the key regions during solving Chinese 'chengyu' riddles by using event–related functional magnetic resonance imaging. Results showed that both insight and noninsight solutions activated the bilateral inferior frontal gyri, middle temporal gyri and hippocampi, and these regions constituted a frontal to temporal to hippocampal neural pathway. Compared with noninsight solution, insight solution had a stronger functional connectivity between the inferior frontal gyrus and middle temporal gyrus in the right hemisphere. Our study reveals the neural pathway of information processing during verbal insight problem solving, and supports the right–hemisphere advantage theory of insight.

Keywords insight; fMRI; functional connectivity; Chinese 'chengyu' riddle; right hemisphere

1 Introduction

Since Köhler observed (1925) that chimpanzees could resolve problems suddenly rather than by an approach of trial and error, the processing of insight has attracted attention of many researchers. Since unsuitable representations of problem would lead to the failure of effective problem solving in many situations, some cognitive psychologists proposed that the representation change such as constraint relaxation and chunk decomposition should be the crucial process of insight (Kaplan and Simon, 1990, Knoblich et al., 2001, Ormerod et al., 2002). Using some visual−representation−based problems such as nine−dot problem and Chinese chunk decomposition problem, researchers found that there were multiple sources of difficulty of particular insight problems, and that early perceptual processes could crucially affect thinking and problem solving (MacGregor et al., 2001, Kershaw and Ohlsson, 2004, Luo et al., 2006, Wu et al., 2013).

With the development of neuroimaging technique, especially from 1990s onwards, the investigations on the neural correlates of insight flourished (Dietrich and Kanso, 2010). However, since a number of homogenous mental events which can be repeatedly observed are required for the neuroimaging approach, the classic insightful paradigms such as nine−dot problem and six−matchstick problem are no longer suitable (Luo, 2004). Thus, a variety of verbal problems have been applied in the studies of insight, such as riddles, logographs and compound remote associates problems (Luo and Niki, 2003, Jung−Beeman et al., 2004, Qiu et al., 2010, Zhao et al., 2013). In these studies, researchers focused on two key components of insight processing, that is to break the mental sets and to form novel, task−related associations among the old nodes of concepts or cognitive skills (Luo and Niki, 2003,

Bowden and Jung−Beeman, 2007).

Studies indicated that the frontal cortex played a key role in breaking mental sets. Therein, the anterior cingulate cortex was highlighted and proposed to monitor the cognitive conflicts resulted from mental sets (Luo et al., 2004, Mai et al., 2004, Qiu et al., 2008, Aziz−Zadeh et al., 2009). After detecting the cognitive conflicts, one should break its mental sets to solve the conflicts in insight. It was found that the lateral prefrontal cortex, including the inferior and middle frontal gyri, was activated in chunk decomposition of Chinese characters (Luo et al., 2006), set−shift problems (Goel and Vartanian, 2005) and insightful riddle solving (Luo and Niki, 2003, Luo et al., 2004, Qiu et al., 2010). Due to its role in establishing and shifting the attentional sets (MacDonald et al., 2000, Luks et al., 2002), the lateral prefrontal cortex was thought to be associated with conflict resolution.

The breaking of mental sets would result in the retrieval of new information pieces. Then, the selection of novel information pieces and forming novel, task−related associations were the keys of insight problem solving (Bink and Marsh, 2000). According to the coarse semantic coding theory, the right−hemisphere engages in coarse semantic coding, weakly and diffusely activating alternative meanings and more distant associates (Faust and Chiarello, 1998, Beeman and Bowden, 2000, Bowden and Jung−Beeman, 2003). Therefore, some researchers highlighted the role of the right hemisphere (especially the right anterior superior temporal gyrus) in making connections across distantly related information during insight (Bowden and Jung−Beeman, 2003, Jung−Beeman et al., 2004). In fact, the right temporal cortex might be mainly in charge of processing novel semantic information (Faust and Mashal, 2007, Mashal et al., 2008, Pobric et al., 2008), while the hippocampus should be the key brain region in forming the novel associations (Luo and Niki, 2003, Zhao et al., 2013), due to its function in path reorientation (Redish, 2001),

relational memory (Cohen et al., 1999, Luo and Niki, 2002) and response to novel stimuli (Knight, 1996, Johnson et al., 2008).

Obviously, verbal insight problem solving activates a distributed neural network including the anterior cingulate cortex, lateral prefrontal cortex, right temporal areas and hippocampus. Although previous studies have identified the roles of these brain regions, the information integrations among them are still unclear. The electroencephalograph study showed that good performance in the divergent thinking task was related to increased functional connectivity of central-parietal areas of both hemispheres and greater ipsilateral connections between the cortex regions of the right hemisphere in the beta2 band (Razoumnikova, 2000). And the study using diffusion tensor imaging reported significant positive relationships between individual creativity as measured by the divergent thinking test and fractional anisotropy in the white matter in or adjacent to the bilateral prefrontal cortices, the body of the corpus callosum, the bilateral basal ganglia, the bilateral temporal-parietal junction and the right inferior parietal lobule (Takeuchi et al., 2010). These studies indicated the importance of the information integration of different brain regions in creativity.

In verbal insight problem solving, the forming of novel associations is dependent on the selection of novel semantic information. Since the lateral prefrontal cortex, temporal areas and hippocampus are respectively associated with conflict resolution, semantic processing and relational memory, it is speculated that the functional connectivity between the lateral prefrontal cortex and temporal areas might reflect the information selection process in insight and that between right temporal areas and hippocampus might underlie the forming of novel association. The current work aims to reveal the functional connectivity among the key brain regions which underlies the cognitive processing in insight.

Additionally, according to the coarse semantic coding theory, the right temporal cortex should play a crucial role in verbal insight problem solving, and this is supported by several studies (Bowden and Jung-Beeman, 2003, Jung-Beeman et al., 2004, Zhang et al., 2011, Zhou et al., 2011). However, there are also some studies revealing bilateral activation patterns associated with insight events (Luo and Niki, 2003, Aziz-Zadeh et al., 2009, Zhao et al., 2013). There might be several reasons why the latter studies do not find the right hemisphere advantage. First, verbal insight problems can not be solved by the conventional semantic information processing, and then the process of retrieving the novel semantic information is the key of insight solving. However, some of the latter studies adopted the paradigm of providing triggers to catalyze the insight processes. This would simplify the retrieval of the novel meanings and distant associates, and then weaken the activation of the right temporal cortex. Second, since insight solution comes to mind suddenly, the right temporal cortex should show greater activation at the time just prior to the solution (Jung-Beeman et al., 2004). However, most of the latter studies focused on the activation throughout the solving period, not exactly catching the key period of the activation in right temporal cortex. It is noticed that all these findings, no matter supporting the right hemisphere theory or not, are from the location analysis of brain functions. Thus, as one of the two patterns of brain functional organization (Tononi et al., 1994), the functional integration analysis may provide something new for the discussions on hemisphere difference in insight.

2 Experimental procedures

2.1 Participants

As paid volunteers, 20 undergraduates or graduates (13 women, 7 men), aged 21 - 35 years (mean age, 23.6 years) from Central China Normal University, participated in the

experiment, and gave their informed consent according to the requirements of Institutional Review Board of Central China Normal University. All participants were healthy, right−handed, and had normal or corrected to normal vision. Two participants were excluded from analysis due to their experiencing of less than 15% normal associations during the experiment. Another participant was excluded due to the excessive head motion during functional magnetic resonance imaging (fMRI) scanning.

2.2 Stimuli and task

In the present study, we adopted the Chinese 'chengyu' (in Chinese pinyin) riddles to explore the underlying neural mechanism of insight. A Chinese 'chengyu' riddle may be a phrase, or a saying, and its answer is a four−character 'chengyu' which is a type of traditional Chinese idiomatic expressions. As each 'chengyu' only has one meaning, the meanings of its four component characters are constrained by the chunk of the 'chengyu'. This prevents the successful riddle solving because the riddles aim at an unconstrained meaning of the key character rather than the meaning of the 'chengyu' as a whole. To solve the riddle, the chunk of 'chengyu' must be decomposed, and extensive meanings of individual characters must be explored and retrieved. For example, the answer of the riddle 'shan zhan er duo mou' (善战而多谋 , means adept at fighting and planning) is the chengyu 'jing da xi suan' (精打细算 , means being very careful in reckoning). The key character in this riddle is 'da' (打 , one of its meanings is to hit), corresponding to 'zhan' (战 , with the meaning to fight). However, inside the 'chengyu', the 'da' (打) is bound with 'suan' (算). And the meaning of their combination 'da suan' (打算) is to plan or to reckon. Obviously, the successful riddle solving is relied on the successful constraint relaxation to the key character or the successful chunk decomposition. This is theoretically similar with the visual chunk decomposition of Chinese characters (Luo et al., 2006). Once extensive meanings of key character were retrieved,

a number of temporary connections between the riddle and the 'chengyu' would be formed, and then the riddle would be solved by the selection process of task−related connections. Since there is a process of representation change when the participants tried to associate the riddle with the original answer, it is considered as the answer with novel association.

In the current work, a control with normal association was produced in a pretest. A group of subjects were asked to report the four−character 'chengyu' that came to mind first when they saw the riddle in the pretest. Mostly, they could not find the novel answer and gave some different answers. The 'chengyu' with the highest frequency was chosen as the control. Thus, there are two answers, one of which is novel, and the other is normal. For example, the novel answer to the riddle 'shan zhan er duo mou' (善战而多谋 , means adept at fighting and planning) is the 'chengyu' of 'jing da xi suan' (精打细算 , means being very careful in reckoning), while its normal answer is 'zu zhi duo mou' (足智多谋 , means being able and adept at planning).

In order to determine the difference between the answers with novel and normal associations, we had another group of subjects (totally 32) to rate their understanding of the Reasonability (matching with the answers to riddles) and Novelty on a scale of 1 to 5 for each of the 120 riddles. In the end, 84 riddles whose answers (both novel and normal ones) were evaluated as reasonable (mean scores > 3.5) were selected as the test riddles. Results showed that there was a significant difference in novelty [paired t−test, t (83) =16.84, p<0.001] between the answers with novel (mean score =3.6) and normal association (mean score=2.6).

To familiarize the participants with the procedure and pace of this task, participants were trained with another set of 10 similar materials before they were put into the scanner. In the formal experiment, 84 test riddles were

Fig. 1. The flow map of the formal experiment

presented one by one with an event−related design. There was not any repetition of stimuli in the test. The Chinese characters, appearing in both the riddles and answers, had a font size of 28 (Song Ti font). The experimental paradigm was illustrated in Fig.1 The trial began with an 8−s black plus, a sign for rest, and a star sign for 1s, followed by a warning of the presentation of riddle. After the riddle was displayed for 4s, the novel association answer, normal association answer and two answers with no associations were presented. Participants were asked to select a novel and reasonable answer among these options within a limited 8−s period. Then was a 1−s blank followed by the next trial. The spatial positions were balanced among the different answers.

Since it is difficult to solve the riddle on participants' own initiative, the paradigm of providing a trigger to catalyze insight processes was adopted by some studies (Luo and Niki, 2003, Luo et al., 2004, Mai et al., 2004). This paradigm is helpful in investigating the processes of breaking mental sets and forming novel associations, but of little effect in exploring how participants retrieve the extensive information to solve problems. Therefore, the present study introduced the answer selection paradigm described above to investigate the information selection in insight. Participants were asked to select the novel and reasonable answer from four options, in which the novelty was more emphasized. According to the selections of participants, the trials are classified into insight and noninsight solutions, respectively. Since the normal answer points to the conventional thinking, it is easy to understand and should be found first. However, the normal answer is of less novelty, and then participants might actively look for the novel one from the others. If participants ultimately

could not find the association between the novel answer and the riddle, they might select the normal one, in which case the problem solving is a simple process in the conventional thinking, and it is considered as noninsight−based solution. Once participants found the association between the novel answer and the riddle, they would select it as asked. In this case, there is a competition between the novel and normal answer, and the selection of novel answer just reflects the breaking of conventional thinking. Additionally, because of including a process of representation change, the selection of novel answer is indeed an insight−based solution.

Indeed, although the answer selection makes participants more active than solution recognition, it is different from the actual problem solving on participants' own. The adoption of the answer selection paradigm is an inevitable compromise.

2.3 fMRI acquirement

During MRI scanning, whole brain T2*−weighted echo planar imaging, based on blood oxygenation level−dependent contrast (EPI−BOLD) fMRI data, was acquired with a Siemens Trio 3.0−T MR−scanner using a standard head coil at the MRI Center of Wuhan Union Hospital. 32 interleaved slices, covering the entire brain, were acquired using a gradient−echo echo−planar pulse sequence. The slice thickness was 3.75 mm and the voxel size was 3 mm × 3 mm (TR = 2s, TE = 30 ms, FA=78°，FOV=192 × 192mm，Matrix size = 64 × 64). Head motion was restricted with plastic braces and foam padding. The whole scanning sequence was divided into two runs, each consisted of 42 trials.

2.4 fMRI data analysis

Preprocessing. The statistical parametric mapping

(SPM5, http://www.fil.ion.ucl.ac. uk/spm/) was used for image preprocessing and voxel–based statistical analysis. Scans were first slice–time corrected, realigned, normalized (using the functional EPI template provided in SPM5), and smoothed (a Gaussian kernel with a full width at the full width at half maximum – FWHM of 8 mm). The resultant images had cubic voxels of $3 \times 3 \times 3$ mm.

Mapping brain activation. Two types of events, the insight solution and the noninsight solution, were defined according to participants' selections to the answers. Since the participants could not solve the riddle by themselves in the initial 4s of riddle presentation (Zhu et al., 2009), the event was defined as the answer selection process for each type which began at the onset of the answers' presentation and ended in the participants' pressing. Then, the time vector of the event was convoluted by the classic haemodynamic response function. Finally, by the general linear model, the activated brain regions associated with the insight solution and the noninsight solution, as well as the differences between the two conditions, were obtained for each participant, and then combined in a random effect analysis to identify differences consistent
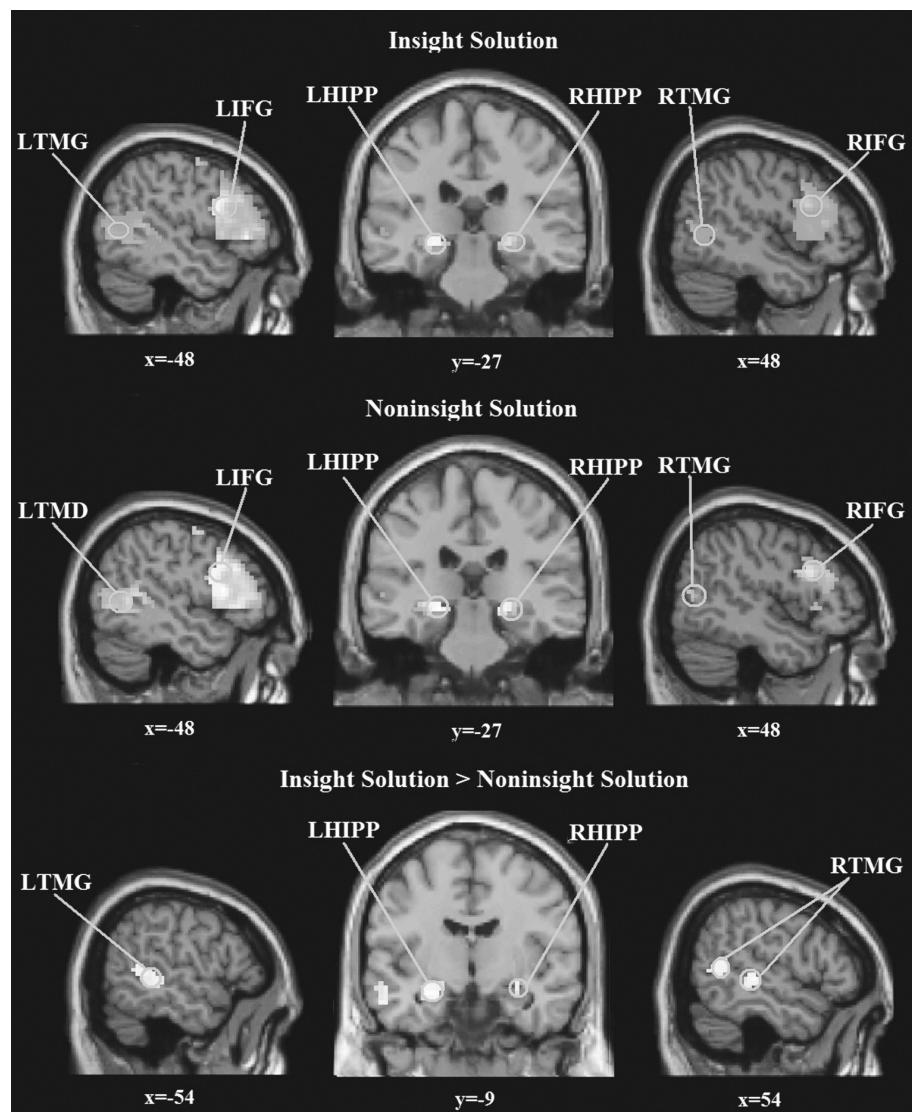
Fig. 2. Brain areas activated for insight solution and noninsight solution as well as the difference between the two conditions in the six regions of interest. For insight vs. rest and noninsight vs. rest, the thresholds were set at $p<0.001$ (False Discovery Rate control for multiple comparisons) and 30 or more contiguous voxels, while for insight vs. nonisight solution, the threshold was set at $p<0.05$ (False Discovery Rate control for multiple comparisons) and 30 or more contiguous voxels.

across all participants. For insight vs. rest and noninsight vs. rest, the thresholds were set at $p<0.001$ (False Discovery Rate control for multiple comparisons) and 30 or more contiguous voxels, while for insight vs. noninsight solution, the threshold was set at $p<0.05$ (False Discovery Rate control for multiple comparisons) and 30 or more contiguous voxels.

Functional connectivity. In the current work, we were interested in the functional connectivity among several brain regions including bilateral inferior/middle frontal gyri, middle/superior temporal gyri and hippocampi. Using the Automated Anatomical Labeling (Tzourio–Mazoyer et al., 2002), the activated voxels in these regions were obtained. The region of interest (ROI) was defined as a cube of $9 \times 9 \times 9$ mm centered at the activation peak in each brain region, and its BOLD time series was defined as the mean of all the voxels in the cube. If there were more than one peak in a region, the most central one located in the activation cluster was selected. Then, the time series associated with insight and noninsight solution

Table 1. Brain areas activated for insight solution and noninsight solution as well as the difference between the two conditions in the six regions of interest

Area	BA	Voxels	x	y	z	T	Z
Insight Solution							
Left inferior frontal gyrus	47	1051	−33	30	3	20.66	7.21
	44		−48	9	24	15.59	6.59
Left hippocampus	20	147	−24	−27	−9	18.13	6.93
	20		−33	−18	−9	9.47	5.42
Right hippocampus	27	123	21	−33	−3	14.28	6.40
	20		24	−27	−9	11.19	5.83
Right inferior frontal gyrus	47	731	36	27	3	10.51	5.68
	48		48	15	24	9.56	5.45
Left middle temporal gyrus	21	524	−60	−42	0	10.10	5.58
	22		−54	−45	9	9.94	5.54
Right middle temporal gyrus	37	113	45	−69	9	7.21	4.74
	37		45	−63	−3	6.72	4.57
Noninsight Solution							
Left hippocampus	20	82	−21	−27	−9	13.80	6.32
Left inferior frontal gyrus	48	914	−48	15	12	13.55	6.28
	44		−42	9	27	13.19	6.21
Right hippocampus	27	65	21	−30	−3	11.41	5.87
Right inferior frontal gyrus	47	348	36	27	3	10.82	5.75
	48		51	18	30	8.42	5.13
Left middle temporal gyrus	21	392	−57	−45	9	9.41	5.41
	21		−60	−33	3	7.58	4.87
Right middle temporal gyrus	37	32	48	−72	12	5.82	4.20
	37		42	−63	9	5.45	4.04
Insight Solution > Noninsight Solution							
Left hippocampus	20	48	−27	−9	−18	6.87	4.63
Right middle temporal gyrus	21	205	54	−30	0	6.41	4.45
	21		54	−54	12	5.39	4.01
Left middle temporal gyrus	22	126	−57	−36	3	5.09	3.87
	22		−54	−48	6	5.00	3.82
Left middle temporal gyrus	21	34	−54	−6	−24	4.96	3.81
Right hippocampus	20	6	30	−9	−12	4.74	3.69

BA, Brodmann area. Coordinates (x,y,z) were the MNI (Montreal Neurological Institute) coordinates. For insight vs. rest and noninsight vs. rest, the thresholds were set at $p<0.001$ (False Discovery Rate control for multiple comparisons) and 30 or more contiguous voxels, while for insight vs. noinsight solution, the threshold was set at $p<0.05$ (False Discovery Rate control for multiple comparisons) and 30 or more contiguous voxels. T–and Z–scores of the activations were also shown. Note that the 6 voxels in right hippocampus were included in a larger cluster with 116 contiguous voxels mostly located in right amygdala. At a looser threshold ($p<0.005$ uncorrected), there were 84 contiguous voxels more activated in right hippocampus for insight than noninsight solution.

was respectively segregated from the whole. Since the BOLD signal delayed for six seconds according to the haemodynamic response function by SPM, and the reaction time was about 4s (see in Results), the time series of each condition was defined from 0s to 10s after the presentation of optional answers. For each condition, the functional connectivity was evaluated by computing the temporal partial correlation between all pair−wise combinations of ROIs controlling the effects of the others (Friston et al., 1993). Finally, the correlation coefficients were compared across conditions after the Fisher transformation.

3 Results

On average, in 61.7% of trials participants selected the answers with novel associations (average reaction time was 3.70s with a standard deviation of 0.78s), and in 26.3% of trials they selected the answers with normal associations (average reaction time was 4.06s with a standard deviation of 0.94s). The larger trial percentage of insight solutions might result from the instruction before the experiment that asked participants to select a novel and reasonable answer.

Our fMRI results demonstrated that both of the insight and noninsight solution induced extensive changes of brain activity in the bilateral inferior frontal gyri, middle temporal gyri and hippocampi (see in Fig.2 and Table 1). Compared with noninsight solution, insight solution activated more in bilateral middle temporal gyri and hippocampi. The activation of the bilateral inferior frontal gyri showed no significant differences between two conditions.

As mentioned in Methods, the peaks of ROIs for insight solution were located at (−60, −42, 0) in left middle temporal gyrus (LMTG), (−48, 9, 24) in left inferior frontal gyrus (LIFG), (−24, −27, −9) in left hippocampus (LHIPP), (21, −33, −3) in right hippocampus (RHIPP), (45, −63, −3) in right middle temporal gyrus (RTMD) and (48, 15, 24) in

right inferior frontal gyrus (RIFG). And the peaks of ROIs for noninsight solution were located at (−57, −45, 9) in LMTG, (−42, 9, 27) in LIFG, (−21, −27, −9) in LHIPP, (21, −30, −3) in RHIPP, (42, −63, 9) in RTMD and (51, 18, 30) in RIFG.

The functional connectivity analysis showed that the ROIs were significantly connected with the ipsilateral regions and the homologous regions in the other hemisphere. In each hemisphere, there was a neural pathway from the inferior frontal gyrus to middle temporal gyrus to hippocampus for both insight and noninsight solution. The functional connectivity between the RIF6, and RTMD was stronger for insight than noninsight solution, while that between the LMTG and RHIPP was stronger for noninsight than insight solution (p<0.05, paired t−test).

4 Disscussion

4.1 Information selection in verbal Insight problem solving

As a type of traditional Chinese idiomatic expressions, the 'chengyu' could be regarded as a chunk. Then, the meanings of four component characters enhanced by the 'chengyu' should be their normal meanings, and their other meanings could be considered as novel ones. Although the left hemisphere is dominant in language processing, it is the right temporal cortex that processes the novel meaning of idioms (Faust & Mashal, 2007; Mashal, Faust, Hendler, & Jung−Beeman, 2008; Pobric, Mashal, Faust, & Lavidor, 2008). In the current work, the greater activation in right temporal gyrus for insight than noninsight solution indicated more retrieval of novel semantic information.

However, the novel semantic information is generally weak in retrieval (Giora, 1997). This conflict might be resolved by the lateral prefrontal cortex due to its role in establishing and shifting the attentional sets (MacDonald

et al., 2000, Luks et al., 2002). Although there was no significant difference of activation strength in bilateral prefrontal cortices at the given threshold between insight and noninsight solution, the RIFG was stronger functionally connected to the RTMD for insight than noninsight solution. The frontal–temporal neural pathway might reflect the selection of novel semantic information in verbal insight problem solving.

4.2 Forming novel association in verbal insight problem solving

Previous studies indicated the hippocampus was involved in the relational memory and its activation strength was associated with the novelty degree of stimuli (Luo and Niki, 2002, 2005, Johnson et al., 2008). Therefore, it was proposed as the key brain region of forming novel association (Luo and Niki, 2003, Zhao et al., 2013). In the present study, the bilateral hippocampi were involved in both insight and noninsight solution, but more activated for insight solution. These might reflect that the associations between the riddle and the selected answer were formed in both insight and noninsight solution, but the novel associations only existed in insight solution.

In insight 'chengyu' riddle solving, the forming of novel associations was dependent on the retrieval of the novel meaning of the key character. Our results showed that

the RHIPP was functionally connected with the RTMD in insight solution, Fig.3 but the strength was not larger than that in noninsight solution. However, the functional connectivity between the RHIPP and LMTG was stronger in noninsight than insight solution. This might reflect that the retrieval of normal semantic information resulted in the normal associations in noninsight solution.

4.3 Right–hemisphere theory in verbal insight problem solving

Since the key of creative problem solving is the selection of novel information (Bink and Marsh, 2000), and the right temporal cortex engages in relatively coarse and novel semantic processing (Beeman and Bowden, 2000, Mashal et al., 2008), it is proposed that there is a right–hemisphere advantage in insight problem solving (Bowden and Jung–Beeman, 2003). In particular, the right superior temporal gyrus is found in several studies and it is considered to facilitate the formation of remote associations (Jung–Beeman et al., 2004, Kounios et al., 2008, Zhang et al., 2011). However, the present study showed the bilateral middle temporal gyri involved in insight solution, even at the time just prior to the solution (Zhao et al., 2013). The paradigm adopted by the present study might be the reason why the advantage of right hemisphere was not found. Although the answer selection paradigm made participants
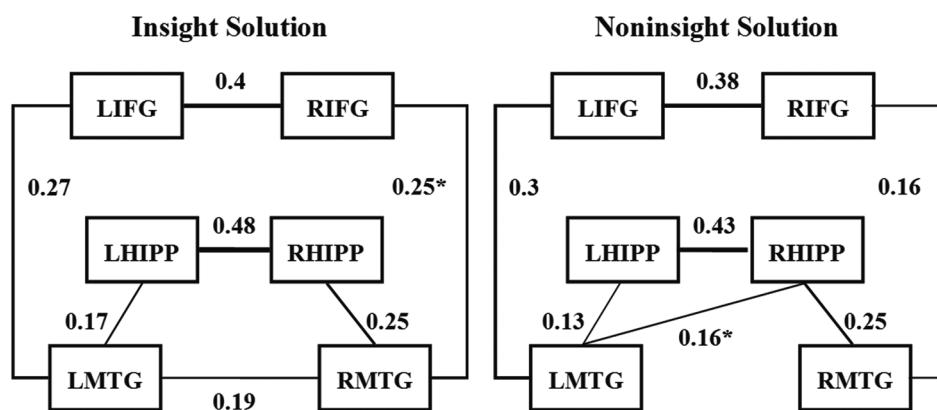
Fig. 3. The functional connectivity among the ROIs for insight and noninsight solution. A line between two regions indicates that the region–to–region correlation is statistically significant, and the thickness of the line reflects the strength of functional connectivity. The star means significant difference between two conditions (p<0.05, paired t–test).

more active than solution recognition, it still simplified the retrieval of the novel meanings and distant associates, and then weakened the activation of the right temporal cortex.

However, the brain activity contains two aspects: the activation of local regions and the interaction among them. Most previous debates about the right–hemisphere advantage in insight were only based on the results of the activation in local regions. In the present study, the functional connectivity among the key brain regions in insight and noninsight solution was evaluated. Results showed that there was a neural pathway from the inferior frontal gyrus to middle temporal gyrus to hippocampus for insight solution in the both hemispheres, but insight solution had greater ipsilateral frontal–temporal connectivity in the right hemisphere than noninsight solution. This implied that verbal insight problem solving needed the participations of both the left and right hemispheres, in which the right hemisphere might be more important compared with common verbal problem solving.

5 Conclusion

Our results demonstrated that verbal insight problem solving activated broad brain regions including the lateral prefrontal cortex, middle temporal gyrus and hippocampus in both hemispheres. There regions constituted a frontal–temporal–hippocampal neural pathway, especially in right hemisphere, which might reflect the selection of novel semantic information and the forming of novel associations. Compared with noninsight solution, insight solution showed greater ipsilateral frontal–temporal connectivity in the right hemisphere. Our result supported the right–hemisphere advantage theory of insight in a new angle of view.

References

Aziz–Zadeh, L., Kaplan, J. T., & Iacoboni, M. (2009). "Aha!": The neural correlates of verbal insight solutions. *Hum Brain Mapp, 30,* 908–916.

Beeman, M. J., & Bowden, E. M. (2000). The right hemisphere maintains solution–related activation for yet–to–be–solved problems. *Memory & cognition, 28,* 1231–1241.

Bink, M. L., & Marsh, R. L. (2000). Cognitive regularities in creative activity. *Review of General Psychology, 4,* 57–78.

Bowden, E. M., & Jung–Beeman, M. (2003). Aha! Insight experience correlates with solution activation in the right hemisphere. *Psychonomic bulletin & review, 10,* 730–737.

Bowden, E. M., & Jung–Beeman, M. (2007). Methods for investigating the neural components of insight. *Methods, 42,* 87–99.

Cohen, N. J., Ryan, J., Hunt, C., Romine, L., Wszalek, T., & Nash, C. (1999). Hippocampal system and declarative (relational) memory: summarizing the data from functional neuroimaging studies. *Hippocampus, 9,* 83–98.

Dietrich, A., & Kanso, R. (2010). A review of EEG, ERP, and neuroimaging studies of creativity and insight. *Psychological bulletin, 136,* 822–848.

Faust, M., & Chiarello, C. (1998). Sentence context and lexical ambiguity resolution by the two hemispheres. *Neuropsychologia, 36,* 827–835.

Faust, M., & Mashal, N. (2007). The role of the right cerebral hemisphere in processing novel metaphoric expressions taken from poetry: a divided visual field study. *Neuropsychologia, 45,* 860–870.

Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. (1993). Functional connectivity: the principal–component analysis of large (PET) data sets. *J Cereb Blood Flow Metab, 13,* 5–14.

Giora, R. (1997). Understanding figurative and literal language: the graded salience hypothesis. *Cognitive*

Linguistics, 7, 183–206.

Goel, V., & Vartanian, O. (2005). Dissociating the roles of right ventral lateral and dorsal lateral prefrontal cortex in generation and maintenance of hypotheses in set–shift problems. *Cereb Cortex, 15,* 1170–1177.

Huang, F. R., Zhou, Z. J., & Zhao, Q. B. (2013). An eye movement study of associate competition in Chinese idiom riddles solving. *Acta Psychologica Sinica, 45,* 36–46.

Johnson, J. D., Muftuler, L. T., & Rugg, M. D. (2008). Multiple repetitions reveal functionally and anatomically distinct patterns of hippocampal activity during continuous recognition memory. *Hippocampus, 18,* 975–980.

Jung–Beeman, M., Bowden, E. M., Haberman, J., Frymiare, J. L., Arambel–Liu, S., Greenblatt, R., Reber, P. J., & Kounios, J. (2004). Neural activity when people solve verbal problems with insight. *PLoS Biol, 2,* E97.

Köhler, W. (1925). The mentality of apes. London: Routledge & Kegan Paul.

Kaplan, C. A., & Simon, H. A. (1990). In search of insight. *Cognitive Psychology, 22,* 374–419.

Kershaw, T. C., & Ohlsson, S. (2004). Multiple causes of difficulty in insight: the case of the nine–dot problem. *Journal of experimental psychology Learning, memory, and cognition, 30,* 3–13.

Knight, R. (1996). Contribution of human hippocampal region to novelty detection. *Nature, 383,* 256–259.

Knoblich, G., Ohlsson, S., & Raney, G. E. (2001). An eye movement study of insight problem solving. *Memory & cognition, 29,* 1000–1009.

Kounios, J., Fleck, J. I., Green, D. L., Payne, L., Stevenson, J. L., Bowden, E. M., & Jung–Beeman, M. (2008). The origins of insight in resting–state brain activity. *Neuropsychologia, 46,* 281–291.

Luks, T. L., Simpson, G. V., Feiwell, R. J., & Miller, W. L. (2002). Evidence for anterior cingulate cortex involvement in monitoring preparatory attentional set. *Neuroimage, 17,* 792–802.

Luo, J. (2004). Neural correlates of insight. *Acta Psychologica Sinica, 36,* 219–234.

Luo, J., & Niki, K. (2002). Role of medial temporal lobe in extensive retrieval of task–related knowledge. *Hippocampus, 12,* 487–494.

Luo, J., & Niki, K. (2003). Function of hippocampus in "insight" of problem solving. *Hippocampus, 13,* 316–323.

Luo, J., & Niki, K. (2005). Does hippocampus associate discontiguous events? Evidence from event–related fMRI. *Hippocampus, 15,* 141–148.

Luo, J., Niki, K., & Knoblich, G. (2006). Perceptual contributions to problem solving: Chunk decomposition of Chinese characters. *Brain Res Bull, 70,* 430–443.

Luo, J., Niki, K., & Phillips, S. (2004). Neural correlates of the 'Aha! reaction'. *Neuroreport, 15,* 2013–2017.

MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science, 288,* 1835–1838.

MacGregor, J. N., Ormerod, T. C., & Chronicle, E. P. (2001). Information processing and insight: a process model of performance on the nine–dot and related problems. *Journal of experimental psychology Learning, memory, and cognition, 27,* 176–201.

Mai, X. Q., Luo, J., Wu, J. H., & Luo, Y. J. (2004). "Aha!" effects in a guessing riddle task: an event–related potential study. *Hum Brain Mapp, 22,* 261–270.

Mashal, N., Faust, M., Hendler, T., & Jung–Beeman, M. (2008). Hemispheric differences in processing the literal interpretation of idioms: converging evidence from behavioral and fMRI studies. *Cortex, 44,* 848–860.

Ormerod, T. C., MacGregor, J. N., & Chronicle, E.P. (2002). Dynamics and constraints in insight problem

solving. *Journal of experimental psychology Learning, memory, and cognition, 28*, 791–799.

Pobric, G., Mashal, N., Faust, M., & Lavidor, M. (2008). The role of the right cerebral hemisphere in processing novel metaphoric expressions: a transcranial magnetic stimulation study. *Journal of cognitive neuroscience, 20,* 170–181.

Qiu, J., Li, H., Jou, J., Liu, J., Luo, Y., Feng, T., Wu, Z., & Zhang, Q. (2010). Neural correlates of the "Aha" experiences: evidence from an fMRI study of insight problem solving. *Cortex, 46*(3), 397–403.

Qiu, J., Li, H., Yang, D., Luo, Y., Li, Y., Wu, Z., & Zhang, Q. (2008). The neural basis of insight problem solving: an event–related potential study. *Brain and cognition, 68,* 100–106.

Razoumnikova, O. M. (2000). Functional organization of different brain areas during convergent and divergent thinking: an EEG investigation. *Brain Res Cogn Brain Res, 10*, 11–18.

Redish, A. D. (2001). The hippocampal debate: are we asking the right questions? *Behavioural brain research, 127*, 81–98.

Takeuchi, H., Taki, Y., Sassa, Y., Hashizume, H., Sekiguchi, A., Fukushima, A., & Kawashima, R. (2010). White matter structures associated with creativity: evidence from diffusion tensor imaging. *Neuroimage, 51*, 11–18.

Tononi, G., Sporns, O., & Edelman, G. M. (1994). A measure for brain complexity: relating functional segregation and integration in the nervous system. *Proc Natl Acad Sci USA, 91*, 5033–5037.

Tzourio–Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single–subject brain. *Neuroimage, 15,* 273–289.

Wu, L., Knoblich, G., & Luo, J. (2013). The role of chunk tightness and chunk familiarity in problem solving: Evidence from ERPs and fMRI. *Hum Brain Mapp, 34*(5), 1173–86.

Zhang, M., Tian, F., Wu, X., Liao, S., & Qiu, J. (2011). The neural correlates of insight in Chinese verbal problems: an event related–potential study. *Brain Res Bull, 84*, 210–214.

Zhao, Q. B., Zhou, Z. J., Xu, H. B., Chen, S., Xu, F., Fan, W. L., & Han, L. (2013). Dynamic neural network of insight：A functional magnetic resonance imaging study on solving Chinese 'chengyu' riddles. *PloS one, 8*, e59351.

Zhou, Z. J., Xu, H. B., Zhao, Q. B., Zhao, L. L., & Liao, M. J. (2011). The processing of novel semantic association in Chinese: Converging evidence from behavior and fMRI studies. *The 4th International Conference on Image and Signal Processing (CISP 2011), Shanghai, 3,* 1588–1592.

Zhu, X. C., Li, R. J., & Zhou, Z. J. (2009). The role of clues in Chinese idiom Riddle solving. *Acta Psychologica Sinica, 41,* 397–405.

人格与社会心理研究所团队成员

郭永玉，博士，教授，人格与社会心理研究所所长，主要研究领域为中西人格理论，社会历史文化背景下的人格研究，社会治理中的心理学问题。

佐斌，博士，教授，华中师范大学社会心理研究中心主任，主要研究领域为文化与社会心理学理论及其应用、青少年儿童对社会与文化的理解、青少年人格发展与心理健康教育等。

李晔，博士，教授，人格与社会心理研究所副所长，主要研究领域为教师心理、师生关系、文化与社会认知；组织公平感与社会公平感等。

贺金波，博士，副教授，主要研究领域为人格的生理机制，网络成瘾的发生机制，网络与青少年心理健康，前注意的神经机制，酒精对前注意的影响等。

王伟，副教授，主要研究领域为社会心理学理论。

黄飞，博士，讲师，主要研究领域为青少年人格发展与评估，人格知觉，人际关系与社会发展，测量等值性等。

温芳芳，博士，讲师，主要研究领域为社会认知刻板印象，社会与群体认同，青少年网络心理与行为，面孔吸引力等。

李静，博士，讲师，主要研究领域为物质主义，收入与幸福，社会阶层心理学等。

人格心理学：人性及其差异的研究 *

郭永玉，李　静，胡小勇

（华中师范大学心理学院，青少年网络心理与行为教育部重点实验室，武汉 430079）

摘　要　人格心理学是心理学学科体系中注重从整体的视角探究人性本质的一个分支，它以人性及其差异作为其核心，研究对象是作为整体的人。人格心理学不仅在心理学的学科体系内部处于重要地位，而且在关于人的所有生命科学、社会科学和人文学科中也处于基础性的位置。经过 100 多年曲折的发展，人格心理学已经步入了一个新的发展和繁荣时期。近 20 年来西方人格心理学的研究进入了快速发展阶段，在研究范式、研究方法以及研究内容（包括人格结构、人格动力、人格发展）方面都有了较大的进展。与此同时，我国学者在大量介绍西方人格心理学的基础上，开始着手研究中国人的人格问题。当前中国社会文化背景下的人格心理学研究主题主要包括人格与创造力、人格与人员选拔及安置、人格与贪腐行为、人格与暴力犯罪、人格与疾病、以及和谐社会的健全人格建构问题。鉴于人格心理学具有重要的科学价值及其在社会进步中所能做出的重要贡献，建议未来中国人格心理学研究应从研究现代化背景下中国人的人格和研究方法多元化等方面予以加强。

关键词　人格心理学；人性及其差异；现代化人格研究

1　引言

我们中国人说到"人格"，神情往往不由自主地严肃起来，因为这个词往往具有法律和道德的涵义。法律上讲"保护人格尊严"，"不能侮辱人格"，是将人格视为权利义务的主体。日常话语中讲"人格高尚"或"人格低下"，甚至"没有人格"，是将人格视为道德品质，与人品、品格或品德同义。事实上，古汉语中并无"人格"一词，这个词是近代从日文中来的，而日文"人格"一词又是对英文"personality"一词的翻译。这个英文词也可以译为"人性"，是指人（person）的各种特征，并没有道德（以至"道德高尚"与否）的含义。它首先是一

个事实性的概念，而不是一个评价性的概念。因此，心理学探讨人格和探讨感知、记忆、思维、情绪、智力等心理现象一样，也是认识人类自身的一种研究活动。只是心理学家研究感知是为了了解感知现象和规律，研究记忆是为了了解识记、保持和遗忘等现象和规律……，人格心理学就是将完整的人作为研究对象，不仅仅是研究人的某一种心理或行为。

人格心理学的研究对象是整体的人。但要研究整体，仍需要对其加以分析。人格心理学家大体从3 个层面分析一个人：第一，人类本性的层面（the human nature level），即一个人首先是人，与所有人相似（like all others）；第二，个体差异和群体差异层面（the level of individual and group differences），

即一个人与部分他人是相似的（like some others），个体之间的差异仅仅是程度的差异，如外向的程度不同而已，并且一个人与其所在群体的其他成员具有相似性，但与其他群体的成员明显不同；第三，个人独特性层面（the individual uniqueness level），即一个人不同于任何人的（like no others）、独特的、不可重复、不可替代的特征。

人格心理学的任务或目的与其他心理学分支一样，都是寻求准确地描述，合理地解释，有效地预测和控制。但人格心理学的独特性在于其是心理学学科体系中特别注重从整体的视角探究人性本质的一门学科，它将人性作为其核心，关注整体的人。正如 Hergenhahn 所说："在把人作为一个整体来研究的心理学中，人格心理学处于独特的地位。绝大多数其他分支的心理学家往往只深入研究人的某一方面。比如，他们只研究儿童发展、老年问题，或知觉、智力、学习、动机、创造性等。只有人格心理学家才试图描绘出人的完整的图画。"人格心理学不仅在心理学的学科体系内部处于重要地位，而且在关于人的所有生命科学、社会科学和人文学科中也处于基础性的位置。它与所有关于人性的学科有关，并整合这些学科关于人性的知识。

人格心理学的研究不仅对于从整体上把握人的心理和行为的独特模式具有深刻的理论意义，而且还具有广泛的实践意义，研究人格有助于提高教育、生产、管理、医疗、资讯、司法、体育和军事等各种活动的效率和绩效。

2 人格心理学的发展脉络

一般认为，现代人格心理学的正式诞生以 Gordon W. Allport（1897-1967）所著《人格：心理学的解释》（1937）和 Henry A. Murray（1893-1988）所著《人格探究》（1938）两书的出版为标志。自这两本书问世后，关于人格心理学的研究才得以蓬勃开展，而且大学心理学系也因此开设了人格心理学课程。人格心理学的历史可以追溯到科学心理学的创始人 Wundt 等人的工作，大致经历了 4 个阶段，即奠基期、理论体系形成期、基本人格结构确立期、质疑与复兴时期。

2.1 奠基期（1937 年之前）

Wundt、James、Freud 等人的工作，无论是对于整个心理学还是人格心理学都是开创性的，意义深远。其中 Freud 对人格心理学发展的贡献尤其巨大。Freud 创立了精神分析理论，该理论内容十分庞杂，是所有人格理论中内容最丰富、影响最大的人格理论。该理论强调潜意识、性本能等人格动力的重要性，其研究的主要问题集中在焦虑、防御机制、早期经验对日后人格发展的重要性以及个人发展出来的处理内部驱力和外部刺激的自我适应功能。精神分析理论的研究方法是临床的个案研究，其评鉴技术主要采用梦的分析、自由联想和投射测验等。

2.2 理论体系形成期（1937–20 世纪 50 年代）

20 世纪 30—50 年代，人格心理学发展成为独立的分支。其标志性事件是 Gordon W. Allport（1897–1967）所著《人格：心理学的解释》（1937）和 Henry A. Murray（1893–1988）所著《人格探究》（1938）两书的出版。这一时期，人格心理学家试图将那些通过相关、临床和实验手段得到的研究结果整合起来，建构可以解释整体人格的"大理论"。其中较具影响力的包括 Allport、Murray 和 Raymond B. Cattell 等人的理论。

Allport 最主要的贡献在于，试图在理论上提出一种架构，用以解释每个人身上共性和特性的方面。他强调个体人格的结构和组织，对特质词汇研究有开拓性贡献，在研究方法上强调特殊规律研究法与一般规律研究法相结合。他的理论直接促成了人格心理学的建立。Murray 经由精神分析和变态心理学而进入人格研究领域，采用折衷的、多方法的研究取向，对 Freud 的精神分析和实验心理学进行了开创性的结合，他的工作对于拓宽人格心理学的研究领域具有广泛的影响。Murray 最主要的理论是动机理论，提出了大约 20 种需要，并创造了主题统觉测验这种动机研究方法。Cattell 对人格特质进行了深入的研究，是

特质心理学史上的伟大人物之一。他深受 Spearman 因素分析法的影响，把因素分析视为确定人格基本单元的最好方法，据此编制了著名的 16PF 问卷。

这一阶段，精神分析学派也在蓬勃发展中，著名的心理学家有 Adler、Jung、Horney、Fromm、Erikson 等人。

2.3 基本人格结构确立期（1950–20 世纪 70 年代）

这一阶段，人格心理学的理论构建基本完成，面临的主要问题是确定基本的人格结构，因此有关人格测量的一些问题就越来越为研究者所关注。这一时期的人格问卷以明尼苏达多项人格问卷(MMPI) 为典型代表。此外，著名的人格问卷还有加利福尼亚心理问卷（CPI）、爱德华个人偏好调查表（EPPS）以及前文提及的 16 人格因素问卷（16PF）等。

这一时期，4 大人格主题（也是人格结构的重要方面）备受关注，获得了大量的实证研究成果，分别是：权威主义（authoritarian）、成就动机（achievement motive）、焦虑（anxiety）和场独立性（field–independence）。这些主题跟美国当时的社会文化环境以及欧美各国对二战历史的反思有关。这段时期人格心理学的研究一方面注重人格结构的理论与实证研究，成果累累，另一方面逐渐远离社会的应用需求，所以也埋下了后一个阶段面临重重危机的种子。

2.4 质疑与复兴时期（20 世纪 70 年代至今）

20 世纪 50 年代中后期，越来越多的自相矛盾的实证研究结果、测量中不断出现的错误以及学科内部统一性的缺乏，使学术界开始质疑人格心理学。随着 Mischel（1968）等人对特质论的批评，对人格心理学的质疑与不满在 70 年代达到顶峰。Mischel 指出，我们几乎不可能依据人格特质去预测一个人的行为。此外，学术界以外的社会与文化变革似乎也不利于人格心理学的发展。70 年代的美国社会发生了巨变，确定人们的基本类型和稳定的人格差异变得不合时宜。种种不利条件的作用，使人格心理学步入低谷时期。

到 80 年代，局面逐步有所改观。自我的研究，认知取向人格心理学的兴起，"大五"的出现，行为遗传学、神经科学和进化心理学的研究，以及各种研究方法的完善，为人格心理学的发展提供了新的动力。由于这些进步，人格心理学在 20 世纪后期步入了一个新的发展和繁荣时期。

3 西方人格心理学的研究进展

万晓霞以美国科技信息研究所出版的《科学引文索引》（SCI）为数据源检索人格心理学文献，对近 10 年 SCI 人格心理学研究文献进行计量分析，结果显示，人格心理学研究近 10 年进入了快速发展阶段，文献量呈逐步上升趋势（如表 1 所示），说明目前人格心理学的研究呈现出一派繁荣景象。

从学科分布来看，人格心理学研究广泛分布在 138 种学科中，其中既包含社会科学领域又涉及到自然科学领域。收录人格心理学论文排名前 10 名的学科为：社会心理学、多学科心理学、精神病学、临床心理学、神经科学、心理学、临床神经学、应用心理学、教育心理学、医学遗传心理学。说明人格心理学具有自然科学和社会科学的交叉学科的特征。

从研究范式来看，精神分析、特质论、行为主义和人本主义是人格研究的传统范式，近 20 年来，这 4 种范式都已扩展了各自的领域，并繁衍出一些新的人格研究范式，包括：社会—认知范式、生物学范式、积极心理学范式，这三者分别源自于行为主义、特质论和人本主义。当代的依恋研究则得益于精神分析的发展。除此以外，进化心理学范式和后现代

表 1　人格心理学 SCI 十年载文量统计

年份	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	总计
文献量	69	61	89	207	169	185	188	223	268	282	1741
%	3.96	3.50	5.11	11.89	9.70	10.63	10.80	12.82	15.39	16.20	100.00

心理学范式（如叙事心理学）被视为人格心理学的新范式 。

从研究内容来看，当代人格心理学在人格结构、人格动力、人格发展等领域都有较大的进展，分别简要介绍如下：

3.1 人格结构

20 世纪末，人格领域最令人欢欣鼓舞的进展应该是 2 个相似的人格分类系统——"大五"结构（"Big Five" Structure）和五因素模型（five-factor model, FFM）的出现。两种模型分别是词汇学取向和理论取向研究成果的结晶，但让人惊叹的是，两种取向的研究殊途同归，最终在人格结构的问题上达成了初步共识。毫无疑问，人格特质的五因素模型（"大五"）是当前人格研究的主导范式，在整个心理学界都是最有影响力的模型之一。正如 McCrae 等人所言，"五因素模型"就像一颗圣诞树，与综合性、稳定性、遗传性、会聚效度、跨文化普适性和预测效度有关的研究成果正是满缀其间的圣诞礼物。直到近两年，五因素人格模型仍然是活跃于权威期刊《人格与社会心理学杂志》（*Journal of Personality and Social Psychology*）上的重要研究主题，具体内容涉及到一些更为细小的争议较多的问题，例如，大五人格因素的代际差异、年龄差异、大五人格因素之间的相关是由测量误差造成的还是由更高阶因素导致的、使用简式问卷测量大五人格特质的有效性，等等。随着这些研究的开展，人们对于人格结构的认识会越来越深入。

3.2 人格动力

人格动力领域有两大较为突出的理论进展。一是 Mischel 基于传统特质理论无法准确地预测和解释跨情境的行为变化的主要缺陷，提出了著名的认知—情感人格系统（cognitive-affective personality system, CAPS）理论。该理论试图在人格研究中引入情境因素，强调个体的人格系统与外部环境的动态交互作用，将人格的基本单元（如特质）视为"如果……那么……"的系统，人格便是大量"如果……那么……"的集合，于是出现在什么样的情境中，不同"如果……

那么……"图式便会指引着人们做出不同的行为。这样不仅考虑了人格的稳定结构，而且还兼顾了人格的动力过程。事实上，研究者已经通过实验研究证实了人际知觉中"如果……那么……"图式的有效性，具体来说，在对他人的社会行为和人格倾向进行解释的过程中，人们会考虑人与情境的交互作用，并以"如果……那么……"的方式进行描述。此外，由于认知—情感人格系统理论是一个元理论（meta-theory），即它是不包含具体内容的。近年来，研究者已将此元理论应用于建构特定领域（如戒烟、精神病理学、组织行为）的包含具体内容的模型。

另一重大的理论进展是由 Deci 和 Ryan 提出的自我决定论（self-determination theory），使人们对人类动机的普遍性有了新的认识。Deci 和 Ryan 认为，人类具有 3 种基本的普遍的心理需要，即自主（autonomy）、胜任（competence）和关系（relatedness）的需要。这 3 种心理需要的满足对于个体的幸福感、心理健康甚至生理健康都是必需的。在自我决定论的理论框架下，研究者开展了大量的实证研究，内容涉及到动机的类型，社会环境对不同类型动机的影响，不同类型动机对一系列结果变量如学习、绩效、认知功能和幸福感的影响，不同的抱负或生活目标与基本心理需要的满足及绩效和幸福感等结果变量的关系，基本心理需要满足的跨文化研究，以及自我决定论在养育、教育、工作和医疗等具体生活领域的应用。

3.3 人格发展

稳定与变化是人格发展的永恒主题。从研究的数量和规模来看，人格特质的发展是当前研究的主要内容，从婴儿气质怎样发展为成人特质，特别是成年期人格特质如何发展，是当今研究得最为广泛的问题。目前研究者进一步关注人格特质稳定性与可变性的深层影响因素，如年龄、生活事件等。除了特质的发展以外，动机和目标的发展及叙事认同的发展也成为人格发展领域新的研究内容。对于人格发展的影响因素，当代的行为遗传学研究、神经科学研究和进化人格心理学研究已为天性的作用提供了越来越多的证据，而教养的作用如家庭环境对人格的

影响也积累了丰富的成果。随着文化心理学的兴起，研究者越来越重视社会文化因素对人格发展的影响。将人格置于特定的社会文化背景下进行研究，有助于获得对人格的更为生动、具体、深刻的理解。总之，人格发展的个人与情境交互作用的观点已深得人心。

4　中国社会文化背景下的人格心理学研究

20 世纪 70 年代末，我国大陆地区开始恢复心理学教学和研究，西方人格心理学也得到介绍；90 年代以来，我国心理学者随着反思西方人格心理学的理论和研究方法论问题，开始着手研究中国社会文化背景下的人格问题。

在理论研究层面，体现为学者对中国人的人格结构、动力和影响因素等一些具体问题开展了探索性的研究。人格结构的研究是人格心理学的一个重要范畴，是了解人格的基本特点、类型以及对个体进行有效评估的基础。杨国枢等较早地进行了相关的本土研究。他从中文人格特质形容词入手，得到了 4–5 个独立的人格维度；王登峰将杨国枢收集到的用于描述稳定人格的形容词与从现代汉语词典和刊物中收集到的词汇合并，用因素分析法进行研究，最后确定中国人人格结构的 7 个维度，并编制了中国人人格量表；张建新等人将他们自己编制的《中国人人格测量表（CPAI）》与西方的五因素问卷（NEO–PI）合起来进行联合因素分析，研究结果显示出一个六因素结构。许燕等运用词汇学方法，通过抽取中文动词 2012 个分析，形成 120 个词汇的动词词表，因素分析结果发现，以动词建立的人格结构有控制、施爱、追求成功 3 个因素；控制与施爱属于关系特质，追求成功属于个人特质。此人格动态结构是以行为为基础的模型，称为中国人人格 CLP 模型。所有这些结果都表明，中国人与西方人的人格结构有共同性也有特殊性。此外，还有许燕、张进辅对价值观等人格动力进行了研究；申继亮对人格发展进行了研究等。

在应用研究层面，体现为我国心理学者们立足

本国实际，借鉴西方心理学的方法，去解决我国经济社会发展中有关人格心理学的问题。这些问题构成了当前中国人格心理学的研究主题。大致说来，有如下一些主题。

4.1　人格与创造力

创造型人物的新发现、新发明和新成果，对整个社会文明进步有着重要的意义。心理学家关注人格对个体创造活动的影响。王极盛用自评法调查了 28 位学部委员和 127 位一般科学工作者，发现影响创造活动的主要人格因素有：事业心、勤奋、兴趣、责任心、求知欲、进取心、意志等。张景焕对 34 位院士进行访谈，发现创造人才心理特征排在前 4 位的是一般智力强、勤奋努力、内在兴趣和研究技能策略。更有研究者在与国外学者的研究结果对照后提出了两类创造型人格特征的假设：一类称作创造型人格特征的内核，是与创造力关系最为密切且比较稳定的部分，另一类称为创造型人格特征的外壳，它是较多受到文化背景影响的创造型人格特征。

4.2　人格与人员选拔及安置

人格测验对组织中的人员安置和选拔具有重要的意义。从 20 世纪 80 年代初，随着改革开放，外资企业进入中国，为中国带来了先进的管理思想、观念和技术，推动了人格测验在人事管理中的应用，一批学者和专家开始关注和着手人格测验在中国企业人事管理中的应用，并致力于研究开发具有自主知识产权、体现中国特色、适用于中国文化的人格测验。例如，王重鸣和陈民科建立的管理胜任力模型，王登峰对中国党政干部的胜任特征的研究。更有研究者考察了 MBTI 人格测验对陆军指挥院校学员心理选拔的预测性，发现 MBTI–G 人格类型量表对陆军指挥院校学员胜任特征评价有一定预测性，可以作为选拔工作使用；他们还建立了初级军官、航天员和陆军学院学员等军队人员胜任特征模型。这些基于工作特性的人格模型的建构，为人事决策提供更全面和更科学的信息，提高了组织人事决策的效率，可帮助组织更有效地进行人员的聘用、选择、训练、开发等。

4.3 人格与贪腐行为

腐败问题是当今中国的严重社会问题。目前，从宏观层面来说，制度建设在不断完善、监管力度在持续加大；在微观层面，对于腐败主体——个人的腐败心理动因认识不清，对于腐败过程的心理机制了解不够。在西方文化背景下展开的研究表明，存在一种固化和内化的，为达目的不择手段、操纵他人、谋取私利的人格特质，即马基雅弗利主义。中国学者已通过实证方法证实了在中国人身上存在这种人格特质，并且设计了信效度良好的测量工具。在西方文化背景下予以证实的马基雅弗利主义和贪腐行为、经济机会主义是正相关的，在信息不对称的情况下，马基雅弗利主义者倾向于利用手中的优势，使自己的利益最大化。但这一发现，在我国文化背景下迄今为止还没有研究过。在贪腐行为较为严重的当今中国，马基雅弗利主义人格在其中到底起到何种作用？如何起作用？如何抑制？亟待回答。许燕等研究了腐败过程的心理机制——心理绑架，初步建立了心理绑架的现象模型。

4.4 人格与暴力犯罪

通过2010年的《法治蓝皮书》可知，现阶段我国的暴力犯罪现象十分严重，并且发展的趋势也越来越严峻。更令人担忧的是：在越来越多的暴力犯罪中，由于反社会人格而导致严重暴力犯罪的案件数量更是比往年有增无减。反社会人格者由于其易于冲动、不吸取经验教训、不能爱别人和缺乏内化了的社会价值系统或良心的特点，非常容易触犯社会规范和法律。在违法犯罪人群中具有反社会人格的人的数量较多，可达30%以上，远高于一般人群中的比例（1%以下），且屡次犯罪以及罪行特别残酷或情节恶劣的现象非常严重。我国学者通过问卷调查发现了反社会人格者的一些特征，但反社会人格如何导致暴力犯罪，其间的过程是怎样的；是否有其他变量的中介作用或交互作用；如何控制这些因素来减弱或防止它在暴力行为发生过程中起到的作用；如何在暴力犯罪发生之后，更加彻底地去了解暴力犯罪人，有效帮助他们改造，减少累犯的几率等研究刚刚起步，

因此迫切需要开展更深入和更广泛的系统性研究。

4.5 人格与疾病

许多研究表明，一个人的人格特征与疾病之间存在十分密切的关系，人格直接或间接地影响个体的心理和生理健康，具有某些人格特征的人面临患某些特定疾病的风险。国内研究者指出与疾病有关的4组人格因素：易发怒和具有敌意；情绪性压抑；有失望经历；悲观与宿命论的态度。在这4组因素中，敌意倾向与发怒对心脏病的发病有影响，情绪压抑与心脏病和癌症的产生有关。在压抑的情境下不愿表达情感以及对抑郁心情的压抑是癌症产生的最主要原因。

正如Eysenk所说："已有足够的证据可以说明在人格与压力以及疾病之间存在着必然的联系，这种联系影响免疫系统的功能……人格与压力因素是癌症产生的重要原因"。然而，该研究领域正处于不断发展和逐步完善的过程中，也面临着许多有待解决和探讨的问题。例如，如何运用心理学知识改进医疗与护理制度，建立合理的保健措施，节省卫生经费和减少社会损失的途径，以及对有关的卫生决策提供建议等。

4.6 和谐社会的健全人格建构问题

当前我们正在致力构建和谐社会，诚如黄希庭所言，构建和谐社会提出了许多亟待解决的人格心理学问题，其中尤其重要的是健全人格的形成问题。健全人格的人能以辨证的态度对待世界、他人、自己、过去、现在和未来、顺境和逆境，是一个自立、自信、自尊、自强、幸福的进取者。黄希庭及其团队开展了一系列研究，对中国人的自我价值感和自立、自信、自强人格等进行探讨；陈建文对健康人格结构、人格功能与人格状态的探讨，为培养和塑造中国人的健康人格提供了理论支持。

5 发展我国人格心理学的建议

中国人格心理学工作刚刚起步，还有许多理论和实证研究要去完成。我国人格心理学应该朝着什么方向发展？我们认为以下两点十分重要。

5.1 现代化背景下的中国人人格研究

英格尔斯在对人的现代化问题作了长达 20 多年研究之后，如是说："一个国家，只有当它的人民是现代人，它的国民从心理和行为上都转变为现代的人格，它的现代政治、经济和文化管理机构中的工作人员获得了某种与现代化发展相适应的现代性，这样的国家才可真正称之为现代化的国家"。可见，国家现代化首先是人的现代化，而人的现代化，最根本的是人格现代化。

随着改革开放政策的实施和市场经济制度的确立，中国正沿着现代化的道路迅猛发展。作为一场深刻的社会变革，中国的现代化建设一方面带来了市场经济的繁荣发展，另一方面又使得整个社会环境产生了躁动起伏的剧烈变化。与此同时，人的问题也变得越来越突出。虽然传统人格在一定程度上还是具有适应性的，但是我们更应该看到，传统人格确有很多特征是不适应甚至阻碍现代化发展的。因此，在我们这个古老的民族从传统负重之下迈向现代化的今天，研究现代化人格是极具现实意义的课题。心理学家杨国枢等做了一些开创性的工作，初步探讨了现代化人格的内涵、特征及影响因素。但是现代化人格的形成机制是怎样的？该如何塑造？等等，都是摆在我国心理学者面前紧迫的研究任务。同时，前文提及的当前中国人格心理学的一些备受关注的研究主题也是现代化背景下的中国人人格研究的重要范畴。

5.2 研究方法的多元化

人格研究目前使用的方法可大致归纳为实验法、临床法与问卷调查，而这些方法本身都面临一对矛盾，即内部效度与外部效度难以两全的问题。严格控制变量的实验研究保证了内部效度，却很难将复杂的社会文化变量还原为个别实验室变量，导致研究的生态效度低下；相反，临床研究较好地还原了人的生活场景，但由于变量不易控制，研究的内部效度不尽人意，难以精确地刻画出变量间的因果联系。此外，问卷法虽然在一定程度上吸收了前两种方法的优点，但仍存在理论基础与现实情境相脱节、被试回答真实

性、量表预测效度、测量目标的含义难以确定等问题，这些问题将直接影响到研究结论的效度。

很显然，上述 3 种研究方法各有侧重和忽略，各有优势和不足。由于人格现象十分复杂，我们必须多种研究方法（如文献分析法、深度访谈法、问卷调查法、测量法、实验法、叙事研究法、故事谚语分析法，以及遗传学和神经科学的方法等）并用，即采用多元化的研究方法才能对所要研究的人格问题有一个全面而深入的把握。

参考文献

陈建文 . (2008). 人格与社会适应 . 合肥：安徽教育出版社 .

陈静，苗丹民，罗正学等 . (2007). MBTI 人格测验对陆军指挥院校学员心理选拔的预测性 . 第四军医大学学报，28(16), 1527–1 529

邓晓芒 . (1989). "人格" 辨义 . 江海学刊，8(3), 116–119

郭永玉 .(2005). 人格心理学：人性及其差异的研究 . 北京：中国社会科学出版社 .

郭永玉，张钊 . (2007). 人格心理学的学科架构初探 . 心理科学进展，15(2), 267–274.

侯玉波，张梦 . (2009). 对中国人自我结构的理论分析 . 心理科学，1, 226–229

黄希庭 . (2007). 构建和谐社会：呼唤中国化人格与社会心理学研究 . 心理科学进展，15(2), 193–195.

黄希庭 . (2002). 人格心理学 . 杭州：浙江教育出版社，5.

黄希庭 . (2006). 时间与人格心理学探索 . 北京：北京师范大学出版社，321.

蒋奖，许燕 . (2007). 罪犯反社会人格障碍的调查 . 中国特殊教育，5, 80–85

况志华，叶浩生 . (2005). 当代西方心理学的三种新取向及其比较 . 心理学报，37(5), 702–709.

刘邦惠，张庆林，谢光辉 . (1994). 创造型大学生人格特征的研究 . (5), 553–557.

申继亮 , 陈勃 , 王大华 . (1999). 成人期人格的年龄特征： 中美比较研究 . *心理科学 , 22*(3), 202–205.

汤舒俊 . (2011). *厚黑学研究* . 华中师范大学博士学位论文 .

王重鸣 , 陈民科 . (2002). 管理胜任力特征分析： 结构方程模型检验 . *心理科学 , 25*(5), 513–516.

王登峰 . (2002). *心理学研究的中国化： 理论与策略* . 北京： 轻工业出版社 , 166–194.

王芳 , 刘力 , 许燕等 . (2011). 社会心理学： 探索人与社会的互动推动社会的和谐与可持续发展 . *中国社科院院刊 , 26*(6), 640–649.

王极盛 , 孙福立 . (1984). 科技工作者创造力的研究 . *科学学研究 , 2*(4), 36–43.

万晓霞 . (2009). 近十年 SCI 人格心理学研究文献计量分析 . *心理科学进展 , 17*(6), 1 291–1 286.

许燕 , 王芳 . (2008). *社会变迁与大学生价值观的演变 . 科学发展 : 社会秩序与价值建构——纪念改革开放 30 年论文集》* (上卷). 北京 ; 北京师范大学出版社 , 277–287.

许燕 , 王萍萍 . (2011). 基于动词分析的人格结构模型探索 . 台北 . 第七届华人心理学家大会论文集 .

杨国枢 . (1974). *中国 "人" 的现代化* . 载杨国枢 : 中国人的蜕变 . 台北 : 桂冠图书公司 , 389.

杨国枢 , 李本华 . (1971). 五百五十个中文人格特质形容词之好恶度、 意义度及熟悉度 . *国立台湾大学心理学系研究报告 , 13*, 36–57.

英格尔斯 , 殷陆军译 . (1985). *人的现代化* . 成都： 四川人民出版社 .

张景焕 . (2005). 科学创造人才心理特征及影响因素研究 (博士学位论文). 北京师范大学 .

张进辅 . (2006). *青少年价值观的特点： 构想与分析* . 北京： 新华出版社 .

张兴贵 , 郑雪 . (2002). 人格心理学研究的新进展与问题 . *心理科学 , 25*(6), 744–745.

张建新 , 周明洁 . (2006). 中国人人格结构探索——人格特质六因素假说 . *心理科学进展 , 14*(4), 574–585.

Chang, L., Connelly, B. S., & Geeza, A. A. (2012). Separating methodfactors and higher order traits of the Big Five: A meta-analytic multitrait multimethod approach. *Journal of Personality and Social Psychology, 102* (2), 408–426.

Credé, M., Harms, P., Niehorster, S. et al. (2012). An evaluation ofthe consequences of using short measures of the BigFive personality traits. *Journal of Personality and Social Psychology, 102*(4), 874–888.

Cheung, F. M., vande, Vijver, F. J. R., Leong, F. T. L. (2011). Toward a new approach to the study of personality in culture. *American Psychologist, 66*(7), 593–603

Deci, E. L., & Ryan, R. M. (2009). Self-determination theory: A consideration of human motivational universals. In P. J. Corr & G. Mathews (Eds.), *The Cambridge handbook of personality psychology* (pp. 441–456). New York, NY: Cambridge University Press.

Eysenk, H. J. (1996). Personality and cancer. In C. L. Cooper (Ed.), *Handbook of stress, medicine and health* (193–215). New York: CRC Press.

Hergenhahn, B. R. (1980). *An introduction to theories of personality*. Englewood Cliffs, NJ: Prentice-Hall. Hergenhahn, B. R. (1988). *现代人格心理学历史导引* (文一 , 郑雪等 编译). 石家庄 : 河北人民出版社 .

Kammrath, L. K., Mendoza, Denton. R., Mischel, W. (2005). Incor-porating if...then...personality signatures in person perception: Beyond the personsituation dichotomy. *Journalof Personality and Social Psychology*, *88*(4), 605–618.

Kluckhohn, C., & Murray, H. A. (1953). Personality formation and its determinants. *In C. Kluckhohn & H. A. Murray (Eds.)*, Personality, its nature, society, and culture (2nd ed., pp. 53–67).

Maria, S., Clive, R., & Yves, T. (2007). Machiavellianism and economic opportunism. *The Journal of Applied Social Psychology, 37*(1), 181–190.

McAdams, D. P., & Olson, B. D. (2010). Personality development: Continuityand change over the life course. *Annual Review of Psychology, 61,* 517–542.

McCrae, R. R. (2009). The five–factor model of personality traits: Consensus and controversy. *In P. J. Corr & G. Matthews (Eds.)*, The Cambridge handbook of personality psychology (pp.148–161). Cambridge: Cambridge University Press.

Shoda, Y., & Mischel, W. (2006). Applying metatheory to achievegeneralisa bility and precision in personality science. *Applied Psychology: An International Review, 55*(3), 439–452.

Mischel, W., & Shoda, Y. (1995). Acognitive affective system theoryof personality: Reconceptualizing situations, dispositions,dynamics, and invariance in personality structure. *Psychological Review, 102*, 246–268.

Smits, I. A. M., Dolan, C. V., Vorst, H. C. M. et al. (2011). Cohort differences in Big Five personality factors over a period of 25years. *Journal of Personality and Social*

Psychology, 100(6), 1124–1138.

Soto, C. J., John, O. P., Gosling, S. D. et al. (2011). Age differences inpersonality traits from 10 to 65: Big Five domains andfacets in a large crosssectional sample. *Journal of Personality and Social Psychology, 100*(2), 330–348.

Specht, J., Egloff, B., & Schmukle, S. C. (2011). Stability and change of personality across the life course: The impact of age and major lifeevents on mean level and rank order stability of the Big Five*Journal of Personality and Social Psychology, 101*(4), 862–882.

Xu, Y. (2011). Psychological Kidnap: A New Model of Corruption Process in China. The 9th Biennial Conference of Asian Associationof Social Psychology, Kunming, China.

Yang, K. S., & Bond, M. H. (1990). Exploring implicit personality theorieswith indigenous or imported constructs: The Chinese case. *Journal of Personality and Social Psychology, 58* (6), 1087–1095.

Personality psychology：The study of human nature and its differences

GUO Yongyu, LI Jing, HU Xiaoyong

(School of Psychology, Central China Normal University 430079 Wuhan; Key Laboratory of Adolescent Cyberpsychology and Behavior, Ministry of Education, Wuhan 430079, China)

Abstract   Personality psychology is the discipline that explores the human nature from the integrality perspective in the psychology discipline system. It focuses on the human nature and its differences as its core, and views the man as integrality. Personality psychology is not only important in psychology discipline system but also in science of life, science of social and humanities about human beings. After one hundred years of tortuous development, personality psychology eventually comes into a new developing and prosperous period. In the last decades, the western personality psychology has jumped into a rapid development period. There are considerable advances in research paradigm, research methods and research contents, including personality structure, personality dynamics and personality development. Meanwhile, domestic scholars began to engage in researching the issues of the Chinese personality on the basis of importing and introducing western personality psychology. Currently, the research themes comprise personality and creativity, personality and personnel selection and resettlement, personality and corruption behavior, personality and violent crime, personality and disease, and the shaping of healthy personality in harmonious society with the Chinese cultural background. Considering the important scientific value of personality psychology and its contributions to the progress of society, we propose that the research of Chinese Personality Psychology should be enhanced in many aspects such as the study of Chinese personality in modernization context and the adopting of diversification of research methods in the future.

Keywords   personality psychology; human nature and its differences; modernization personality research

男性化与女性化对面孔偏好的影响
—— 基于图像处理技术和眼动的检验 *

温芳芳，佐　斌

（华中师范大学心理学院；青少年网络心理与行为教育部重点实验室，武汉 430079）

摘　要　采用图像处理技术和眼动探讨了性别二态线索对面孔偏好的影响。实验 1 发现非面孔线索未掩蔽和掩蔽时，感知男性化技术与原始照片条件下女性化的男性面孔更有吸引力和信任度；性别二态技术条件下，非面孔线索未掩蔽时男性化的男性面孔更有吸引力和信任度。实验 2 表明被试对男性面孔的平均瞳孔大小和注视次数均大于和多于女性面孔，首次注视时间短于女性面孔；被试对男性化面孔的首次注视时间和首次注视持续时间均长于女性化面孔。

关键词　面孔吸引力；性别二态；感知男性化；图像处理技术；眼动

分类号　B842

1　引言

面孔吸引力心理学作为一个研究主题，过去的几十年有着快速的发展，成为了文化人类学、进化心理学和社会心理学等领域广泛关注的焦点。面孔吸引力 (facial attractiveness) 是指目标人物面孔所诱发的一种积极愉悦情绪体验并驱使他人产生接近意愿的程度 (转引自李鸥，陈红，2010)。研究者从面孔吸引力与个体社会交往、择偶、求职、健康的关系等方面进行了探讨，发现面孔吸引力对个体社会生活的多个方面都有极其重要的影响 (Buckingham et al., 2006; 张妍等，2010)。男性和女性经过青春期第二性征的发展后逐步形成了体态上的性别二态性 (马华维，俞琴燕，陈浩，2007)。性别二态线索 (sexual dimorphism cues) 即男性化 ‒ 女性化线索 (Masculinity‒Femininity Cues)，是决定面孔吸引力的重要指标 (Enquist, Ghirlanda, Lundqvist, & Wachtmeister, 2002)。Rhodes (2006) 通过元分析表明性别二态线索对面孔吸引力会产生很大影响。

以往研究一致表明女性化的女性面孔更有吸引力。女性的吸引力几乎是 “女性化” 的代名词，因为吸引力能够预测判断一张女性面孔的性别所需的时间，同样也能预测判断此面孔女性化程度的时间 (Rhodes, 2006)。研究者发现被认为有吸引力的女性面孔，确实具有更多的女性化特征 (梁静，李红，2010)。当人们用计算机生成的更有吸引力的女性面孔同样被认为更女性化 (Johnston & Franklin, 1993)。

但女性偏好男性化还是女性化的男性面孔一直

存在争议，根据进化心理学中性选择的优良基因理论 (good genes theory of sexual selection)，女性偏好男性化的男性面孔。多元化策略权衡模型 (the trade off model of strategic pluralism) 表明女性对男性化面孔的偏好存在"好基因"收益 (benefits) 和"坏父亲"代价 (costs) 之间的权衡，并且文化的差异会对女性的权衡决策产生影响。大部分研究者从产生面孔刺激的方法 (性别二态技术 (sexual dimorphism method) 和感知男性化技术 (perceived masculinity method)) 差异的角度来解释女性对性别二态线索的男性面孔的不同偏好，从男性化和女性化人格维度上看，性别二态技术更像是单维观点，感知男性化技术则更像是双维观点。而性别二态技术和感知男性化技术等方法差异是否导致了女性对性别二态线索的男性面孔偏好的不一致？心理学界却对此一直存在争论。除此之外，考虑到以往关于性别二态线索对面孔吸引力影响的相关研究均让被试对配对呈现的面孔刺激进行吸引力评价，这些行为指标具有自我报告法的主观局限性。眼动作为一种新兴的重要的认知方法，可为认知加工过程提供即时的客观眼动指标。因此，本研究拟首次在中国的文化背景下，通过两个实验对此进行探讨，在实验一中采用性别二态和感知男性化两种计算机图像处理技术考察女性对性别二态线索的男性面孔的偏好，为方法差异之争提供中国的证据。在实验二中首次尝试采用眼动仪探究性别二态线索对面孔吸引力的影响，以期从客观因素的角度，一方面验证以往研究对女性化的女性面孔偏好的一致性，另一方面为性别二态线索对男性面孔吸引力的影响提供客观的眼动指标。

1.1 性选择的优良基因理论

根据进化心理学中性选择的优良基因理论，人们认为具有健康基因，较强免疫能力和生殖优势标志的面孔更有吸引力。男性化的男性面孔被认为是优良基因的指标，由于男性化的第二性特征是在青春期性激素的影响下发育的，因此可以作为可靠信号 (Thornhill & Gangestad, 1999)。男性化的面孔特征与男性长期的健康状况呈正相关，也可能标示着基因的健康，会产生更健康的后代，从而使女性在择偶中间接受益 (Fink

& Penton-Voak, 2002; Gangestad & Simpson, 2000)。可见，基于优良基因理论女性应该更偏好男性化的男性面孔 (Gangestad & Scheyd, 2005)，一些研究结果也支持了这一理论假设，即女性感知到男性化的男性面孔更有吸引力 (Johnston, Hagel, Franklin, Fink, & Grammer, 2001; Penton-Voak et al., 2001; DeBruine et al., 2006; DeBruine, Jones, Smith, & Little, 2010)。然而，另一些研究结果则与之相反，表明女性更偏好女性化的男性面孔 (Perrett et al., 1998; Rhodes, Hickford, & Jeffery, 2000; Little, Burt, Penton-Voak, & Perrett, 2001; Little & Hancock, 2002; Penton-Voak et al., 2003)。鉴于这些相矛盾的研究结果，优良基因理论似乎不能完全解释女性对男性化的偏好。

1.2 多元化策略权衡模型

多元化策略权衡模型表明，对男性化偏好的这种不一致性可能是由于女性对男性化偏好存在更优良健康基因和更低的预期亲代投资之间的权衡 (trade off) (Gangestad & Simpson, 2000)，也就是说，女性选择男性化所获得的优良基因的"收益" (benefits) 可能要以亲代投资的减少为"代价" (costs)。尽管男性化标示着基因的健康，但女性化的男性有更多的亲社会人格特质 (Perrett et al., 1998)。比方说，Perrett 等人 (1998) 发现男性化的男性面孔具有更低的诚信和热情，更可能成为坏父亲，更不愿意为亲代投资，男性化的男性可能会被感知为不可信任的，冷漠的 (Boothroyd, Jones, Burt, & Perrett, 2007; Oosterhof & Todorov, 2008)。与男性化的男性相比，女性化的男性则报告过去有更长期的浪漫关系 (Rhodes, Simmons, & Peters, 2005)。多元化策略权衡模型指出那些能改变女性选择男性伴侣的"代价"和"收益"的相对重要性的因素会影响女性偏好男性化和女性化的强度 (Gangestad & Scheyd, 2005)。

1.3 吸引力偏好的文化差异

根据前述理论，女性对男性化的偏好是优良基因线索和预期亲代投资线索两者间"权衡"的结果，但这种"权衡"会根据个人情况而改变，大规模的环境、文化和种族等因素更会影响"权衡"的决策。

Gangestad 和 Buss (1993) 根据 29 个不同文化背景的被试在择偶时对一系列属性重要性的等级评价发现，来自高病原体流行区的个体对身体吸引力的评价要高于较低区域个体，Low (1990) 研究发现在疾病更普遍的地区，女性更偏好标示着健康基因的男性化男性面孔 (DeBruine, Jones, Crawford, Welling, & Little, 2010)。生态和文化因素的方面也会影响女性对男性化偏好的"权衡"，Penton-Voak, Jacobson 和 Trivers (2004) 发现牙买加比英国的寄生负担 (parasite load) 高，医疗普及程度低，男性的亲代投资不明显，从而牙买加女性比英国女性更偏好男性化的男性面孔。Swami 和 Tovée (2005) 对英国和马来西亚男性身体吸引力的研究也发现了对男性化偏好存在文化差异。由此，文化和环境因素对女性男性化偏好的"权衡"决策的影响在东西方两种不同的文化背景下可能更加明显。中西文化在审美上的也存在很大差异，孙银玉 (2004) 指出中国文化在艺术追求上是以"韵"为主旨，而以古希腊为源头的西方则以"美"作为其最高审美范畴。如西方男性普遍追求强壮和肌肉型身体 (McCabe & Ricciardelli, 2001), Botta (2003) 通过对美国大学生的调查发现，随着影视杂志等媒体的曝光，人们更容易致力于成为"肌肉男"。与西方男性以肌肉为美相比，中国文化则更含蓄些，西方学者在对文化的划分中提出过东方文化属于"女性化"文化，与重视成功和成就的"男性化"文化不同，东方文化更注重关系与和谐 (转引自王登峰，崔红，2007)。并且，中国为典型的长期取向文化社会，即着眼于未来，强调长期性的承诺，尊重传统，推崇节俭和持久力，倾向于做长期规划和投入，相信这样虽有风险但最终会带来丰厚回报 (Hofstede, 2008)。当前大多数关于面孔吸引力的研究都在西方文化背景下进行，因此，探索中国女性对性别二态线索的男性面孔的偏好将是一个有趣的问题，并且根据中国文化"女性化"和"长期取向"的特点，中国女性在考虑"收益"和"代价"的"权衡"时会更看重长远的亲代投资，从而偏好女性化的男性面孔。

1.4 性别角色理论对方法差异的解释

关于女性对男性化偏好的不一致结果更多的研究主要从产生面孔刺激的方法差异角度进行解释，这可用性别角色理论来理解。以往研究一般采用基于原型的计算机图像处理技术 (性别二态技术和感知男性化技术) 来操作男性化 – 女性化维度 (Tiddeman, Perrett, & Burt, 2001)。性别二态技术最初由 Perrett 等人 (1998) 使用，具体来讲，首先分别对一组男性和一组女性面孔照片进行平均化处理，得到男性和女性的平均化面孔图像原型，然后采用夸大或减少男性和女性的平均化面孔图像原型形状上的差异来操作男性化 – 女性化的面孔刺激，女性化的男性面孔图像是通过将男性的平均化面孔图像原型朝向女性的平均化面孔图像原型扭曲 50% 而获得，而男性化的男性面孔图像则是通过将男性的平均化面孔图像原型偏离女性的平均化面孔图像原型扭曲 50% 而得到 (Perrett et al., 1998; Rhodes et al., 2000)。感知男性化技术是由 Johnston 等人 (2001) 对性别二态技术进行的修正，首先让被试对一组男性面孔照片进行男性化程度的等级评价，然后分别将被试感知到的高男性化与低男性化照片进行平均化处理，从而得到男性化和女性化两种男性面孔图像刺激。对比两种获取男性化 – 女性化面孔图像刺激的不同操作技术原理可知，性别二态技术将男性化与女性化看作一个连续体对立的两端，与传统的性别角色理论一致，该理论认为性别角色概念是一维的结构，男性特质和女性特质分别位于同一维度的两端，每个个体均处于该维度的某一点上 (蔡华俭，杨治良，2002)。感知男性化技术将男性化特质与女性化特质看作是两个不同的维度，与 Constantinople (1973) 率先从理论层面提出的双性化理论 (男性特质和女性特质分别是独立的单一维度) 相一致 (徐振华等，2010), 也符合 Bem (1974) 提出的性别角色观点和钱铭怡、张光健、罗珊红和张莘 (2000) 研究所支持的性别图式双维说。因此，从男性化与女性化人格维度上看，性别二态技术更像是单维观点，符合传统的性别角色模式理论，而感知男性化技术则与双性化理论相一致，更像是双维观点。

1.5 以往研究争论的焦点

目前关于面孔偏好的争论焦点是采用单维性别

二态技术和双维感知男性化技术产生的面孔刺激是否会影响女性对性别二态线索的男性面孔的偏好。DeBruine 等人 (2006) 首次将两种技术进行了比较,结果发现不管采用哪种技术都表明男性化的男性面孔更有吸引力。Rennels, Bronstad 和 Langlois (2008) 则同样采用这两种技术发现了不一致的结论,性别二态技术条件下,更偏好女性化的男性面孔,而在感知男性化技术条件下,认为男性化的男性面孔更有吸引力。同时让被试对未经改变的原始照片分别进行男性化和吸引力评价,发现男性化与吸引力评价之间呈显著正相关,这与感知男性化技术条件下的结果一致,而与性别二态技术条件的研究结果相反。Rhodes (2006)采用元分析也得出了同样的结论,即采用性别二态技术操作的面孔刺激的研究,男性化与吸引力评价之间呈中等的负相关,而采用未经改变的原始照片的研究表明男性化与吸引力的评价之间呈中等的正相关。然而,DeBruine 等人 (2010) 认为 Rennels 等人所采用的配对面孔刺激材料存在非面孔线索 (如发型等) 的差异,从而可能对男性面孔的偏好造成干扰,并通过两个实验进行检验。实验 1 采用 Rennels 等人 (2008) 的面孔图像材料并未对非面孔线索进行掩蔽时,实验结果验证了 Rennels 等人的主要结论,即不同的技术导致女性偏好不同的男性面孔;当非面孔线索掩蔽时,感知男性化技术条件下,女性由非面孔线索未掩蔽时偏好男性化面孔转向偏好女性化面孔,而性别二态技术条件则没有改变。实验 2 则采用新材料结果发现,非面孔线索未掩蔽时,感知男性化技术条件下,女性偏好男性化面孔,性别二态技术条件下,对男性化的

偏好不存在显著差异;当非面孔线索掩蔽时,感知男性化技术和性别二态技术条件对男性化的偏好都不显著,非面孔线索的掩蔽对两种技术都产生了影响。并且实验 1 和 2 都表明当非面孔线索掩蔽时,两种技术条件所产生的男性化偏好的差异都变得不显著,从而认为两种技术条件所产生的不同结果应归因于非面孔线索 (如发型等) 的作用 (如表 1 所示)。

由表 1 可知,当非面孔线索未掩蔽时,以往研究采用双维的感知男性化技术都产生了一致的结果,即偏好男性化的男性面孔,与 Rhodes (2006) 元分析中采用原始照片的结果一致,支持了西方性别角色理论将男性特质和女性特质作为独立维度的男性化模型,认为男性化成分在双性化与心理健康的关系上做主要贡献,符合西方的"男性化"文化。但是,采用单维的性别二态技术产生的结果与双维的感知男性化技术相比却存在争议,DeBruine 等人 (2006) 和 DeBruine 等人 (2010) (非面孔线索掩蔽时) 都认为两种技术产生一致的结果,而 Rennels 等人 (2008) 和 DeBruine 等人 (2010) (非面孔线索未掩蔽时) 则发现两种技术产生的结果相反。那么采用单维的性别二态技术和双维的感知男性化技术产生的面孔刺激是否会影响女性对男性化面孔偏好呢?采用双维的感知男性化技术在中国"女性化"文化背景下是否会产生相反的结果呢?这些问题都值得研究。另外,当非面孔线索掩蔽时,DeBruine 等人 (2010) 在两个实验的结果也存在不一致性,非面孔线索的掩蔽是否对两种技术都会造成混淆值得进一步探究。

此外,信任度评价与吸引力判断之间呈高相关

表 1 以往采用两种技术的研究对男性化和女性化的男性面孔吸引力偏好情况

作者及年份	实验	非面孔线索 (是否掩蔽)	性别二态技术		感知男性化技术	
			男性化	女性化	男性化	女性化
DeBruine 等 (2006)		否	√		√	
Rennels 等 (2008)		否		√	√	
DeBruine 等 (2010)	1	否		√	√	
		是		√		√
	2	否	√ ns		√	
		是		√ ns		√ ns

注:"√"被试认为对应面孔更有吸引力而且差异显著;"√ ns"被试认为对应面孔更有吸引力但是差异不显著。

(Oosterhof & Todorov, 2008; Theodoridou, Rowe, Penton–Voak, & Rogers, 2009)。对他人的信任是成功交往的一个重要因素，信任度是一种社会态度，通常可仅通过观察面孔而得出 (Zaidel, Bava, & Reis, 2003)。根据晕轮效应 (The Halo Effect)，人们往往会自动地将积极的人格特质 (如，信任度等) 归于有吸引力的个体 (Dion, Berscheid, & Walster, 1972)。在此，本研究在考察面孔吸引力的同时考察面孔信任度主要基于三方面的考虑：首先，信任度感知能很好地对面孔的总体效价进行近似评估，信任度与其他的积极特质 (如，责任感、情绪的稳定性、社交性、细心、聪明、自信等) 存在高相关 (Oosterhof & Todorov, 2008)；其次，信任度感知是社会交流互动中的一种门控机制 (Gating Mechanism)，是社会交往中确定接近还是回避 (approach/avoidance) 基本反应的最重要决定因素之一 (Todorov, 2008)；最后，信任度感知比其他特质判断更有效率，在面孔呈现仅 100ms 时就能做出判断，研究表明信任度感知与吸引力的判断同样有效率 (Willis & Todorov, 2006)。

鉴于以往关于女性对男性化面孔偏好研究的争论，并且进一步考虑到以往关于男性化与女性化对面孔吸引力影响的研究均采用的自我报告法，存在一定的主观局限性，本研究拟在此通过两个实验共同探讨性别二态线索对面孔偏好的影响。实验 1 采用性别二态和感知男性化两种计算机图像处理技术，考察女性对性别二态线索的男性面孔的吸引力和信任度评价。根据中国"女性化"和"长期取向"的文化特点和已有研究结果提出以下假设：中国女性对原始照片的男性化与吸引力评价呈显著负相关；感知男性化技术条件与对原始照片情况下女性均偏好女性化的男性面孔；性别二态技术条件下，女性偏好男性化的男性面孔。非面孔线索对两种技术条件下的被试的面孔偏好都会产生干扰。实验 2 则尝试结合眼动仪考察性别二态线索对面孔吸引力的影响，一方面验证以往研究中女性化女性面孔更有吸引力的结论，另一方面通过客观指标的方法为女性对男性化面孔争议的研究提供新的思路，在一定程度上弥补自我报告法的局限性。

2　实验 1

2.1　研究方法

2.1.1　被试　53 名女大学生参与本实验，年龄范围在 19~22 岁 (M=19.89, SD=0.93)，所有被试都报告为异性恋者。被试获得公共心理学课程的学分。

2.1.2　刺激　采用与 Rennels 等人 (2008) 和 DeBruine 等人 (2010) 相似的程序，分别采用性别二态技术和感知男性化技术两种方法得到男性化、平均化和女性化的六种合成的男性面孔刺激版本。

实验所选面孔刺激材料来自某高校毕业生信息采集登记照片数据库，所有照片是同一批在相同背景，标准光线条件下拍摄的正面像，共 321 张，其中，男 144 张，女 177 张。从 144 张男性照片中剔除佩戴眼镜，有胡须，面带表情的部分，共筛选出 60 张男性彩色登记照，然后从 60 张 22~26 岁 (M=23.70, SD=0.83) 的男性的彩色照片中随机挑选 32 张作为实验材料，同样，从 177 张女性照片中挑选出 32 张 (M=23.56, SD=0.84)。所有目标照片均显示中性面部表情，未戴眼镜，无配饰，无胡须。挑选出的各 32 张男性和女性面孔照片将分别作为后面采用计算机图像处理技术处理的面孔平均合成材料。

计算机图像处理技术的操作过程如下：采用 FantaMorph 4.0 软件，分别将以上所准备的各 32 张男、女面孔照片合成男性和女性的平均化面孔图像原型，用鼠标在每张面孔上标注出 179 个关键点，这些关键点标识出了面孔的形状、内部特征的位置和轮廓 (如图 1 所示)。具体来讲，先在一张面孔照片上标注出这些关键点，同时将相对应的关键点在另一张面孔照片的相应位置标注出来，从而将两张照片上关键点的数值 (像素) 平均化而产生一张由 2 张面孔合成的图像，然后采用同样标注关键点的方法将这 2 张面孔合成图像与另外的 2 张面孔合成图像进一步合成，从而产生一张由 4 张面孔合成的图像，依次类推，进而是 8 张，16 张，最终得到 32 张面孔的合成图像，采用此方法最

终得到 2 张男性和女性的平均化面孔图像原型。

　　然后，采用 Perrett 等人 (1998) 所发展的性别二态技术分别将 2 张平均化面孔图像原型照片进行男性化 – 女性化处理。基本原理是通过夸大或减少男性与女性的平均化面孔图像原型之间的形状差异来分别创建男性化和女性化的男性面孔。操作过程在阿伯丁大学心理学院 DeBruine 等人创建的网站 www.faceresearch.org 上实现。将男性和女性的平均化面孔图像原型上传到网站的处理程序接口，然后采用性别二态技术将照片进行处理。男性化的男性面孔图像是通过夸大男性与女性的平均化面孔图像原型的形状差异，将男性的平均化面孔图像原型偏离女性的平均化面孔图像原型方向扭曲 50% 而创建，女性化的男性面孔图像则是通过减少男性和女性的平均化面孔图像原型的形状差异，将男性的平均化面孔图像原型朝向女性平均化的面孔图像原型方向扭曲 50% 得到 (Perrett et al., 1998; Rhodes et al., 2000)，最终得到经过性别二态技术操作的男性化和女性化的男性面孔图像刺激 (见图 2 左上)。

　　由 26 名 18~31 岁 (M=23.54, SD=3.99) 的女性被试对最初筛选出的 60 张男性照片进行男性化特质
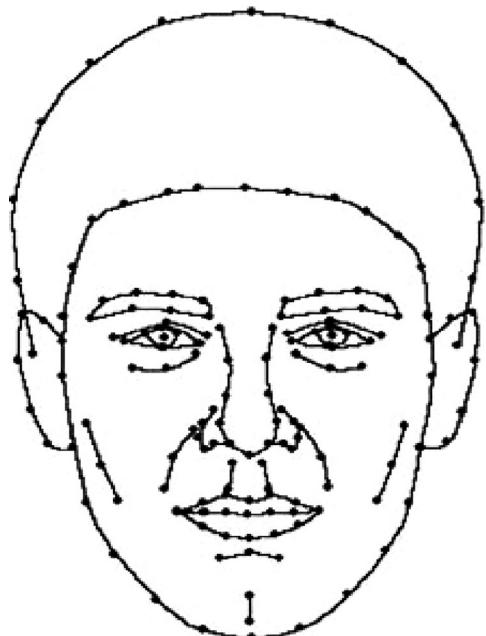
图 1　创建平均化面孔图像原型所采用过的关键点 (Rhodes, 2000, 2006)

的 7 分等级评价，其中，1 表示非常不男性化，7 表示非常男性化。将所有个体非面孔线索掩蔽，排除了衣着，发型等非面孔线索的干扰。等级评价范围从 2.23 到 5.81(M=4.42, SD=0.76)，评价结果具有很高的信度 (α =0.91)。14 张最男性化面孔的等级评价的平均值为 4.37, 14 张最不男性化的等级评价平均值为 3.33。由另外不同的 35 名 18~32 岁之间 (M= 23.34, SD=3.69) 的女性被试对上述照片进行吸引力 7 分等级评定，其中，1 表示非常没有吸引力，7 表示非常具有吸引力。吸引力评价具有很高的信度 (α = 0.93)。男性化评价与吸引力评价之间呈显著的负相关 (r = –0.36, p<0.01)，与 Rennels 等人 (2008) 和 DeBruine 等人 (2010) 的结果相反。

　　感知男性化技术是将前面男性化等级评价的 60 张照片中得分最高部分 (高男性化) 或最低部分 (低男性化) 的面孔照片平均化来分别创建男性化和女性化的男性面孔。与 DeBruine 等人 (2010) 相似，选取 60 张男性面孔中评价高男性化的 14 张照片 (顶端的 23.3%) 合成男性化的男性面孔图像，同样选取评价低男性化的 14 张照片 (底端的 23.3%) 合成女性化的男性面孔图像。平均化的男性面孔是将低男性化的 14 张照片中较高的一半 (7 张) 与高男性化的 14 张图像中较低的一半 (7 张) 合成创建平均化的男性面孔图像。合成同样是用 FantaMorph 4.0 软件，采用与 Rennels 等相似的程序，最终得到感知男性化技术条件下的男性化、平均化和女性化的男性面孔图像刺激 (见图 2 右上)。所有面孔图像均标准化，图像大小 400×300 像素，蓝色背景，沿下巴和发线切出面孔的形状，得到将非面孔线索掩蔽的图像 (见图 2 下)。

　　2.1.3　程序　采用组内设计，被试分别完成两个简短的吸引力和信任度评价的面孔感知实验，这两个实验的顺序是随机呈现的。在吸引力评价实验中，被试评价 12 对随机程序的面孔，从中选择更有吸引力的面孔。每组配对中的面孔刺激都来自以下四种条件的一种：未经掩蔽的感知男性化技术，掩蔽的感知男性化技术，未经掩蔽的性别二态技术和掩蔽的性别二态技术。每个条件均显示 3 类配对：男性化 vs. 平均
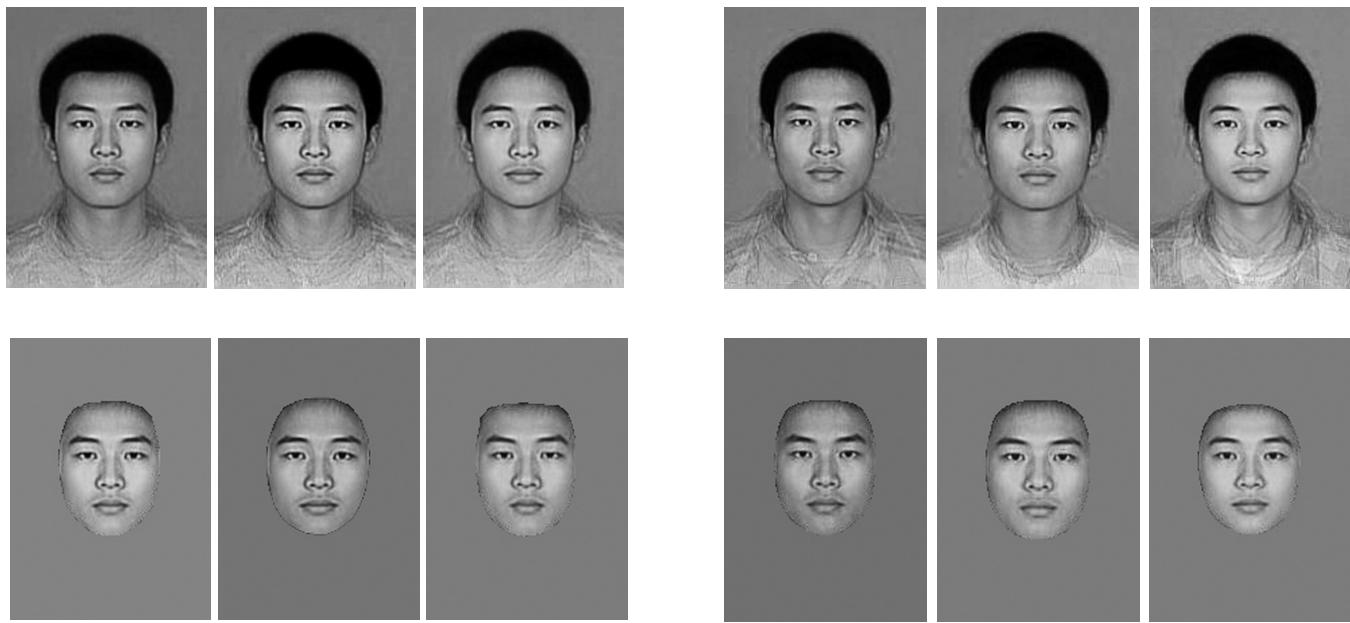
图 2　性别二态技术条件和感知男性化技术条件下的面孔图像刺激

注：左半部分为性别二态技术条件下的男性面孔，依次为男性化，平均化和女性化的面孔图像刺激；右半部分为感知男性化技术条件下的男性面孔，依次为男性化，平均化和女性化的面孔图像刺激。上半部分为非面孔线索（如发型等）未经掩蔽时的面孔图像刺激，下半部分为非面孔线索掩蔽时的面孔图像刺激。

化，男性化 vs. 女性化，平均化 vs. 女性化，即共产生 12 组配对面孔。在信任度评价实验中，被试评价 18 对随机呈现的面孔，其中除了包括吸引力评价实验中相同的 12 对照片外，增加了原始照片条件（照片未经任何处理），即从前面的 60 张男性照片中挑出三张原始照片，分别是评价最高男性化，最低男性化和中等程度男性化的照片。也就是说在信任度评价实验中增加了非面孔线索未经掩蔽和掩蔽的原始照片条件，然后让被试从随机呈现的配对照片中选择更有信任度的面孔。通过随机呈现配对照片在屏幕上呈现的左右位置和顺序，平衡了位置和顺序效应。

2.2　结果

2.2.1　吸引力评价　被试分别对性别二态技术和感知男性化技术两种条件下，同时考虑面孔刺激非面孔线索未经过掩蔽和掩蔽两种情况下的男性面孔吸引力进行评价，通过组合共产生四种操作条件，分别是性别二态技术条件下非面孔线索未掩蔽、感知男性化条件下非面孔线索未掩蔽、性别二态技术条件下非面孔线索掩蔽和感知男性化技术条件下非面孔线索掩蔽。四种技术条件下被试对性别二态线索的

男性面孔的吸引力选择频率如表 2 所示。评价结果记录方法如下，如果被试在平均化和男性化面孔配对中的选择平均化面孔，结果记录选择的是更女性化面孔，如果被试在平均化和女性化面孔配对中选择平均化面孔，则记录选择的是更男性化面孔。

当非面孔线索未掩蔽时，经 Wilcoxon 带符号秩检验发现，在性别二态技术条件比感知男性化技术条件下被试认为男性化更有吸引力 (Z=−5.18, p<0.001)。同样，当面孔掩蔽时，经 Wilcoxon 带符号秩检验发现，在性别二态技术条件比感知男性化技术条件认为男性化更有吸引力 (Z=−4.32, p<0.001)。

此外，在感知男性化技术条件下，非面孔线索掩蔽比未掩蔽时被试认为女性化的男性面孔更有吸引力 (Z=−2.51, p<0.05)。同样，在性别二态技术条件下，非面孔线索未掩蔽时比掩蔽时被试认为男性化的男性面孔更有吸引力 (Z=−3.34, p<0.01)。

经过二项式分析发现，当非面孔线索未掩蔽时，在感知男性化技术条件下被试认为女性化的男性面孔更有吸引力 (p<0.001)；而在性别二态技术条件下被试则认为男性化的男性面孔更有吸引力 (p< 0.001)。

表 2　被试在四种操作条件下对性别二态线索的男性面孔吸引力的选择频率

条件	选择	性别二态技术条件		感知男性化技术条件	
		人数 (n)	频率 (%)	人数 (n)	频率 (%)
非面孔线索未掩蔽	男性化	111	69.8	52	32.7
	女性化	48	30.2	107	67.3
非面孔线索掩蔽	男性化	76	47.8	29	18.2
	女性化	83	52.2	130	81.8

当非面孔线索掩蔽时，在感知男性化技术条件下被试同样认为女性化的男性面孔更有吸引力 ($p<0.001$)；而在性别二态技术条件下两者不存在显著的差异 ($p=0.58$)。

2.2.2 信任度评价　被试分别对性别二态技术条件、感知男性化技术条件和原始照片三种条件下，同时考虑面孔刺激非面孔线索未经过掩蔽和掩蔽两种情况下的男性面孔信任度进行评价，通过组合共产生六种操作条件，分别是性别二态技术条件非面孔线索未掩蔽、感知男性化技术条件非面孔线索未掩蔽、原始照片条件非面孔线索未掩蔽、性别二态技术条件非面孔线索掩蔽、感知男性化技术条件非面孔线索掩蔽和原始照片条件非面孔线索掩蔽。六种操作条件下被试对性别二态线索的男性面孔的信任度选择频率如表 3 所示。评价结果记录方法如下，如果被试在平均化和男性化面孔配对中选择平均化面孔，结果记录选择的是更女性化面孔，如果被试在平均化和女性化面孔配对中选择平均化面孔，则记录选择的是

更男性化面孔。

当非面孔线索未掩蔽时，经 Wilcoxon 带符号秩检验发现，在性别二态技术条件比感知男性化技术条件下被试认为男性化更有信任度 ($Z=-2.48, p<0.05$)；性别二态技术条件与原始照片条件下被试对男性化信任度的选择存在显著差异 ($Z=-3.04, p<0.01$)；而感知男性化技术条件与原始照片条件下被试对男性化的信任度的评价不存在显著差异 ($Z=-0.82, p=0.41$)。

同样，当非面孔线索掩蔽时，经 Wilcoxon 带符号秩检验发现，在性别二态技术条件比感知男性化技术条件下被试认为男性化更有信任度 ($Z=-3.70, p<0.001$)；性别二态技术条件与原始照片条件下被试对男性化信任度的选择存在显著差异 ($Z=-3.25, p<0.01$)；而在感知男性化技术条件与原始照片条件下被试对男性化的信任度评价不存在显著差异 ($Z=-0.11, p=0.91$)。

此外，在感知男性化技术条件下，非面孔线索掩蔽时与未掩蔽时被试对男性面孔信任度的选择不存在显著差异 ($Z=-1.58, p=0.11$)；在性别二态技术条件下，

表 3　被试在六种操作条件下对性别二态线索的男性面孔信任度的选择频率

条件	选择	性别二态技术条件		感知男性化技术条件		原始照片条件	
		人数 (n)	频率 (%)	人数 (n)	频率 (%)	人数 (n)	频率 (%)
面孔未掩蔽	男性化	89	56.0	65	40.9	58	36.5
	女性化	70	44.0	94	59.1	101	63.5
面孔掩蔽	男性化	86	54.1	53	33.3	52	32.7
	女性化	73	45.9	106	66.7	107	67.3

非面孔线索未掩蔽时与掩蔽时被试对信任度的偏好也不存在显著差异 ($Z=-0.43, p=0.67$)；在原始照片条件下，非面孔线索未掩蔽时与掩蔽时被试对信任度的偏好也不存在显著差异 ($Z=-0.84, p=0.40$)。

经过二项式分析发现，当非面孔线索未掩蔽时，在感知男性化技术条件下被试认为女性化的男性面孔更有信任度 ($p=0.05$, 边缘显著)；在性别二态技术条

件下被试则认为男性化的男性面孔的信任度评价边缘显著高于女性化的男性面孔 ($p=0.098$, 边缘显著)；在原始照片条件下被试认为女性化的男性面孔更有信任度 ($p<0.01$)。当非面孔线索掩蔽时，在感知男性化技术条件下被试同样认为女性化的男性面孔更有信任度 ($p<0.05$)；在原始照片条件下被试同样认为女性化的男性面孔更有信任度 ($p<0.01$)；在性别二态技

术条件下两者不存在显著的差异 ($p=0.41$)。

3 实验 2

3.1 研究方法

3.1.1 被试 某高校 53 名学生，其中，男 19 名，女 34 名，年龄范围在 18~28 岁 ($M=22.66$, $SD=2.16$)。无任何心理疾病史，均为右利手，其裸视力或矫正视力正常，均无色盲或色弱，无眼动实验经验，且均自愿参加，实验结束均得到一份小礼品。

3.1.2 实验仪器 Eyelink1000 型眼动仪一台 (SR Research, Mississauga, Ontario, Canada)，采样率 (Sample Rate) 为 1000 Hz，噪音小于 0.022，被试下额固定在下巴托上，校准模式是 Hv9，即 9 个栅格点校准。正式实验程序采用 Eprime 软件编程，实验显示器为 17 英寸，分辨率为 1024×768 像素，刷新率为 100 Hz，图片大小 400×300 像素，均为单眼记录。

3.1.3 实验变量及设计 采用 2×2 的组内实验设计，其中，两个组内变量分别为目标性别 (男 vs. 女) 和性别二态线索 (男性化 vs. 女性化)。因变量包括行为指标和眼动指标，其中，行为指标为被试对配对呈现照片吸引力评价的按键反应，眼动指标包括瞳孔直径 (pupil size)，首次注视时间 (first fixation time)，首次注视持续时间 (first fixation duration) 和注视次数 (fixation count) 等。

3.1.4 实验刺激材料 从实验 1 相同的数据库和筛选方法，挑选出 40 张原始照片，采用实验 1 中的性别二态技术将这些原始照片 (男女各半) 分别创建出男性化与女性化图像，同时包括实验 1 中采用的性别二态条件下非面孔线索未掩蔽的刺激材料 3 对男性原型面孔图像 (男性化 vs. 平均化，男性化 vs. 女性化，平均化 vs. 女性化)，采用同样的方法得到 3 对女性原型面孔图像。最终得到 46 对经性别二态技术所得到的男性化和女性化的面孔图像刺激。

首先需要对所得的面孔刺激材料进行操作性检验，挑选出性别二态线索显著的面孔图片。具体方法如下：将得到的 46 对男性化和女性化的面孔图片随机排序，并且平衡男性化与女性化照片所在的左右位置，在计算机上用 Powerpoint 软件放映呈现。73 名大学生被试进行男性化评价，其中，男 35 人，女 38 人，平均年龄 22.56 岁，标准差 2.19。如果被试认为左边的照片更男性化，则选择 1, 右边的照片更男性化，则选择 2。对结果进行二项式统计分析，选取配对呈现时男性化评定显著的 20 对照片 (男女各半) 作为正

表 4 确定为正式实验面孔刺激的男性化评价情况

图片编号	1	2	3	6	9	10	11	15	17	18	19	24	27	29	30	31	35	37	38	40	41	42
目标性别	男	女	女	女	女	男	男	男	女	女	女	女	男	男	女	男	女	男	男	女	男	男
Mn	54	54	52	49	53	50	55	46	54	56	49	55	46	47	51	54	51	54	50	47	67	66
Fn	19	19	21	24	20	23	18	27	19	17	24	18	27	26	22	19	22	19	23	26	6	7
p	***	***	***	***	**	**	***	*	***	***	**	***	*	*	**	***	**	***	**	*	***	***

注：Mn 为男性化刺激被选人数，Fn 为女性化刺激被选人数。*$p<0.05$, **$p<0.01$, ***$p<0.001$

式实验材料。所选择的男性化评定的具体情况如表 4 所示。最终确定用于实验的材料为 20 对所挑选的原始照片经性别二态技术转化所得的刺激材料，和 6 对原型图像经性别二态技术所得的刺激材料 (6 对原型面孔图像的男性化评价也均显著，$p<0.01$)。

3.1.5 实验程序 实验地点在华中师范大学心理学院眼动实验室，环境安静，无干扰，由笔者和眼动实验室的专业人员作为主试，对被试进行单独施测。实验分为两个阶段：

实验准备阶段：被试坐在距离显示器 65 cm 处的被试椅上，下额放在 U 型托上，固定头部，确保实验数据的准确性。首先，主试先告知被试指导语："一会儿，请注视屏幕中央出现的注视点，之后屏幕上将会配对出现两张面孔图片，这 2 张面孔图片之间很相似，只有非常细微的差别，请仔细观察，选择您认为更有吸引力的那个面孔。认为左边面孔更有吸引力按电脑的 C 键，右边面孔更有吸引力则按电脑的 M 键。如果您已理解以上说明，请按 Y 键开始"，使其明确

任务和要求。再进行眼动仪的校正调适，采用九点进行视线追踪系统的校正 (Calibra-tion)。

刺激呈现阶段：利用 Eprime 2.0 软件编程，连接在眼动仪记录程序，26 对配对出现的男女面孔刺激以随机的方式呈现，平衡了男性化与女性化的位置和不同性别面孔刺激的呈现顺序。实验流程如图 3 所示，注视点持续时间为 1s，每张配对出现的图片在屏幕上持续时间为 5s，刺激呈现时，被试需要做出的按键反应，按键后，图片自动消失，紧接着出现一张空屏，持续时间为 1s。实验正式开始前，让被试左手放在 C 键位置，右手放在 M 键位置，做好按键准备，避免实验中被试为寻找按键头动对实验结果的精确性造成干扰。

3.2 结果

3.2.1 面孔吸引力评价行为指标分析 将被试对男性化和女性化的男女面孔刺激的吸引力评价选择频率经过二项式分析发现，对于男性面孔刺激，被试认为男性化的男性面孔吸引力 (N=26) 显著高于女性化 (N=7)，$p<0.01$；而对于女性面孔刺激，被试则认为女性化的女性面孔吸引力 (N=28) 显著高于男性化 (N=10)，$p<0.01$。

3.2.2 整体的眼动指标分析 在实验过程中，删除由于眼睛疲劳或其它原因，使眼动仪无法记录到数据或不准确数据记录，通过筛选，统计中的各项眼动数据来自 49 个被试。用 EDF Data Viewer 对数据进行处理分析后，用 SPSS 16.0 对数据进行统计处理。

对被试判断不同性别二态线索的男女面孔吸引力时的平均瞳孔大小、首次注视时间、首次注视持续时间和注视次数等指标进行比较，见表 5。
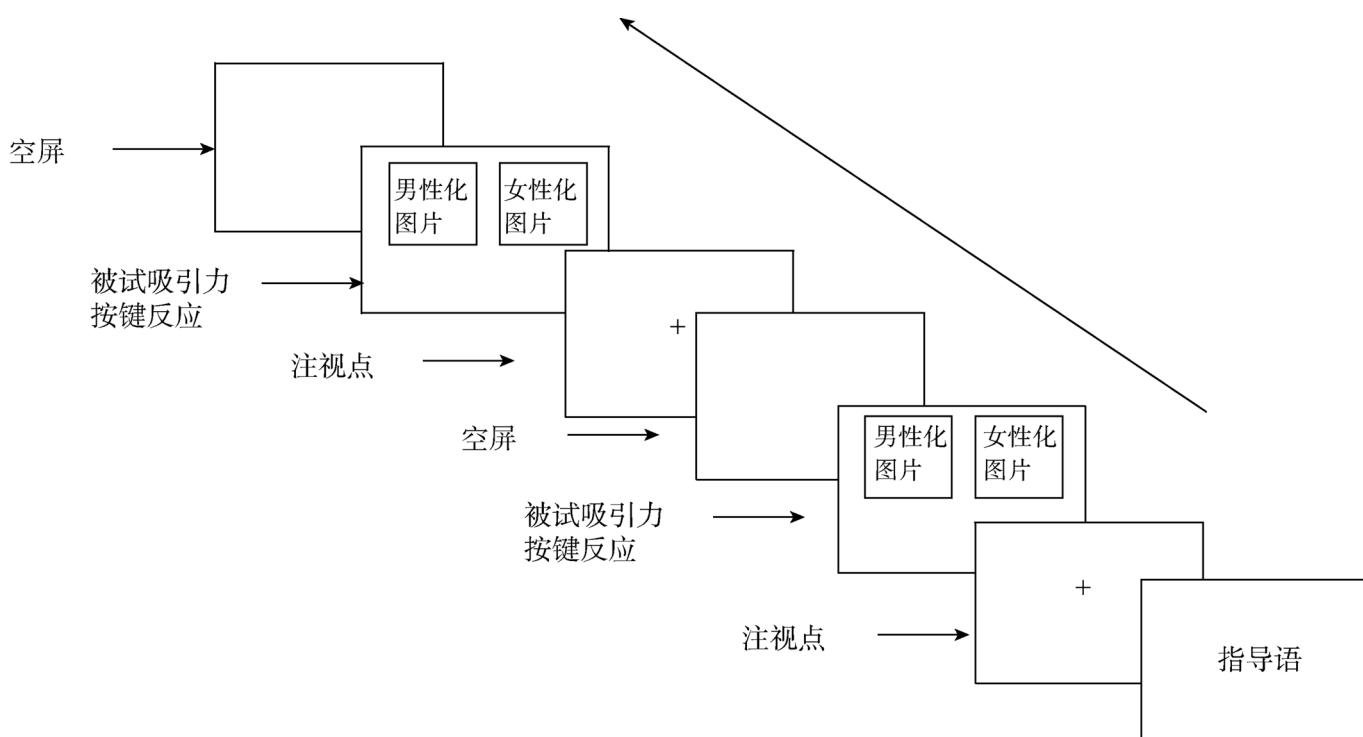
图 3 实验流程图

表 5 被试判断不同性别二态线索面孔吸引力的主要眼动指标（M±SD）

目标性别	性别二态线索	平均瞳孔大小	首次注视时间	首次注视持续时间	注视次数
男	男性化	615.94 ± 149.14	1607.17 ± 92.58	258.79 ± 58.57	4.72 ± 1.64
	女性化	616.43 ± 146.14	1605.64 ± 132.66	249.20 ± 43.60	4.71 ± 1.58
女	男性化	607.73 ± 142.60	1629.51 ± 148.85	266.65 ± 52.99	4.45 ± 1.22
	女性化	606.40 ± 147.42	1586.06 ± 96.41	252.01 ± 56.33	4.68 ± 1.40

（1）平均瞳孔大小

平均瞳孔大小的重复测量方差分析结果表明，平均瞳孔大小的目标性别变量的主效应显著 $F(1,48)=11.17$, $p < 0.01$，具体为被试在性别二态线索条件下对男性面孔吸引力的平均瞳孔大小 (M=(615.94+616.43)/2=616.19) 显著大于女性面孔 (M=(607.73+606.40)/2=607.06)。瞳孔大小的性别二态线索变量的主效应不显著，$F(1,48)=0.22$, p=0.64；目标性别变量和性别二态线索变量两项交互作用不显著 $F(1,48)=0.41$, p=0.53。

（2）首次注视时间

首次注视时间的重复测量方差分析结果表明，首次注视时间的性别二态线索变量的主效应显著 ($F(1,48)=6.42$, p=0.02)，被试对男性化面孔的平均首次注视时间 (M=(1607.17+1629.51)/2=1618.34) 显著长于对女性化面孔的平均首次注视时间 (M=(1605.64+1586.06)/2=1595.85)，首次注视时间的目标性别变量的主效应不显著 $F(1,48)=0.03$, p=0.86。目标性别变量与性别二态线索变量交互作用不显著，$F(1,48)=1.36$, p=0.25。

（3）首次注视持续时间

首次注视持续时间的重复测量方差分析结果表明，首次注视持续时间的性别二态线索变量的主效应显著 $F(1,48)=7.86$, p=0.01，即被试对男性化面孔的首次注视持续时间 (M=(258.79+266.65)/2= 262.72) 显著长于女性化面孔 (M=(249.20+252.01)/2= 250.61)。首次注视时间的目标性别变量的主效应边缘显著，$F(1,48)=3.01$, p=0.09，即被试对女性面孔的首次注视持续时间 (M=(266.65+252.01)/2=259.33) 边缘显著长于男性面孔 (M=(258.79+249.20)/2= 253.99)。目标性别变量与性别二态线索变量交互作用不显著，$F(1,48)=0.26$, p=0.62。

（4）注视次数

注视次数重复测量方差分析结果表明，注视次数的目标性别变量主效应边缘显著，$F(1,48)=3.20$, p=0.08，即被试对男性面孔的注视次数 (M=(4.72+4.71)/2=4.72) 边缘显著多于女性面孔 (M=(4.45+4.68)/2=4.56)。注视次数的性别二态线索变量的主效应不显著，($F(1,48)=2.37$, p=0.13)。注视次数的目标性别变量和性别二态线索变量交互作用不显著 $F(1,48)=2.30$, p=0.14。

4 讨论

4.1 基于图像处理技术对性别二态线索的男性面孔偏好分析

4.1.1 对性别二态线索的男性面孔的吸引力偏好分析 选择面孔刺激时，从对原始照片的男性化与吸引力等级评价的显著负相关结果，可以看出，女性认为女性化的男性面孔更有吸引力。当将原始照片分别经过性别二态和感知男性化两种技术处理后，在非面孔线索未掩蔽的情况下，感知男性化技术条件的吸引力评价结果与选择面孔刺激时对原始照片吸引力等级评价一致，性别二态技术条件的结果则相反，这与 Rennels 等人 (2008) 和 DeBruine 等人 (2010) 存在一致性。但本研究中，感知男性化技术条件和原始照片情况下，女性偏好女性化的男性面孔，性别二态技术条件下，女性则认为男性化的男性面孔更有吸引力，这些发现与 Rennels 等人和 DeBruine 等人的结果不同，他们发现在感知男性化技术条件和原始照片情况下，女性偏好男性化的男性面孔，而在性别二态技术条件下则是女性化的男性面孔更有吸引力。

在非面孔线索掩蔽的情况下，本研究发现，在感知男性化技术条件下同样更多选择女性化的男性面孔更有吸引力；在性别二态技术条件下则由未掩蔽时偏好男性化转为不存在显著偏好差异，本研究中非面孔线索是否掩蔽对性别二态技术和感知男性化技术两种条件都会造成干扰，这与 DeBruine 等人 (2010) 的结果中非面孔线索造成干扰部分一致。在 DeBruine 等人的研究中，实验 1 发现非面孔线索掩蔽主要在感知男性化技术条件产生作用，即由非面孔线索未掩蔽时的男性化偏好转向掩蔽时的女性化偏好，而非面孔线索掩蔽对性别二态技术条件未产生影响；实验 2 中则发现非面孔线索对感知男性化技术条件和性别二

态技术条件都产生了干扰，即在两种条件下，非面孔线索未掩蔽时比掩蔽时都更偏好男性化。可以看出，本研究中非面孔线索的作用与 DeBruine 等人实验 2 的结果一致，即非面孔线索的掩蔽在性别二态技术和感知男性化技术两种条件下都降低了对男性化面孔的偏好，提高了对女性化面孔的偏好。此发现进一步验证了 DeBruine 等人的结论，一方面非面孔线索可能对面孔吸引力的感知产生影响，另一方面非面孔线索的有无（如头发等）会影响男性化偏好，换句话说，对有非面孔线索面孔的男性化偏好要系统高于无非面孔线索面孔。非面孔线索可能对面孔男性化偏好和面孔吸引力产生影响的这一发现为今后的研究提供了方向，正如 DeBruine 等人所提出的进一步探讨头发线索的有无以及头发的样式（如，短的，男性化头发 vs. 长的，女性化头发）对男性化偏好的影响将是未来面孔感知研究的一个有趣课题。以上结果可以看出，性别二态技术和感知男性化技术两种方法产生的面孔刺激不论是在非面孔线索掩蔽和未掩蔽的情况下，都存在显著差异。这一结果支持了 Rennels 等人所认为的男性面孔吸引力评价产生的不一致结果可能部分是由不同方法所产生的不同面孔刺激所造成的结论。

4.1.2 对性别二态线索的男性面孔信任度分析

从信任度的评价结果可以看出，不论是在非面孔线索掩蔽或未掩蔽的情况下，在感知男性化技术条件下均与对原始照片的评价结果一致，女性都认为女性化的男性面孔更有信任度，而在性别二态技术条件下，女性对性别二态线索的男性面孔的信任度评价在非面孔线索未掩蔽时，存在边缘显著差异，即女性偏好男性化的男性面孔。可见，感知男性化技术条件与原始照片的信任度评价结果一致，而性别二态技术与感知男性化技术两种条件所获得的面孔刺激所产生的信任度评价也存在不一致性，这与前面对吸引力的结果一致，同样支持了 Rennels 等人 (2008) 的结论。而非面孔线索对男性面孔信任度的评价并未造成干扰，在原始照片、感知男性化技术条件和性别二态技术条件下，非面孔线索的掩蔽和未掩蔽情况都得到了相似的结果。

值得注意的是，在前面吸引力的评价中，非面孔线索的掩蔽对男性面孔吸引力的评价有一定的影响，而对信任度的评价却没有什么影响。这一发现与 Smith 等人 (2009) 考察女性自身吸引力对男性化和女性化的男性面孔在吸引力和信任度两个指标上感知的预测结果具有类似之处，他们发现相对于高腰臀比（低吸引力）的女性，低腰臀比（高吸引力）的女性认为男性化的男性面孔更有吸引力，而女性自身的吸引力对男性面孔的信任度感知上却没有什么影响。我们可以看出，本研究中非面孔线索掩蔽与否对男性面孔吸引力和信任度的评价，与前人关于女性条件依赖偏好 (condition-dependent preferences) 的发现具有一致性，正如相对于低吸引力的女性而言，高吸引力的女性对男性化的面孔吸引力感知增强，而不是提高对一般积极特质（如信任度）的感知，同样，非面孔线索掩蔽与否主要对男性面孔的吸引力评价产生干扰，而不是广泛意义上对亲社会特质都会产生感知偏差，表明非面孔线索的掩蔽对男性面孔感知存在领域的特殊性，那么，非面孔线索的掩蔽是否只对男性面孔的吸引力评价产生干扰，而对其他的亲社会特质（如责任感等）是否会产生影响呢？这值得未来研究进一步探究。

4.1.3 权衡模型和文化差异的解释 本研究对性别二态线索的男性面孔的吸引力和信任度评价的主要发现为，不论在非面孔线索掩蔽或未掩蔽时，感知男性化技术条件和对原始照片的评价结果一致，即相对于男性化的男性面孔，中国女性都认为女性化的男性面孔更有吸引力和信任度。而国外采用双维的感知男性化技术都产生了一致的结果，即都偏好男性化的男性面孔 (DeBruine et al., 2006; Rennels et al., 2008; DeBruine et al., 2010)，Rhodes (2006) 通过元分析也发现未经改变的原始照片男性化与吸引力的评价之间呈中等的正相关。可见，本研究中国女性偏好女性化的男性面孔的结果与西方偏好男性化面孔存在不一致性，这可以通过前文所提的多元化策略权衡模型和中西文化差异的视角来理解。根据女性对男性化偏好的权衡理论，那些能改变女性选择男性伴侣的"收益"

和"代价"的相对重要性的因素会影响女性偏好男性化和女性化的强度，而文化因素会影响女性的"权衡"决策。与西方相比，中国为典型的集体主义和长期取向文化的社会，即中国人更关心群体利益而不是个人利益，倾向于做长期规划和投入。王登峰、崔红 (2007) 提出中国文化是女性化文化，同时又是反男性化文化的假设，并且在女性的"温柔"气质占主导的社会，人们谦虚、温和且有教养，注重保持良好人际关系，这与中国文化中的重视关系取向相一致，中国社会注重人情，而人情是由各种各样的人与人之间的关系构成的 (佐斌，2009)。由此可见，与西方女性相比，中国女性受群体利益，和谐关系和长期承诺的文化因素的影响，更愿意重视亲代投资，从而更多偏好女性化的面孔，并且，男性面孔的女性化标示着亲社会特质，从而有利于促进良好的人际关系。

　　4.1.4　性别角色理论解释及方法建议　本研究中女性对性别二态技术和感知男性化技术两种方法所产生的性别二态线索的男性面孔刺激不同偏好结果可以从性别角色理论的角度来理解。正如前文所介绍的，性别二态技术符合传统的性别角色模式理论，感知男性化技术与双性化理论相一致。本研究中，在性别二态技术条件对吸引力和信任度的评价时，女性都总体偏爱男性化的男性面孔，这可能与传统性别角色模式理论中"具有男性特质的男性和具有女性特质的女性在心理上更为健康"有关 (钱铭怡等，2000)。而在感知男性化条件下，女性在对吸引力和信任度评价时都更偏好女性化的男性面孔，这可能与"中国被试中女性化个体心理社会适应最好"有关 (王登峰，崔红，2007)。此外，女性对女性化男性面孔的偏好这一结果与佐斌、刘晅 (2006) 采用内隐联想测验 (IAT) 和刻板解释偏差 (SEB) 两种内隐社会认知的方法测量性别刻板印象的发现存在相似之处，即女性持有的性别刻板印象与传统的认知习惯有所不同。

　　值得一提的是，从以上对性别二态线索的男性面孔的吸引力和信任度评价的分析可以看出，当非面孔线索 (如发型等) 未掩蔽和掩蔽时，感知男性化技术与原始照片两种条件下认为女性化的男性面孔更有

吸引力和信任度；在性别二态技术条件下，非面孔线索未掩蔽时认为男性化的男性面孔更有吸引力和信任度，其它条件下对吸引力和信任度的评价不存在显著差异。总体来看，双维的感知男性化技术条件的结果与原始照片的研究一致，而与单维的性别二态技术条件的结果相反，这与 Rennels 等人 (2008) 的发现一致，也符合 Rhodes (2006) 元分析的结果，即性别二态技术产生的面孔刺激与原始照片所得结果相反。故在此认为以往关于女性偏好男性化还是女性化的男性面孔的研究结果不一致性，可能部分原因是由不同方法产生的不同面孔刺激类型所造成的。单维的性别二态技术在得到男性化和女性化的男性面孔图像时，采用的是不同的操作，男性化的男性面孔是通过将男性的平均化面孔图像原型扭曲偏离女性的平均化面孔图像原型获得，女性化的男性面孔则是将男性的平均化面孔图像原型扭曲朝向女性的平均化面孔图像原型而得到，这种转化操作上的差异可能让男性化和女性化的男性面孔本身都存在很大的不同。Johnston 等人 (2001) 表明性别二态技术假定男性化的面孔形状能通过将男性与女性的平均化面孔图像原型形状差异线性外推法 (linear extrapolation) 而产生，此方法可能是无效的，因为两性在面孔形状上的差异是由生长激素，雄激素和雌激素复杂的作用所导致的骨骼生长的结果 (Grumbach, 2000)，另有一些研究者表明男性面孔与女性面孔本身在某种程度上具有不同的神经表征 (Little, DeBruine, & Jones, 2005)，可以看出，此方法可能缺乏生态效度，不能准确反映出吸引力与男性化 – 女性化之间的关系。而双维的感知男性化技术是将被试所认为的高男性化和低男性化的面孔进行平均化处理，保留了面孔样本男性化 – 女性化的信息，并且与未经改变原始照片的结果一致，能为探究性别二态线索对面孔吸引力的作用提供有用的面孔刺激。故在此建议心理学者在使用单维的性别二态技术时需采取谨慎的态度，正如 Rennels 等人 (2008) 所提议，如果心理学者的目的是为了更好地理解感知过程 (如分类知觉)，那么使用单维的性别二态技术是合适的，然而，如果研究目标是模仿自然的面孔变化，那么心

理学者在将他们的研究结论推论到真实面孔之前，需要确保所采用面孔的生态效度。

4.2 基于眼动分析性别二态线索对面孔偏好的影响

4.2.1 性别二态线索对面孔吸引力影响的行为指标分析 从被试对性别二态线索的男女面孔刺激的吸引力的行为指标评价结果可以看出，对于男性面孔刺激，被试认为男性化面孔比女性化更有吸引力，这一结果支持了实验1中采用性别二态技术被试更偏好男性化男性面孔的结论。对于女性面孔刺激，相对于男性化面孔，被试认为女性化面孔更有吸引力，这一结果与 Perrett 等人 (1998) 以及 Koehler, Simmons, Rhodes 和 Peters (2004) 的发现一致，即女性化的女性面孔更有吸引力。

4.2.2 性别二态线索对面孔吸引力影响的眼动指标分析 平均瞳孔大小是眼动研究的一个重要指标，一般而言，瞳孔直径指标与人的情绪、兴趣和认知负荷有关 (韩映虹，闫国利，2010), 认知负荷的增加会引起瞳孔的扩张 (周颖，刘俊升，2009)。从认知加工的角度来看，被试比较不同性别二态线索的男性面孔吸引力时的平均瞳孔大小显著大于女性面孔，这符合以往研究结论，关于性别二态线索对女性面孔影响存在更多共识，即一致表明女性化的女性面孔更有吸引力，而关于性别二态线索对男性面孔的影响却存在争论，说明人们更容易判断性别二态线索对女性面孔吸引力的影响，需要更少的加工负荷，相对于女性面孔，被试在评价不同性别二态线索的男性面孔吸引力时给予更多的加工负荷。也就是说当评价不同性别二态线索的男性面孔刺激时比女性面孔需要更多的心理资源对吸引力的信息进行深层的分析。这也可为以往研究结果一致表明女性化的女性面孔更有吸引力，而人们对性别二态线索的男性面孔吸引力评价却一直存在争论提供一定的解释。

首次注视时间是在某个区域内第一个注视点开始的时间。被试对男性化面孔的首次注视时间显著长于女性化面孔，表明相对男性化面孔，人们对女性化面孔会更优先注意。首次注视持续时间是在某一

个区域内第一个注视点持续的时间。首次注视持续时间代表了对材料的早期识别过程以及对材料加工难度的敏感 (丁小燕，孔克勤，王新法，2007)。从首次注视持续时间结果可以看出，被试对男性化面孔的首次注视持续时间显著长于女性化面孔，并且他们对女性面孔的首次注视时间边缘显著长于男性面孔，表明被试对男性化面孔和女性照片的面孔在早期识别过程中对材料加工难度的敏感要高于女性化面孔和男性面孔的刺激。这也可能说明，性别二态线索对于吸引力判断的影响中，被试对于男性化线索会有更高的加工水平和负荷。注视次数是某个区域内所有注视点的个数，反映了被试对刺激的处理能力，刺激本身的难易程度等。从注视次数的结果发现，被试对男性面孔刺激的注视次数边缘显著多于女性面孔。注视次数反映加工难度，表明与女性面孔相比，被试对男性面孔刺激材料判断的难度更大，加工程度更深。

4.3 综合讨论

本研究通过两个实验考察了人们对男性化与女性化的面孔偏好情况。实验一探讨女性对性别二态线索的男性面孔的吸引力和信任度评价，比较性别二态技术和感知男性化技术条件在吸引力和信任度评价上的差异，主要发现为感知男性化技术条件与对原始照片的吸引力和信任度评价均存在一致的结果，即女性都认为女性化的男性面孔更有吸引力和信任度，这与西方偏好男性化的结果不一致性可从中国的文化背景和权衡模型的角度来理解。而性别二态技术条件在吸引力和信任度评价上，产生了与感知男性化技术条件不一致的结果，即在性别二态技术条件下，人们认为男性化的男性面孔更有吸引力和信任度，这一结果则符合性别角色理论，即从男性化与女性化人格维度上看，性别二态技术更接近于单维观点，而感知男性化技术则更接近于双维观点。此外，非面孔线索的掩蔽与否对性别二态技术条件和感知男性化技术条件的吸引力评价都产生混淆。因此，如果心理学者采用性别二态技术来考察对性别二态线索下的男性面孔的偏好时，则需采取谨慎的态度，一方面要考虑性别二态技术可能与感知男性化技术和对原始照片

评价的结果存在不一致性，另一方面值得注意的是非面孔线索 (如发型，衣着等) 可能会对性别二态技术和感知男性化技术两种条件的面孔吸引力评价产生干扰。实验二则结合眼动，采用性别二态技术操作面孔刺激材料，探索了性别二态线索对面孔吸引力影响的眼动指标，行为指标表明被试更偏好男性化的男性面孔和女性化的女性面孔。眼动指标则表明被试比较不同性别二态线索的男性面孔吸引力时的平均瞳孔大小显著大于女性面孔；被试对男性化面孔的首次注视时间和首次注视持续时间均显著长于女性化面孔，并且他们对女性面孔的首次注视时间边缘显著长于男性面孔；被试对男性面孔刺激的注视次数边缘显著多于女性面孔，可以看出，眼动指标为性别二态线索对面孔吸引力影响提供了一定的佐证。

4.4 本研究的局限与今后研究的展望

本研究为国内对性别二态线索的面孔偏好的初步探索，存在一定的局限性，主要有以下几个方面值得注意和进一步研究：(1) 被试方面，本研究所选取的被试主要是大学生，样本代表性具有一定的局限，未来研究值得扩大样本，进一步考察不同性别，年龄和职业等中国人群体是否存在相似的结果；(2) 面孔数据库方面，本研究所采用的照片材料主要来自某高校大学生信息采集的数据库，面孔刺激照片材料的样本和像素等存在一定的限制，进一步丰富面孔刺激数据库来源，扩大面孔吸引力研究的生态效度值得未来进一步探究；(3) 面孔刺激方面，本研究中被试主要对静态出现的配对面孔进行吸引力评价，进一步采用男性化 – 女性化的动态转化面孔刺激，考察在性别二态转化过程中被试的眼动情况将成为一项有趣的课题；(4) 指标评价方面，随着神经影像技术的兴起与运用，进一步结合采用功能磁共振成像术 (fMRI)，事件相关电位 (ERP) 等认知神经科学技术设备，探讨对性别二态线索的面孔感知的神经生理机制，为面孔感知研究提供更客观科学的指标，也将成为今后研究的方向。

参考文献

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology, 42*, 155–162.

Boothroyd, L. G., Jones, B. C., Burt, D. M., & Perrett, D. I. (2007). Partner characteristics associated with masculinity, health and maturity in male faces. *Personality and Individual Differences, 43*, 1161–1173.

Botta, R. A. (2003). For your health? The relationship between magazine reading and adolescents' body image and eating. *Eating Disorders Review, 48*(9–10), 389–399.

Buckingham, G., DeBruine, L. M., Little, A. C., Welling, L. L. M., Conway, C. A., Tiddemand, B. P., et al. (2006). Visual adaptation to masculine and feminine faces influences generalized preferences and perceptions of trustworthiness. *Evolution and Human Behavior, 27*, 381–389.

Constantinople, A. (1973). Masculinity–femininity: An exception to a famous dictum? *Psychological Bulletin, 80*, 389–407.

Cai, H. J., & Yang, Z. L. (2002). Using IAT to study the structure of gender self–concept. *Acta Psychologica Sinica, 34*, 168–174.

[蔡华俭，杨治良 . (2002). 大学生性别自我概念的结构 . *心理学报 , 34*, 168–174.]

DeBruine, L. M., Jones, B. C., Little, A. C., Boothroyd, L. G., Perrett, D. I., Penton–Voak, I. S., et al. (2006). Correlated preferences for facial masculinity and ideal or actual partner's masculinity. *Proceedings of the Royal Society of London B, 273*, 1355–1360.

DeBruine, L. M., Jones, B. C., Crawford, J. R., Welling, L. L. M., & Little, A. C. (2010). The health of a nation predicts their mate preferences: Cross–cultural

variation in women's preferences for masculinized male faces. *Proceedings of the Royal Society of London B, 277*, 2405 – 2410.

DeBruine, L. M., Jones, B. C., Smith, F. G., & Little, A. C. (2010). Are attractive men's faces masculine or feminine? The importance of controlling confounds in face stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 36*, 751 – 758.

Ding, X. Y., Kong, K. Q., & Wang, X. F. (2007). The relationship between eye movements and reading achievements in English fast reading. *Psychological Science, 30*, 535 – 539.

[丁小燕 , 孔克勤 , 王新法 . (2007). 英文快速阅读的眼动特点与阅读成绩的关系 . *心理科学 , 30*, 535 – 539.]

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology, 24*, 285 – 290.

Enquist, M., Ghirlanda, S., Lundqvist, D., & Wachtmeister, C. (2002). An ethological theory of attractiveness. In G. Rhodes & L. A. Zebrowitz (Eds.), *Facial attractiveness: Evolutionary, cognitive, and social perspectives* (pp. 127 – 151). Westport, CN: Ablex.

Fink, B., & Penton–Voak, I. (2002). Evolutionary psychology of facial attractiveness. *Current Directions in Psychological Science, 11*, 154 – 158.

Gangestad, S. W., & Buss, D. M. (1993). Pathogen prevalence and human mate preferences. *Ethology and Sociobiology, 14*, 89–96.

Gangestad, S. W., & Simpson, J. A., (2000). The evolution of human mating: Tradeoffs and strategic pluralism. *Behavioral and Brain Sciences, 23*, 573 – 644.

Gangestad, S.W., & Scheyd, G. J. (2005). The evolution of human physical attractiveness. *Annual Review of Anthropology, 34*, 523 – 548.

Grumbach, M. M. (2000). Estrogen, bone, growth and sex: A sea change in conventional wisdom. *Journal of Pediatric Endocrinology and Metabolism, 13*, 1439 – 1455.

Han, Y. H., & Yan, G. L. (2010). The application of the method of eye movement analysis in preschoolers' cognitive research. *Psychological Science, 33*, 191 – 193.

[韩映红 , 闫国利 . (2010). 眼动分析法在学前儿童认知研究中的应用 . *心理科学 , 33*, 191 – 193.]

Hofstede, G. (Eds.), Xu, L. S. (Review). (2008). *Culture⑤s consequence: Comparing values, behaviors, institutions and organizations across nations (2nd ed.). Shanghai*, China: Shanghai Foreign Language Education Press.

[Hofstede, G. (着). 许力生导读 . (2008). *文化之重 : 价值、行为、体制和组织的跨国比较 (第二版)*. 上海 : 上海外语教育出版社 .]

Johnston, V. S., & Franklin, M. (1993). Is beauty in the eye of the beholder? *Ethology and Sociobiology, 14*, 183 – 199.

Johnston, V. S., Hagel, R., Franklin, M., Fink, B., & Grammer, K. (2001). Male facial attractiveness: Evidence for hormone–mediated adaptive design. *Evolution & Human Behavior, 22*, 251 – 267.

Koehler, N., Simmons, L. W., Rhodes, G., & Peters, M. (2004). The relationship between sexual dimorphism in human faces and fluctuating asymmetry. *Proceedings of the Royal Society of London Series B: Biological Sciences, 271*, S233 – S236.

Low, B. (1990). Marriage systems and pathogen stress in human societies. A*merican Zoologist, 30*, 325 – 339.

Li, O., & Chen, H. (2010). The retrospect and prospect of facial attractiveness. *Advances in Psychological Science, 18*, 472 – 479.

[李鸥 , 陈红 . (2010). 面孔吸引力的回顾与前瞻 . *心理科学进展 , 18*, 472 – 479.]

Liang, J., & Li, H. (2010). A review on the researches of facial attractiveness. *The Guide of Science and*

Education, (1), 120－121.

[梁静 , 李红 . (2010). 面孔吸引力研究的概览 . *科教导刊 (中旬刊)*, (1), 120－121.]

Little, A. C., Burt, D. M., Penton-Voak, I. S., & Perrett, D. I. (2001). Self-perceived attractiveness influences human female preferences for sexual dimorphism and symmetry in male faces. *Proceedings of the Royal Society of London Series B, 268,* 39－44.

Little, A. C., & Hancock, P. J. B. (2002). The role of masculinity and distinctiveness in judgments of human male facial attractiveness. *British Journal of Psychology, 93,* 451－464.

Little, A. C., DeBruine, L. M., & Jones, B. C. (2005). Sex-contingent face after-effects suggest distinct neural populations code male and female faces. *Proceedings of the Royal Society of London Series B, 272,* 2283－2287.

Ma, H. W., Yu, Q. Y., & Chen, H. (2007). A review on the research methods of facial attractiveness. *Psychological Science, 30,* 906－908.

[马华维 , 俞琴燕 , 陈浩 . (2007). 面孔吸引力研究方法综述 . *心理科学 , 30,* 906－908.]

McCabe, M. P., & Ricciardelli, L. A. (2001). Body image and body change techniques among young adolescent boys. *European Eating Disorders Review, 9,* 1–13.

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the USA, 105,* 11087–11092.

Penton-Voak, I. S., Jacobson, A., & Trivers, R. (2004). Populational differences in attractiveness judgements of male and female faces: Comparing British and Jamaican samples. *Evolution and Human Behavior, 25,* 355－370.

Penton-Voak, I. S., Jones, B. C., Little, A. C., Baker, S., Tiddeman, B., Burt, D. M., et al. (2001). Symmetry, sexual dimorphism in facial proportions and male facial attractiveness. *Proceedings of the Royal Society of London Series B, 268,* 1617－1623.

Penton-Voak, I. S., Little, A. C., Jones, B. C., Burt, D. M., Tiddeman, B. P., & Perrett, D. I. (2003). Female condition influences preferences for sexual dimorphism in faces of male humans (Homo sapiens). *Journal of Comparative Psychology, 117,* 264－271.

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, D., Yoshikawa, S., Burt, D. M., et al. (1998). Effects of sexual dimorphism on facial attractiveness. *Nature, 394,* 884－887.

Qian, M. Y., Zhang, G. J., Luo, S. H., & Zhang, S. (2000). Sex role inventory for college students (CSRI). *Acta Psychologica Sinica, 32,* 99－104.

[钱铭怡 , 张光健 , 罗珊红 , 张莘 . (2000). 大学生性别角色量表 (CSRI) 的编制 . *心理学报 , 32,* 99－104.]

Rennels, J. L., Bronstad, P. M., & Langlois, J. H. (2008). Are attractive men's faces masculine or feminine? The importance of type of facial stimuli. Journal of Experimental Psychology: *Human Perception and Performance, 34,* 884－893.

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology, 57,* 199－226.

Rhodes, G., Hickford, C., & Jeffery, L. (2000). Sex-typicality and attractiveness: Are super-male and super-female faces super-attractive. *British Journal of Psychology, 91,* 125－140.

Rhodes, G., Simmons, L. W., & Peters, M. (2005). Attractiveness and sexual behavior: Does attractiveness enhance mating success? *Evolution and Human Behavior, 26,* 186－201.

Smith, F. G., Jones, B. C., Welling, L. L. W., Little, A. C., Vukovic, J., Main, J. C., et al. (2009). Waist-hip ratio predicts women's preferences for masculine male faces, but not perceptions of men's trustworthiness. *Personality and Individual Differences, 47,* 476－480.

Sun, Y. Y. (2004). See difference of Chinese and Western culture from aesthetic way. *Journal of Changchun*

University, 14, 38 – 40.

[孙银玉 . (2004). 从审美方式看中西文化的差异 . *长春大学学报 , 14,* 38 – 40.]

Swami, V., & Tovée, M. J. (2005). Male physical attractiveness in Britain and Malaysia: A cross-cultural study. *Body Image, 2,* 383 – 393.

Theodoridou A., Rowe, A. C., Penton-Voak, I. S., & Rogers, P. J. (2009). Oxytocin and social perception: Oxytocin increases perceived facial trustworthiness and attractiveness. *Hormones and Behavior, 56,* 128 – 132.

Thornhill, R., & Gangestad, S. W. (1999). Facial attractiveness. *Trends in Cognitive Sciences, 3,* 452 – 460.

Tiddeman, B. P., Perrett, D. I. & Burt, D. M. (2001). Prototyping and transforming facial textures for perception research. *IEEE Computer Graphics and Applications, 21,* 42 – 50.

Todorov, A. (2008). *Evaluating faces on trustworthiness: An extension of systems for recognition of emotions signaling approach/avoidance behaviors. In M. Miller & A. Kingstone (Eds.), 1124,* 208 – 224.

Wang, D. F., & Cui, H. (2007). Development of Chinese sex role scale and relations between sex role and psychosocial adjustment. *Journal of South west University (Social Sciences Edition), 33,* 1 – 9.

[王登峰 , 崔红 . (2007). 中国人性别角色量表的建构及其与心理社会适应的关系 . *西南大学学报 (社会科学版), 33,* 1 – 9.]

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science, 17,* 592 – 598.

Xu, Z. H., Liu, D. Z., Huang, X., Liu, J. G., Zhang, J., & Li, Y. Q. (2010). The development of the gender role of college students: A comparative study between undifferentiation and androgyny. *Psychological Science, 33,* 219 – 222.

[徐振华 , 刘电芝 , 黄欣 , 刘金光 , 张姣 , 李宇青 . (2010). 大学生性别角色形成研究—— 双性化与未分化个案的对比 . *心理科学 , 33,* 219 – 222.]

Zaidel, D. W., Bava, S., & Reis, V. A. (2003). Relationship between facial asymmetry and judging trustworthiness in faces. *Laterality, 8,* 225 – 232.

Zhang, Y., Kong, F. C., Chen, H., Xiang, Y. H., Gao, X., & Chen, M. Y. (2010). Cognitive bias toward female facial attractiveness in males: Evidences from an ERP Study. *Acta Psychologica Sinica, 42,* 1060 – 1072.

[张妍 , 孔繁昌 , 陈红 , 向燕辉 , 高笑 , 陈敏燕 . (2010). 男性对女性面孔吸引力的认知偏好 : 来自 ERP 的证据 . *心理学报 , 42,* 1060 – 1072.]

Zhou, Y., & Liu, J. S. (2009). Using eye-movement to study implicit aggressiveness. *Psychological Science, 32,* 858 – 860.

[周颖 , 刘俊升 . (2009). 运用眼动指标探测个体的内隐攻击性 . *心理科学 , 32,* 858 – 860.]

Zuo, B. (Eds.). (2009). *Social Psychology* (pp. 408 – 409). Beijing: Higher Education Press.

[佐斌 . (主编). (2009). *社会心理学* (pp. 408 – 409). 北京 : 高等教育出版社 .]

Zuo, B., & Liu, X. (2006). The researches of implicit gender stereotype based on IAT and SEB. *Psychological Development and Education, 22,* 57 – 63.

[佐斌 , 刘晅 . (2006). 基于 IAT 和 SEB 的内隐性别刻板印象研究 . *心理发展与教育 , 22,* 57 – 63.]

The effects of transformed gender facial features on face preference of college students: Based on the test of computer graphics and eye movement tracks

WEN Fangfang; ZUO Bin

（ School of Psychology, Huazhong Normal University, Wuhan 430079, China ）

Abstract   Perceived facial attractiveness can influence people's social interactions with one another, including mate selection, intimate relationship, hiring decision, and voting behavior. People evaluate faces using multiple trait dimensions such as attractiveness and trustworthiness both of which are affected by facial masculinity or femininity cues. However, studies manipulating the computer graphics of sexual dimorphism on facial attractiveness have yielded inconsistent results. Some found that feminine facial features in male faces were more attractive than masculine ones. Some others found that women prefer masculine male faces. And still others found that women preferred femininity in male faces. The current study used the computer graphics and the eye tracker to assess the effect of the dimorphic cues on the perception of facial attractiveness among Chinese college students through two experiments. Experiment 1 assessed women's perceptions of attractiveness and trustworthiness of men's faces under the condition of either perceived masculinity vs. femininity or the sexual dimorphism. Results showed that, when non-face cues (e.g., hairstyle) were masked, women perceived femininity in men's faces as more attractive and trustworthy than the masculinity. However, in the sexual dimorphism condition in which the non-face cues were not masked, women found masculinity in men's faces more attractive and trustworthy. Experiment 2 used the eye tracker to assess the effects of the dimorphic cues on the evaluation of facial attractiveness. Results showed that the subjects preferred the masculine male faces obtained by the sexual dimorphism and feminized female face. Eye movement tracking showed that average pupil dilation and average fixation count on a male face were significantly higher than on a female face. The first fixation time was significantly greater for the masculine faces than for the feminine ones, but the first fixation time was significantly shorter for the male faces than the female ones. The first fixation time and first fixation duration for masculine faces were both significantly longer than for feminine ones. These indicators of eye movement provide some evidence for the effect of the sexual dimorphism on the facial attractiveness.

Keywords   facial attractiveness; sexual dimorphism; perceived masculinity; computer graphics; track of eye movement

大学生个人奋斗、人格特质与主观幸福感的关系

杨慧芳[1,2]，郭永玉[1]

（[1] 华中师范大学心理学院暨湖北省人的发展与心理健康重点实验室 武汉 430079）

（[2] 广东湛江师范学院教务处 广东湛江 524048）

摘 要 本研究采用问卷调查法，以武汉市 6 所高校的 533 名本科生为被试，通过结构方程建模的方法，探讨了人格特质、个人奋斗与主观幸福感三者之间的关系，并比较了个人奋斗和人格特质对主观幸福感的影响作用。研究发现，个人奋斗和人格特质都对主观幸福感有影响；人格特质对主观幸福感的影响大于个人奋斗对主观幸福感的影响；人格特质对个人奋斗有影响。因此，人格的特质因素与动机因素对主观幸福感有不同的影响作用。

关键词 个人奋斗；人格特质；主观幸福感；大学生

分类号 B844

1 问题提出

人格心理学中有两种不同的概念：特质和动机，分别对应于人格心理学中两种重要的研究取向：人格的特质取向与动机取向。Cantor 追溯了人格心理学从诞生直到当代的演变进程，发现人格心理学中一直存在这两种研究取向。这两种研究取向的区别在于：特质取向强调行为的特点和习惯性倾向，而动机取向更强调人类行为在情境下的目的性和目标导向性，侧重于从动力性和特异性的角度理解人格。人格特质和个人奋斗分别代表了人格中的特质因素与动机因素、静态结构和动力特征。

个人奋斗 (personal striving) 是个体当前正努力去做的事，代表了个体通过其行为或目标试图去独特地完成的事，由一组目标组成。个人奋斗构念作为动机的"载体单元"之一，体现了人格心理学中逐渐兴起的意动观点，代表了动机研究的新取向。Emmons 指出，拥有有意义的生活目标并朝向其前进是主观幸福感的一个先决条件。个体目标的各种特征对主观幸福感有重要的影响。

回顾关于主观幸福感的研究，大多是关于人格特质与主观幸福感的。大量研究认为，外向性和神经质这两种人格特质解释了人格和主观幸福感之间主要的相关。然而，有研究发现，个人奋斗的特征比人格特质能解释主观幸福感中更多的变异。Emmons 指出，个人奋斗是人格和主观幸福感关系研究中一种新取向。随着主观幸福感研究的深入，仅仅从人格的特质取向来研究势必不能全面地陈述人格和主观幸福感之间的复杂关系。因此，如果从人格的动机取向来研究，可以丰富人格与主观幸福感之间关系的研究。

个人奋斗的概念来自于西方，现有关于个人奋

斗与主观幸福感关系的研究都是在国外进行的。无论东方还是西方，每个人都有自己独特的个人奋斗，即都有自己努力去追求的目标和想要去做的事。研究中国大学生的个人奋斗对主观幸福感的影响，不仅可弥补个人奋斗在中国心理学研究中的空白，而且对于大学生的主观幸福感研究具有重要的实践意义。

此外，人格特质与个人奋斗作为两种不同的人格分析单元，使我们得以触及人生的"静止"与"流变"两个不同的层面。因此，研究人格特质、个人奋斗与主观幸福感之间的关系，可以了解并比较人格的特质因素、"静止"层面与人格的动机因素、"流变"层面对主观幸福感的影响作用。

综上所述，本研究拟借助于结构建模的方法，考察人格特质、个人奋斗与主观幸福感三者之间的关系；并比较人格特质、个人奋斗对主观幸福感的预测力。由于对特质的分类至今尚未取得共识，本研究选取了较有影响的五因素模型 (FFM)，文中的人格特质采用的是五因素模型所提出的五种人格特质。

2 研究方法

2.1 被试

从武汉地区 6 所高校，按照文科、工科、理科、农学、医学五个层次进行分层随机抽样，随机抽取大学本科生，得到各年级的有效被试 533 名。被试的平均年龄在 18 ~ 25 岁之间。由经过培训后的研究生主试采用团体施测的方式进行；为了控制额外因素，主试的指导语是统一的。正式施测用时 30 分钟左右，问卷当堂发放，当堂回收。有关被试的专业、性别、家庭来源、是否独生等情况见表 1。

2.2 研究工具

2.2.1 个人奋斗问卷

该问卷修订自 Emmons 编制的个人奋斗量表

(Striving Assessment Scales)。原量表要求被试列出 15 条个人奋斗，并对每条个人奋斗从 15 个维度进行评价。本研究要求被试列出 5 条个人奋斗，并从 5 个维度进行评价。之所以选取 5 个维度，是基于以下两个原因：首先，Emmons 曾直接从 15 个维度中选取 6 个评价维度；其次，本研究在预测中对部分被试进行了深度访谈，在回答"列出多少条个人奋斗较为合理"的问题时，85.2% 的被试回答为 5 条。根据已有文献和对大学生的访谈结果，本研究最后确定了 5 个维度：重要性、成功可能性、清晰性、难度、以往成就。重要性是指个人奋斗对个体的重要程度如何。难度是指要想在个人奋斗上取得成功，个体认为会有多大的难度。清晰性是指，为了在个人奋斗上取得成功，个体对需要做的事情应该有一个大体的思路，该思路的清晰程度如何。成功可能性是指，未来要在该个人奋斗上取得成功，个体认为可能性有多大。以往成就是指，最近几个月中，个体在个人奋斗上取得的成功程度如何。

由于每个维度的分数是由 5 条个人奋斗在该维度上的得分平均而来的，那么，每个维度实则包含了 5 个项目。原问卷部分项目采用利克特 5 点量表计分，部分项目采用利克特 9 点量表计分，还有部分项目采用利克特 7 点量表计分；本研究中，统一采用 5 点评定，从 1 到 5 采用正向计分。问卷共 25 个项目，包括重要性、成功可能性、清晰性、难度、以往成就 5 个维度。各维度的表达式为：在重要性维度上，1 表示"非常不重要"，5 表示"非常重要"；在难度维度上，1 表示"非常容易"，5 表示"非常难"；在清晰性维度上，1 表示"非常不清晰"，5 表示"非常清晰"；在成功可能性上，1 表示"非常小"，5 表示"非常大"；在以往成就维度上，1 表示"非常失败"，5 表示"非常成功"。各维度的得分在 1~5 分之间。

2.2.2 人格问卷

采用"五因素"人格问卷 (NEO– FFI) 简式版。

表 1 被试情况一览表

专业					性别		家庭来源		是否独生		总计
文科	工科	理科	农学	医学	男	女	农村	城镇	独生	非独生	
123	117	104	100	89	272	261	212	309	136	397	533

该问卷是 Costa 和 McCrae 编制的，由中科院心理健康实验室张建新翻译修订而成。在华人地区试用的结果表明，它具有较好的信效度。该问卷共 60 个题目，由 5 个分量表组成，分别测量神经质（N）、外向性（E）、开放性（O）、随和性（A）和尽责性（C）。在一个 5 点量尺上（从 1 "非常不同意" 到 5 "非常同意"），要求被试对每个项目进行评定。本研究以每个维度的粗分作为测量指标，每个维度的得分在 1~5 分之间。根据 Costa 和 McCrae 的报告，5 个分量表的内部一致性系数分别是：神经质 0.92、外向性 0.89、开放性 0.88、随和性 0.86、尽责性 0.90。通过对五种人格维度的信度进行检验，本研究发现，被试在五种人格维度上的内部一致性系数分别是：神经质 0.83、外向性 0.80、开放性 0.86、随和性 0.86、尽责性 0.84。

2.2.3 主观幸福感问卷

包括两部分：一为情绪量表，测量主观幸福感的情感成分；二是生活满意度量表，测量主观幸福感的认知成分。第一部分的情绪量表，是本研究根据 Diener 编制的情绪量表修订而成。该量表是形容词评定量表，包括积极情感和消极情感两个分量表，共 9 个项目，其中 4 个项目测量积极情感的体验频率，5 个项目测量消极情感的体验频率。在一个 5 点量尺上（从 1 "从来没有" 到 5 "经常"），要求被试对过去一个月内感受到某种情绪的次数进行评价，在某种情绪上的得分越高表明该种情绪体验越多。积极情感和消极情感的得分都在 1~5 分之间。积极情感维度的内部一致性系数为 0.86，消极情感维度的内部一致性系数为 0.80。第二部分的生活满意度量表，采用的是张兴贵编制的青少年学生生活满意度量表，由 36 个项目组成，采用 7 级计分。生活满意度得分范围在 1~7 分之间，得分越高，说明生活满意度越高。

2.3 统计分析

本研究采用 LISREL8.30 和 SPSS11.5 对研究数据进行统计分析和处理。其中，五因素人格问卷（NEO-FFI）的数据由中科院心理所心理健康实验室帮助统计。

3 研究结果

3.1 个人奋斗、人格特质与主观幸福感的基本状况

个人奋斗、人格特质与主观幸福感的得分情况见表 2。在个人奋斗的各个维度上，大学生在个人奋斗重要性上的得分为 4.58~0.43，在个人奋斗难度上的得分为 3.60~0.53，在个人奋斗清晰性上的得分为 3.81~0.69，在个人奋斗可能性上的得分为 3.60~0.59，在个人奋斗以往成就上的得分为 3.21~0.56。可见，大学生普遍认为，个人奋斗对自己非常重要；而要实现个人奋斗，则有一定难度；对于如何在个人奋斗上取得成功以及应该做哪些事情以达到成功，大学生没有较清晰的思路；大学生认为，在个人奋斗上取得成功的可能性比较大；此外，大学生认为近期在个人奋斗上取得了一定的成就。

在大学生主观幸福感的各个成分中，生活满意度的平均得分为 4.81~0.66，积极情感的平均得分为 3.64~0.82，消极情感的得分为 2.49~0.76。这些说明大学生的生活满意度较高，体验到较多的积极情感和较少的消极情感，即大学生的主观幸福感水平较高。

另外，在大五人格各维度上，神经质的平均得分为 2.75~0.61，外向性的平均得分为 3.18~0.49，开放性的平均得分为 3.37~0.45，随和性的平均得分为 3.61~0.37，尽责性的平均得分为 3.42~0.50。由大学生在大五人格各维度上的得分可知，大学生在神经质这一人格特质上得分较低，而在外向性、开放性、随和性和尽责性四种人格特质上的得分较高。

3.2 个人奋斗、人格特质与主观幸福感观测变量之间的相关分析

对个人奋斗、人格特质及生活满意度的观测变量进行相关分析，结果（见表 2）表明，生活满意度、积极情感、消极情感几个变量都与人格特质的五个维度、个人奋斗的各个维度有显著相关。神经质与生活满意度、积极情感的相关都为负，与消极情感的相关为正，而外向性则正好相反。开放性、随和性、尽责性 3 个人格维度都与生活满意度、积极情感正

表 2　大学生人格特质、个人奋斗、主观幸福感的描述性统计及相关矩阵（ N = 533 ）

	1	2	3	4	5	6	7	8	9	10	11	12	13
1.N													
2.E	−0.42**												
3.O	−0.14**	0.18**											
4.A	−0.35**	0.21**	0.08										
5.C	−0.35**	0.29**	0.23**	0.34**									
6.i	−0.07	0.10*	0.03	0.11*	0.12**								
7.d	0.09	−0.12**	0.01	−0.04	−0.11**	0.10*							
8.c	−0.21**	0.17**	0.14**	0.06	0.25**	0.28**	−0.08						
9.p	−0.23**	0.20**	0.15**	0.11*	0.25**	0.32**	−0.31**	0.52**					
10.a	−0.27**	0.19**	0.12**	0.13**	0.28**	0.22**	−0.30**	0.39**	0.56**				
11.ls	−0.48**	0.49**	0.27**	0.39**	0.49**	0.15**	−0.20**	0.30**	0.38**	0.38**			
12.pa	−0.40**	0.34**	0.13**	0.18**	0.29**	0.03	−0.17**	0.24**	0.23**	0.25**	0.46**		
13.na	0.50**	−0.27**	−0.03*	−0.21**	−0.28**	−0.10*	0.13**	−0.16**	−0.17**	−0.19**	−0.40**	−0.52**	
M	2.75	3.18	3.37	3.61	3.42	4.58	3.60	3.81	3.60	3.21	4.81	3.64	2.49
SD	0.61	0.49	0.45	0.37	0.50	0.43	0.53	0.69	0.59	0.56	0.66	0.82	0.76
ALPHA	0.83	0.80	0.86	−0.86	0.84						0.88	0.86	0.80

注：N= 神经质，E= 外向性，O= 开放性，A= 随和性，C= 尽责性，i= 个人奋斗的重要性，d= 个人奋斗的难度，c= 个人奋斗的清晰性，p= 个人奋斗的成功可能性，a= 个人奋斗的以往成就，ls= 生活满意度，pa= 积极情感，na =消极情感；** 表示 $p < 0.01$ ，* 表示 $p < 0.05$ 。

相关，而与消极情感负相关。生活满意度、积极情感只与个人奋斗的难度负相关，与个人奋斗的其余 4 个维度相关都为正，而消极情感正好与之相反。

从表 2 中可以看到，生活满意度与积极情感、消极情感的相关分别为 0.46、−0.40，且积极情感与消极情感之间有显著的负相关（−0.52）。根据邱林的研究，因子结构的层次性成立必须有两个条件：(1) 一阶因子定义良好，它们相互之间能被区分开来；(2) 一阶因子之间有中等程度的相关，从而使它们有被一个共同的二阶因子解释的可能性。由以上相关分析可以说明，主观幸福感可以作为生活满意度、积极情感和消极情感 3 个维度的高阶因子。

3.3 各量表的信效度

由表 2 可以看出，"五因素"人格问卷、生活满意度量表和情绪量表的测量问卷具有良好的信度，Cronbach（系数均在 0.80 以上。其中，"五因素"人格问卷五个维度的 Cronbach α 系数分别为：神经质 0.83、外向性 0.80、开放性 0.86、随和性 0.86、尽责性 0.84；生活满意度量表的 α 系数为 0.88；情绪量表的积极情感维度的 Cronbach α 系数为 0.86，消极

情感维度的 Cronbach α 系数为 0.80。这说明，"五因素"人格问卷、生活满意度量表和情绪量表均具有较好的测量信度，可以作进一步的统计分析。

鉴于个人奋斗问卷是一个半开放式问卷，每个维度的分数是由被试所列出的 5 条不同的个人奋斗平均而得出的，而每个维度下的个人奋斗项目之间的同质性可能较低，因而不能计算个人奋斗的 Cronbach 系数。本研究采用专家判断法求得个人奋斗问卷的内部效度。由 2 名心理学家及 4 名心理学专业的研究生根据个人奋斗的定义对问卷的项目进行判断，均认为本问卷基本能测量所要测量的个人奋斗特征，该问卷具有良好的内容效度。

以上分析为变量间假设关系的存在提供了初步证据，使我们得以对它们之间的关系作进一步的探讨。

3.4 人格特质、个人奋斗与主观幸福感之间的关系

根据已有的相关文献，本研究建构了有关人格特质、个人奋斗、主观幸福感三者之间关系的两个模型。为了检验理论构想与实际情况的符合程度，本研究采用 LISREL8.30 统计分析软件对研究结果进行验证性因素分析和模型比较，选取最佳模型。

本研究为了使模型路径简化，用"人格特质"这一因子代表 5 个人格维度。但要说明的是，人格特质并不是 5 个人格维度的高阶因子，因为人格特质并不具有实质意义。在此，外向性、随和性、开放性、尽责性、神经质五种人格特质既是人格特质的观测变量，并具有实际意义。

（一）模型一：人格特质和个人奋斗为外生潜变量，主观幸福感为内生潜变量；人格特质和个人奋斗分别直接影响主观幸福感；人格特质的观测变量为

从表 3 可以看出，模型一和模型二的拟合指数完全相同。侯杰泰和温忠麟将这种用同样个数参数所产生的，与样本数据有相同拟合程度、但结构不同的模型称为等同模型（equivalent models）。等同模型中各个不同模型可能含有非常不同的意义。虽然模型一与模型二等同，但两个模型的意义不同。模型一表示主观幸福感受人格特质和个人奋斗的影响。模型二表示主观幸福感受人格特质和个人奋斗的直接影响，同时人格特质经由个人奋斗间接影响主观
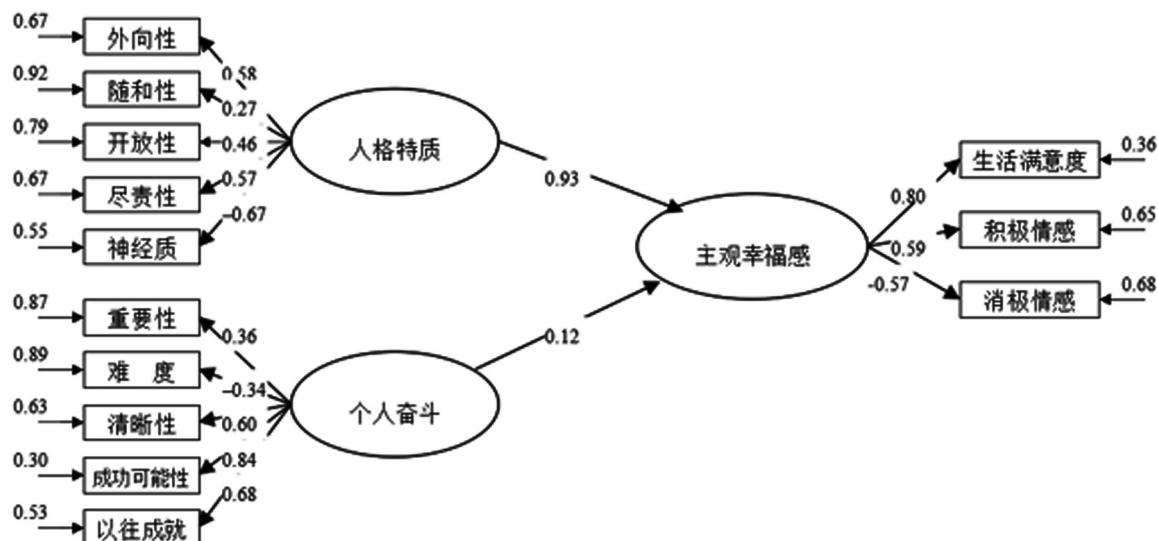
图 1 模型一的验证性因素分析结果

表 3 模型一和模型二验证性因素分析的拟合指数

模型	x^2	df	x^2/df	RMSEA	SRMR	GFI	IFI	CFI	NNFI
M1	263.17	62	4.25	0.08	0.06	0.93	0.89	0.88	0.85
M2	263.17	62	4.25	0.08	0.06	0.93	0.89	0.88	0.85

注：M1 表示模型一，M2 表示模型二。

外向性、随和性、开放性、尽责性、神经质；个人奋斗的观测变量为重要性、难度、清晰性、成功可能性、以往成就；主观幸福感的观测变量为生活满意度、积极情感、消极情感。模型一的验证性因素分析结果，见图 1，其拟合度指标见表 3。

（二）模型二：人格特质为外生潜变量，个人奋斗和主观幸福感为内生潜变量；人格特质直接影响主观幸福感，另外，人格特质还通过个人奋斗间接影响主观幸福感；各潜变量的观测变量与模型一相同。模型二的验证性因素分析结果见图 2，其拟合度指标见表 3。

幸福感。

由于等同模型的拟合指数相同，此时无法依靠指标来评估模型的拟合优度。但我们也不能说等同模型就毫无价值。Lee 和 Hershberger 提出，对等同模型应谨慎，应留意其存在，并检视其它模型。侯杰泰和温忠麟也提出，需审视各模型所表达的意义，依据学科理论选取有意义的模型。

根据 little 的人格社会生态学模型，人格特质可以影响个人奋斗。因此，依据该理论以及模型二的意义，本研究将选取模型二作为最佳模型。从表 3 可
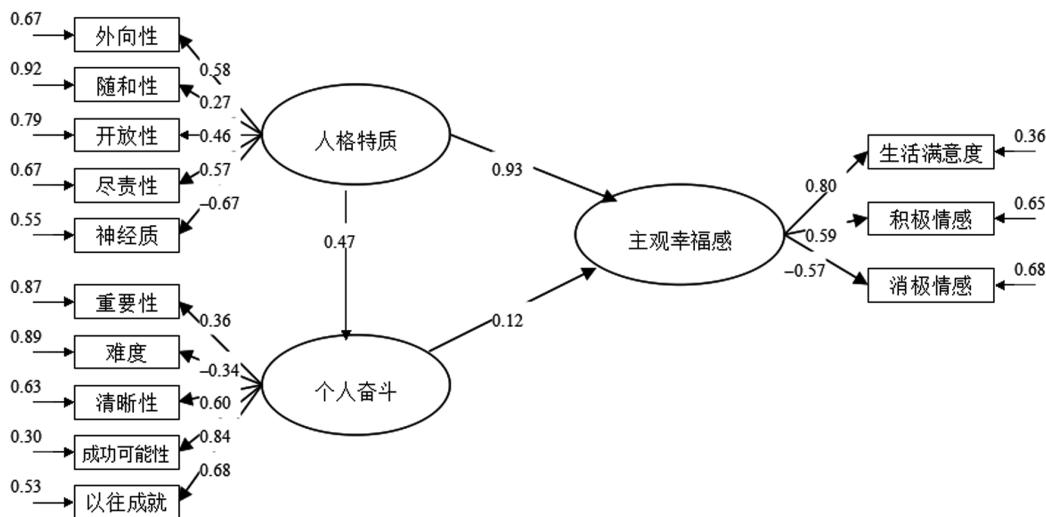
图 2　模型二的验证性因素分析结果

以看出，虽然模型二的 x^2/df 大于 3，但小于 5，说明模型整体上还可以接受；模型二的 RMSEA 小于 0.08，达到了临界标准；GFI 达到了 0.90 的临界标准；虽然 IFI、CFI、NNFI 没有达到 0.90 的临界标准，但也非常接近。可见，模型二是一个较好的模型。

由图 2 可以看出，通过比较外生变量和内生变量之间的通径系数，人格特质对主观幸福感的影响要大于个人奋斗对主观幸福感的影响。此外，人格特质对个人奋斗有显著的影响作用。

4　讨论

4.1 个人奋斗、人格特质对主观幸福感的影响分析

目前，国内对主观幸福感的研究主要是从人格特质的角度着手，而从动机的角度入手的却较少。虽然人格特质对主观幸福感具有强大的预测力，但并不能对主观幸福感作出完整的预测与解释，其原因既有理论上的问题，也有方法上的问题。从理论上讲，人格特质只是人格的基本单元之一，不能反映人格这个复杂的系统。随着主观幸福感研究的深入，仅仅从人格的特质取向来研究势必不能全面地陈述人格和主观幸福感之间的复杂关系。

本研究就是针对这一局限，同时从人格心理学中的特质取向和动机取向来研究大学生的主观幸福感，以从静态和动态两种角度全面研究主观幸福感，从而丰富人格与主观幸福感之间关系的研究。

根据已有理论，本研究建立了两个有关人格特质、个人奋斗与主观幸福感三者之间关系的结构方程模型。结果表明这两个模型是等同模型，各项拟合指数都相等。在此种情况下，仅凭数据无法判定哪个模型更好。根据 Lee 和 Hershberger、侯杰泰和温忠麟的观点，应根据各模型所表达的意义，以及相关理论选取有意义的模型。本研究根据已有理论选择了模型二作为最佳模型。首先，根据 little 的人格的社会生态学模型，人格特质可以影响个人奋斗。其次，本质上，个人奋斗是个人行动建构单元之一，也即动机单元之一。因此，由一系列目标所组成的个人奋斗实则是动机的外显表现形式。从特质与动机的关系来看，特质代表行为的心理生理机制，规定动机以一致性的方式表现出来。所以，人格特质会直接影响个人奋斗。根据以上理论以及模型二的意义，可以确定模型二为最佳模型。

根据结构方程模型，发现人格特质对主观幸福感的影响要远远大于个人奋斗对主观幸福感的影响。这一点与 Emmons、Diener 和 Lucas 的结论相反。

但研究结果支持了我们所熟知的观点，即，正如 Diener 所言，在预测幸福感时，人格特质因素即使不是最好的预测指标，至少也是最可靠、最有力

的预测指标之一。因此，心理学者可以培养有利于人们幸福生活的人格特征，从而促进和提高人们的主观幸福感水平。

4.2 个人奋斗对主观幸福感的影响作用

研究结果表明，个人奋斗对生活满意度、积极情感、消极情感都有显著的预测力，但不同维度的影响作用又不相同。进一步分析发现，个人奋斗已经取得的成就，如何在个人奋斗上取得成功，以及其思路的清晰性都对主观幸福感的三个成分有显著的影响作用。

就消极情感来看，个人奋斗的以往成就和清晰性都具有显著的负向预测力。当大学生在以往成就上较为失败时，则会体验到较多的消极情感；而当过去在个人奋斗上越是成功，则体验到的消极情感就越少。另外，当大学生对在个人奋斗上取得成功的思路不清晰时，会体验到较多的消极情感。这可能是因为个人奋斗的清晰性会影响到个体的自我控制感，当大学生对于未来在个人奋斗上取得成功的思路不清晰时，也就不相信自己能控制个人奋斗的发展，不相信自己能达成个人奋斗，因而会有较多的消极情感体验。

在生活满意度上，大学生在个人奋斗上已经取得的成就、成功可能性和清晰性都具有显著的正向预测力。当大学生以往在个人奋斗上越成功，或预期自己未来较有可能在个人奋斗上取得成功时，或对在个人奋斗上取得成功的思路越清晰，对生活的满意度就越高。

在积极情感上，清晰性、以往成就，成功可能性有显著的正向预测力。本研究发现，以往成就的获得能导致积极情感的增加，这一结论与Emmons的研究结论一致。Emmons还提出，这一现象可以用Bandura的自我效能理论来解释。Bandura认为，以往成功的经验会使个体产生较强的自我效能感或胜任感，从而提高积极情感体验。此外，个人奋斗的难度对积极情感有显著的负向预测作用，即当大学生认为自己难以在个人奋斗上取得成功时，会体验到较少的积极情感。

4.3 人格特质对主观幸福感的影响作用

本研究发现，人格特质对主观幸福感的三种成分都有显著的预测作用。神经质对生活满意度、积极情感都有显著的负向预测作用。由于在神经质上得高分的人，情绪不稳定，往往会对微小的挫折和问题情境产生强烈的情绪反应，而且需要很长一段时间才能平静下来。因此，神经质得分高的人可能较少体验到积极情感。神经质对消极情感有显著的正向预测力，在神经质上得分高的人往往更容易体验到消极情感。因为这些人往往比一般人更易激动、动怒和沮丧。这一点验证了Costa和McCrae关于神经质与消极情感相关密切的研究结论。

外向性、随和性、开放性、尽责性都对生活满意度、积极情感有正向预测作用，而对消极情感有负向预测作用。这一点与我们的常识相一致，一般来说，外向性上得分高的人往往更为快乐，更易于体验到积极的情感，对生活更为满意。而在尽责性上得分高的人，因为其生活有条理、勤奋而有抱负，这种人一般会过上较好的生活，因而会体验到更多积极情感。

4.4 本研究的局限性与未来研究的方向

本研究采用结构方程建模的方法，较为系统地考察了人格特质和个人奋斗对大学生主观幸福感的影响作用，以比较人格的特质因素与动机因素哪种对主观幸福感的影响作用更大，并得出了一些有意义的研究结果。当然，本研究也有一些不足，需要在未来的研究中加以完善。

首先，对个人奋斗的测量上尚有不足。个人奋斗问卷是一个半开放式问卷，要求被试对自己所列出的个人奋斗进行评价。因此，个人奋斗问卷在测量过程中能否测到我们想测量的个人奋斗，主要依靠于被试对个人奋斗概念的理解。未来在个人奋斗研究中，如何找到合适的测量方法，有待进一步的研究。

在收回的个人奋斗问卷中，我们发现，大学生所列出的个人奋斗主要集中于学习、交友、恋爱、锻炼能力等方面。有些学生所列的个人奋斗有较强的家庭倾向，如更好地孝顺父母、成立美满家庭等，

这些都反映了中国传统文化的影响，也反映了中国大学生具有较强的人际关系取向。由此我们可以考虑：能否去收集大学生的个人奋斗项目，从这些个人奋斗项目中找到一些规律性的东西，并编制一个中国大学生个人奋斗问卷。因此，如何有效地在中国文化背景中运用个人奋斗构念，将是未来研究的努力方向之一。

其次，在考察人格特质、个人奋斗与主观幸福感三者之间的关系时，本研究只建构了两个结构模型。侯杰泰等认为，在面对等同模型时，研究者应根据有关理论列举多个可能模型，并采用模型比较法选取最佳者。本研究没有列出所有可能的模型，因此所选择的模型只是较好的，而不是最佳的，这一点有待进一步研究。

5　结论

本研究获得了以下结论：

（1）人格特质与个人奋斗都对主观幸福感有显著影响。

（2）人格特质对主观幸福感的影响作用要大于个人奋斗对主观幸福感的影响作用，即相对于人格的动机因素，人格的特质因素对主观幸福感有更强的预测力。

（3）人格特质对个人奋斗有显著的影响作用。

参考文献

侯杰泰，温忠麟，成子娟. (2005). 结构方程模型及其应用. 北京：教育科学出版社，143– 145.

李育辉，张建新. (2004). 中学生人格特质、主观应激与应对风格之间的关系. *心理学报，36*(1), 71–77.

Pervin L A, Oliver P J 著. 黄希庭主译. (2003). 人格手册：理论与研究，华东师范大学出版社，677–682.

邱林，郑雪. (2005). 主观幸福感的结构及其与人格特质的关系. *应用心理学，4*, 21–26.

杨子云，郭永玉. (2004). 人格分析的单元 –– 特质、动机及其整合. *华中师范大学学报 (人文社会科学版)，43*(6), 131–135.

张兴贵，何立国，郑雪. (2004). 青少年学生生活满意度的结构和量表编制. *心理科学，27*(5), 1257–1260.

张钊，郭永玉. (2006). 个人奋斗及其相关研究，*心理科学进展，14*(6), 950–955.

Baumann N, Kaschel R, Kuhi J. (2005). Striving for unwanted goals:stress–dependent discrepancies between explicit and implicit achievement motives reduce subjective well–being and increase psychosomatic symptoms. *Journal of Personality and Social Psychology, 89*(5), 781–799.

Cantor N. (1990). From thought to behavior: "having" and "doing" in the study of personality and cognition. *American Psychologist, 45*(6), 735–750.

Diener E, Fujita F. (1995). Resources, personal strivings, and subjective well–being: a nomothetic and idiographic approach (in Chinese). *Journal of Personality and Social Psychology, 68*(5), 926–935.

Diener E, Suh E M, Lucas R E, Smith H L. (1999). Subjective well–being: three decades of progress. *Psychological Bulletin, 125*(2), 276–302.

Emmons R A. (1992). Abstract versus concrete goals: personal striving level, physical illness, and psychological well–being. *Journal of Personality and Social Psychology, 62*(2), 292–300.

Emmons R A, Diener E. (1985). Factors influencing satisfaction judgements: a comparative examination. *Social Indicators Research, 16,* 157–167.

Emmons R A. (1986). Personal strivings: an approach to personality and subjective well–being. *Journal of Personality and Social Psychology, 51*(5), 1058–1068.

Emmons R A. (2005). Striving for the sacred: personal goals, life Meaning, and religion. *Journal of Social Issues, 61*(4), 731–745.

Jepson E, Forrest S . (2006). Individual contributory factors in teacher stress: the role of achievement

striving and occupational commitment. *British Journal of Educational Psychology, 76*(1), 183–197.

McCrae R R, Cost a P T. (1991). Adding liege and albeit: the full five factor model and well–being. *Personality and Social Psychology Bulletin, 17*, 227–232.

McCrae R R, Costa P T, Jr del Pilar G H, Rolland J P, Parker W D. (1998). Cross–cultural assessment of the five factor model: The Revised NEO Personality Inventory. *Journal of Cross–Cultural Psychology, 29*, 171–188.

The relations among personal strivings, personality traits and subjective well-being in undergraduates

YANG Huifang[1,2], GUO Yongyu[1]

([1] Department of Psychology, Hua Zhong Normal University, and Hubei Human Development and Mental Health Key Laboratory, Wuhan 430079, China)

([2] Location of Educational Administration, Guangdong Zhanjiang Normal College, Zhanjiang 524048, China)

Abstract By applying the structural equation modeling analysis, the present study examined the relationships among personal strivings, personality traits and subjective well−being, and compared the influence of personal strivings and persibakity traits on subjective well−being. 533 undergraduates from 6 universities of Wu Han completed the NEO personality − Five Factor Inventory、Striving Assessment Scales、Students Life Satisfaction Scales、Mood Forms. Results of the study indicated as follows: both personal strivings and personality traits had influence on subjective well−being; the influence of personality traits on subjective well− being was greater than the influence of personal strivings; personality traits had influence on personal strivings. The different relations between subjective well− being and two aspects of personality, which were personal strivings and personality traits, were discussed.

Keywords personal strivings; personality traits; subjective well−being(SWB); undergraduates

评价性条件反射效应：无条件刺激的呈现时长、效价强度与关联意识的作用 *

赵 显[1,2]，李 晔[2]，刘 力[1]，曾红玲[2]，郑 健[1]

([1] 北京师范大学心理学院 , 北京市应用实验心理重点实验室 , 北京 100875)

([2] 华中师范大学心理学院 , 湖北省人的发展与心理健康重点实验室 , 武汉 430079)

摘 要 以真实商标图案为条件刺激，情绪图片为无条件刺激，探索无条件刺激呈现时长、效价强度与关联意识对评价性条件反射效应的影响。实验通过结合四图再认测验与基于项目分析，对关联意识的作用进行了详细探讨。结果表明，评价性条件反射效应只发生在无条件刺激长呈现水平与无条件刺激强效价水平；评价性条件反射效应的产生需要基于被试的关联意识。关联意识在呈现时长 (效价强度) 与评价性条件反射效应间的中介作用不显著。结果不支持评价性条件反射的内隐错误归因机制及联想 – 命题评价模型的相关论断，部分支持命题性解释模型。

关键词 态度形成；评价性条件反射；呈现时长；效价强度；关联意识

分类号 B842

1 引言

态度形成与态度改变是心理学家长期关注的一个重要课题。其中，评价性条件反射 (evaluative conditioning, EC) 效应更是近年来研究的热点。在态度的联想 – 命题评价模型 (associative–propositional evaluation, APE) 中，EC 也是其建构的核心成分 (Gawronski & Bodenhausen, 2006)。EC 指当某一个刺激 (条件刺激，conditioned stimulus, CS) 与另一个积极或消极刺激 (无条件刺激，unconditioned stimulus, US) 反复配对后，人们对原刺激 (指条件刺激) 的喜好程度发生改变的现象 (De Houwer, 2007; De Houwer, Thomas, & Baeyens, 2001)。英文俗语" 杀死报信者" (shooting the messenger) 即与此相关：报信者似乎难以避免地与他们所传递信息的效价相关联，坏消息的报信者也获得了坏消息的消极效价，从而招来负面评价。与经典条件反射 (Pavlovian conditioning, PC) 不同，EC 是人们对喜欢与不喜欢的学习，习得的是一种选择偏好，并且具有抗消退等特点 (De Houwer, et al.2005,2001; Walther, 2002; Walther, Nagengast, & Trasselli, 2005)。

EC 效应的研究者建构了不同的理论模型。

新近提出的有内隐错误归因机制 (Jones, Fazio, & Olson, 2009) 与命题性解释模型 (De Houwer, 2007,2009,2011;De Houwer et al., 2005; Mitchell, De Houwer, & Lovibond, 2009)。内隐错误归因机制认为 EC 的产生是 US 刺激唤起了一个特定的评价，被试因为来源混淆感 (source confusability) 而错误且内隐地将 US 所产生的评价性反应归因于 CS 的过程。CS 与 US 刺激图片的相对大小、空间临近程度和效价强度都会调节 CS 相比于 US 的凸显程度，影响内隐错误归因过程，最终影响 EC 效应。EC 效应不需要依赖注意与意识，可以在无意识水平上发生。此观点与 APE 模型对 EC 的相关论断类似 (Gawronski & Bodenhausen, 2006)。命题性解释模型认为，EC 效应中 CS 效价的改变是因为个体有意识地形成了 CS 与 US 配对的命题，并且个体通过判断 CS–US 命题真假来确定他们对 CS 的偏好。EC 效应的产生一定要基于被试对刺激进行有意识的加工，需要一定的注意水平。

可以看出，这两种自下而上建构的模型存在很大程度的对立，而两种理论模型框架下的实验结果差异可能来自不同的实验范式。相关实验研究主要集中在探索注意与意识对 EC 效应的影响上。对于 US 注意时长 (刺激呈现时长) 这个因素，在阈上呈现条件下，研究者普遍选取 1000ms 或以上来同时或继时呈现 US 刺激 (Olson & Fazio, 2001, 2002，2006; Pleyers, Corneille, Luminet, & Yzerbyt, 2007; Pleyers, Corneille, Yzerbyt, & Luminet, 2009; Stahl, Unkelbach, & Corneille, 2009; Walther, 2002; Walther, Gawronski, Blank, & Langer, 2009; Walther & Nagengast, 2006)；而在阈下呈现条件下通常选取一个屏幕刷新周期的时间，例如 17ms(Dijksterhuis, 2004; Field & Moore, 2005; 陈进，梁宁建，杨福义, 2009)，这中间缺乏对 US 阈上短时长呈现的讨论，而被试的注视时长或刺激的呈现时长是影响知觉加工的重要因素，例如 Chua, Boland 和 Nisbett（2005）发现，在阈上 3000ms 呈现条件下，0ms–420ms 时间段与 420ms–3000ms 时间段相比，被试对图片中客体与背景的注视选择会有差异。阈上短

时长是从阈上到阈下的过渡，这段呈现时长的 EC 效应支持哪种理论模型尚缺乏相应的实验支持。从内隐错误归因机制推论，首先，在 CS 与 US 配对过程中，如果减少阈上 US 的呈现时间，CS 的呈现时间不变，这使得 CS 更为凸显，应该有助于内隐错误归因的产生，从而增大 EC 效应；其次，如果真正促使 CS 效价发生转变的是 US 图片效价而非 US 图片内容 (Stahl, et al., 2009)，那么被试要获取图片效价，可能只需较短时间，因而缩短 US 的呈现时长不影响 EC 效应的大小，即在低于 1000ms 的阈上水平呈现 US 也可产生与长时间呈现 US 相同甚至更大的 EC 效应。但从命题性解释模型推论，EC 效应依赖于被试的注意水平，那么阈上短时间呈现 US 会导致被试注意时间过短，EC 效应会降低甚至不发生。基于命题性解释模型，本研究假设：在阈上水平，长呈现时长的 US 相比于短呈现时长的 US 会产生更大的 EC 效应。

对于 US 的效价强度，Jones 等人 (2009) 的研究发现效价强度更弱的刺激相比于强效价的无条件刺激，更容易引发条件刺激产生 EC 效应，他们认为此结论符合内隐错误归因机制。而在 Baeyens Crombes, Van den Bergh 和 Eelen(1988) 早期研究中，强效价的 US 与中等效价的 US 产生了相近大小的 EC 效应。但在 APE 模型中，EC 被视为一个联想学习过程 (Gawronski & Bodenhausen, 2006)，由此推测，强效价刺激应该能激活更强的自动情感反应 (automatic affective reactions) 并形成联想评价 (associative evaluations)，进而作用于外显态度。在这个问题上，目前仅有两篇实验报告，再无额外的例证，而 Jones 等人 (2009) 的研究是支持内隐错误归因机制的关键实验之一。效价强度对于刺激知觉与加工十分重要，这个因素也可能与注意时长存在交互作用，共同影响 EC 效应。命题性解释模型中虽然对此无针对性的论述，但可以推论强效价 US 相比于弱效价 US 会在 CS 与 US 之间建立更为牢固的命题关系。因此，本研究假设：强效价的 US 相比于弱效价的 US 会产生更大的 EC 效应。

对于 EC 效应是否依赖被试的关联意识 (被试有意识地记住 CS 与 US 的配对关系，contingency

awareness)，存在更多的争议。内隐错误归因解释机制认为，如果被试存在关联意识，那么这种意识会使得 US 及其效价更为凸显，进而会降低错误归因的可能性，从而减小 EC 效应，因此 EC 效应的产生并不需要关联意识 (Baeyens, Eelen, & Bergh, 1990; Fulcher & Hammerl, 2001; Gibson, 2008; Hammerl & Fulcher, 2005; Walther et al., 2009; Walther & Nagengast, 2006)。命题性解释模型则认为，关联意识是形成命题必不可少的基础，对这一观点有大量实验结果支持 (Dawson, Rissling, Schell, & Wilcox, 2007; Lipp & Purkis, 2005; Lovibond & Shanks, 2002; Pleyers et al., 2007,2009; Purkis & Lipp, 2001; Stahl & Unkelbach, 2009; Wardle, Mitchell, & Lovibond, 2007)。然而，对关联意识与其他因素共同作用 EC 效应的发生机制，以往研究缺乏整合性的探讨。根据命题性解释模型，本研究假设：形成强关联意识的被试相比于弱关联意识的被试，会产生更大的 EC 效应。呈现时长与效价强度通过关联意识的中介作用影响 EC 效应。

在方法上，以往关联意识的研究普遍采用的是以被试为单位的统计分析方法，Pleyers 等 (2007) 建议，采用以项目 (基于被试对 CS–US 配对所产生的每个反应项目) 为单位来进行数据分析，以提高数据分析的利用率和敏感性。近年来，已有研究者基于项目分析来深入探索关联意识的作用 (Pleyers, et al., 2007, 2009; Stahl & Unkelbach, 2009; Stahl, et al., 2009)。此外，Walther 与 Nagengast(2006) 提出可以通过四图再认测验的方法来检测被试的关联意识。所谓四图再认测是指在每一个再认题目中，屏幕左侧呈现每一个配对中的 CS，而右侧呈现四个小图片 (与 CS 匹配过的 US，与 US 相同效价的刺激，与 US 相反效价但程度相同的刺激，与 CS 效价接近的中性刺激)。四个小图片在右侧呈现的位置随机；要求被试要从中选择匹配过的 US。这种测验方法有助于将被试作答进行细致分类并对关联意识进行深入探讨。然而，在 Walther 与 Nagengast 的研究中，他们仅以被试为单位进行统计分析，发现 EC 效应只在无关联意识的被试上发生。这个结果与 Pleyers 等人 (2007) 的结果

不一致，而且其研究的统计检验力也受到了其他研究者的质疑 (Stahl & Unkelbach, 2009)。目前，尚没有研究者将四图再认测验与基于项目分析的方法结合起来使用。此外，很多研究者认为，被试可能会猜测出实验目的，并根据与之配对过的 US 效价对条件反射后的 CS 进行评价 (即需求意识，demand awareness) (De Houwer et al., 2001)。同时，在检测 CS–US 关联意识的再认测验中，被试也可能从 CS 的喜好感受 (评价) 来推测与之匹配过的 US(即推论策略，inference strategy) (Bar–Anan, De Houwer &No sek, 2010; Pleyers, et al., 2007; Stahl, et al., 2009)。针对这些问题，本研究拟通过综合使用不同方法来进一步检验与意识相关的变量对 EC 效应的影响。

综上，本研究将探索 US 呈现时长、US 效价强度对 EC 效应的影响，并结合四图再认测验与基于项目分析检验关联意识的作用，探讨这些自变量共同影响 EC 效应的机制。实验为三因素混合设计：US 呈现时长 (长呈现，短呈现) × US 效价强度 (强效价，弱效价) × CS 类型 (与积极图片匹配过的 CS+，与消极图片匹配过的 CS–)。其中 CS 类型自变量为被试内变量，另两个自变量为被试间变量。

2 方法

2.1 被试

本实验被试选取 122 名本科生及研究生，其中男生 38 人。被试报告无色盲、色弱。被试被随机分配到四个实验组中 (US 短呈现强效价组、US 长呈现强效价组、US 短呈现弱效价组、US 长呈现弱效价组)。实验结束时，每位被试获得一份礼物。

2.2 实验材料

2.2.1 正式实验材料 　正式实验材料包括商标图片、中性情绪图片、积极情绪图片 (包括强积极与弱积极)、消极情绪图片 (包括强消极与弱消极) 各 8 张，共计 32 张图片。CS(商标刺激) 与 US(积极 – 消极图片刺激) 进行拉丁方平衡化匹配，以此避免主试基于主观知觉的刺激特性 (比如，知觉流畅性) 对 CS 和

US 进行配对而带来混淆因素 (Field & Davey, 1999)。最终组成 16 个 CS–US 刺激对，强效价组 8 对，弱效价组 8 对。每个效价组内，被试被随机分配到 8 种 CS–US 配对关系中。

2.2.2 实验材料的选取与评定 图片刺激一部分采自国际情绪图片系统 (International Affective Pictures System, IAPS)。参照 IAPS 图片系统中愉悦度维度并排除掉明显西方色彩的图片、色情图片和重度伤残图片，从中挑选出 69 张图片作为积极 – 消极 – 中性效价刺激候选图片 (Lang, Bradley, & Cuthbert, 2008)。从互联网上挑选 11 张图片，作为补充。从中国商标网 (sbj.saic.gov.cn/sbgg/sbgg2.asp《商标公告》) 上选取黑白商标 20 张，作为 CS 刺激候选图片。这些商标全部采自大学生比较陌生的产品类别 (如农用器械等)，商标不包含文字，图片大小均为 768×576 像素。

预研究选取 48 位本科生和研究生，对 80 张情绪图片和 20 张商标图片的效价进行评定，以区分出强效价、弱效价图片，并对商标效价进行检验。评定者在李克特 11 点量表 (–5, +5) 上对情绪图片的效价进行评定，在 5 点量表 (–2, +2) 上对商标图片的效价进行评定。其中一半评定者先评定情绪图片，另一半先评定商标图片。主试将情绪图片效价范围转换为 1–11，商标图片效价范围转换为 1–5。对 80 张情绪图片和 20 张商标图片的评定结果在性别差异上进行独立样本 t 检验，发现有 8 张图片存在明显的性别评定差异，删去此 8 张图片。对商标图片的评定顺序差异 (先评定商标图片还是先评定情绪图片) 进行独立样本 t 检验，发现有 1 张图片存在明显的评定顺序差异，删去此张图片。商标图片从 2.5–3.5 效价范围内选取，选取的 8 张商标图片效价均值为 3.22。选取后的每张商标图片与理论中值 3 相比，进行 t 检验，发现在 8 张商标图片中，有 2 张图片的效价与理论中值存在显著差异 (M_1=3.40，t_1=2.72, p=0.009; M_2=3.41, t_2=2.29, p=0.027)。在权衡了研究的生态效度后，最终保留了这两张图片。

参考 80 张情绪图片的效价分布，选出中性情绪图片、积极情绪图片、消极情绪图片各 8 张，作为正

式实验材料。对五个情绪图片组中相邻两组效价的平均数进行差异检验，以保证效价强度自变量有效，结果表明差异均极其显著：强消极 – 弱消极，$t(45) = -13.77$, $p<0.001$, Cohen's d =2.04；弱消极 – 中性，$t(44) =-11.20$, $p<0.001$, Cohen's d=2.13；中性 – 弱积极，$t(43) =-7.64$, $p<0.001$, Cohen's d=1.40；弱积极 – 强积极，$t(44) =-11.23$, $p<0.001$, Cohen's d =1.86。将每张情绪图片的效价减去 6 并转换成正数，进行强积极与强消极组之间、弱积极与弱消极组之间的效价平均数差异检验，发现它们之间仅效价方向不同，效价强度差异不显著：强积极 – 强消极，$t(43) = 0.10$, p=0.923, Cohen's d =0.02；弱积极 – 弱消极，$t(46)=-0.28$, p=0.777, Cohen's d =0.05。结果表明实验材料满足本实验要求。

2.3 实验程序

实验程序由四部分组成：条件反射阶段、评价测量阶段、再认测验阶段、填答问卷阶段。前三个阶段被试在计算机上完成，所有刺激均呈现在 17 寸计算机显示器上，被试通过按键和点击鼠标进行操作。屏幕刷新频率为 85HZ，显示器亮度、对比度和色彩均为统一设置，被试距离屏幕约 0.5 米。第四阶段被试填答纸质问卷。

2.3.1 条件反射阶段 CS 与 US 系列呈现，并采用 EC 研究中的普适程序：前行条件反射程序 (forward–conditioning)，即 CS 先于 US 呈现 (De Houwer, et al., 2001)。实验中，每对 CS–US 随机呈现 5 次。在每次呈现过程中，屏幕上先呈现注视点 500ms，然后呈现 CS 1000ms，继而呈现 US。短呈现组 US 呈现时间为 120ms，然后呈现掩蔽刺激 880ms；长呈现组 US 呈现时间为 1000ms。每对 CS–US 间隔空屏时间为 1500ms。

2.3.2 评价测量阶段 研究使用外显评价测量来收集被试对商标的评价，此阶段屏幕上会随机呈现商标图片。在商标图片下方呈现 (–3, +3) 的李克特量表 (–3 表示非常消极，+3 表示非常积极)。被试通过鼠标点击量表上对应的区块对商标图片进行积极 – 消极评价，每张图片仅评价一次。实验后，将数据

转换成 1–7 点计分。

2.3.3 再认测验阶段 再认测试采用 Walther 与 Nagengast(2006) 提出的四图再认测验。除中性刺激，所有图片均在之前的任务中呈现过。被试使用鼠标从四个小图片中选出与 CS 配对过的 US。

2.3.4 填答问卷阶段 要求被试写出他所认为的本实验目的，以此探测被试的需求意识程度。

3 结果

9 名被试没有完成实验，剔除后使用 SPSS11.5 软件进行基于被试的统计分析。

3.1 EC 效应

对两类 CS(CS+、CS–) 评价数据进行配对样本 t 检验，发现 CS 类型效应显著，M(CS+)= 4.49, M(CS–) = 4.06, $t(112)$ = 3.65, $p < 0.001$, $Cohen´s\ d$ = 0.51。基于以往研究的统计分析标准 (Olson & Fazio, 2001; Pleyers et al., 2007，2009;Walther & Nagengast, 2006)，结果表明实验在总体水平上发生了显著的 EC 效应，即与积极图片匹配过的商标的评价显著高于与消极图片匹配过的商标的评价。

3.2 性别及需求意识对 EC 效应的影响

对两类 CS 评价数据进行 MANOVA 2(性别) × 2(CS 类型) 的重复测量方差分析，以考察性别对 EC 效应的影响。发现性别主效应不显著，$F(1,111)=0.00$, p = 0.986, $\eta^2_p < 0.001$。性别与两类 CS 评价的交互作用不显著 $F(1,111) < 0.01$，p=0.982，$\eta^2_p < 0.001$，即性别对 EC 效应无显著影响。主试对问卷结果进行评定，发现仅有 14 个被试猜出了实验目的，占有效数据的 12.2%。将这 14 个被试归类于"有需求意识"，其余被试归类于"无需求意识"。对两类 CS 评价数据进行 MANOVA 2(需求意识程度) × 2(CS 类型) 的重复测量方差分析。需求意识程度主效应不显著 $F(1,111) < 0.01$，p=0.964，$\eta^2_p < 0.001$；需求意识与两类 CS 评价的交互作用也不显著，$F(1,111)=0.85$, p=0.358, $\eta^2_p < 0.01$，即总体上需求意识对 EC 效应无显著影响。

3.3 呈现时长、效价强度与关联意识对 EC 效应的影响

根据四图再认测验结果，统计四种选项的分布情况（"正确 US""相同效价 US""相反效价 US""中性刺激"）。发现被试有超过一半的选择都能将正确 US 选中 (58%)，排在第二是选择相反效价 US(23%)，其次是相同效价 US(14%)，最后是中性刺激 (4%)。将正确识别出 5 个及 5 个以上配对过的 US 的被试归类于"强关联意识"(n_1=57)，将识别出 4 个及 4 个以下的被试归类于"弱关联意识"(n_2=56)。

为了考察呈现时长、效价强度、关联意识对 EC 效应的影响，进行 MANOVA 2(CS 类型) × 2 (呈现时长) × 2(效价强度) × 2(关联意识) 四因素重复测量分析。发现 CS 类型主效应显著，$F(1,105)$ =13.92, $p<0.001$, $\eta^2_p < 0.12$；CS 类型与呈现时长的交互作用显著，$F(1,105)=11.72$, $p=0.001$, $\eta^2_p < 0.10$；CS 类型与效价强度的交互作用显著，$F(1,105)=7.29$, $p=0.008$, $\eta^2_p < 0.07$；CS 类型与关联意识的交互作用显著，$F(1,105)=15.33$, $p<0.001$, $\eta^2_p < 0.13$;效价强度与关联意识的交互作用显著，$F(1,105)$ =6.10, $p=0.015$, $\eta^2_p < 0.06$。其他主效应与交互作用均不显著。下面逐个
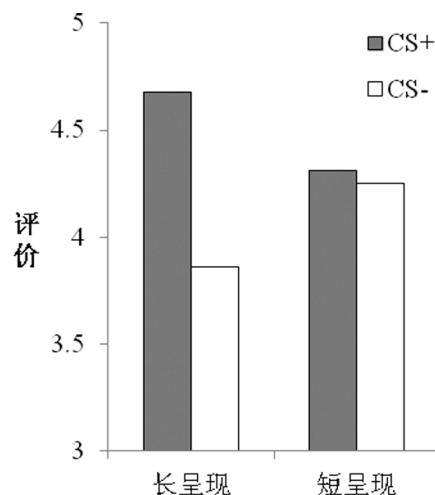
图 1 呈现时长对 EC 效应的影响

分析交互作用显著项的简单效应。

3.3.1 CS 类型与呈现时长的交互作用 简单效应分析发现，在长呈现水平上，CS 类型效应显著，M (CS+) =4.68，M (CS–)=3.86，$F(1,111)=25.19$,

$p<0.001$，$\eta_p^2=0.21$；在短呈现水平上，CS 类型效应不显著，$M(CS+)=4.31$，$M(CS-)=4.25$，$F(1,111)=0.14$，$p=0.706$，$\eta_p^2<0.01$（图 1）。说明当 US 呈现时间为 120ms 时，EC 效应并没有发生，而在 1000ms 的长呈现水平上发生了 EC 效应。

3.3.2 CS 类型与效价强度的交互作用　简单效应分析发现，在强效价水平上，CS 类型效应显著，$M(CS+)=4.57$，$M(CS-)=3.93$，$F(1,111)=14.57$，$p<0.001$，$\eta_p^2<0.16$；在弱效价水平上，CS 类型效应不显著，$M(CS+)=4.42$，$M(CS-)=4.19$，$F(1,111)=1.88$，$p=0.173$，$\eta_p^2=0.01$（图 2）。说明实验只在强效价水平上发生了 EC 效应。
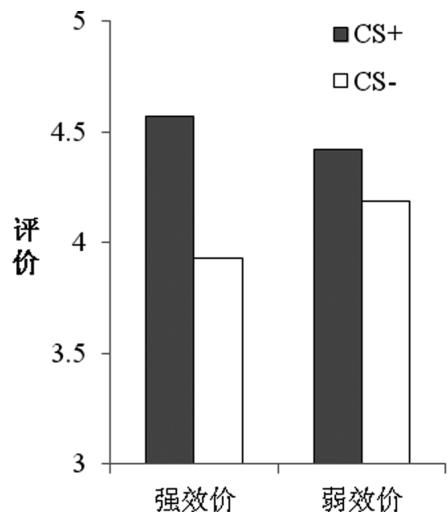
图 2　效价强度对 EC 效应的影响
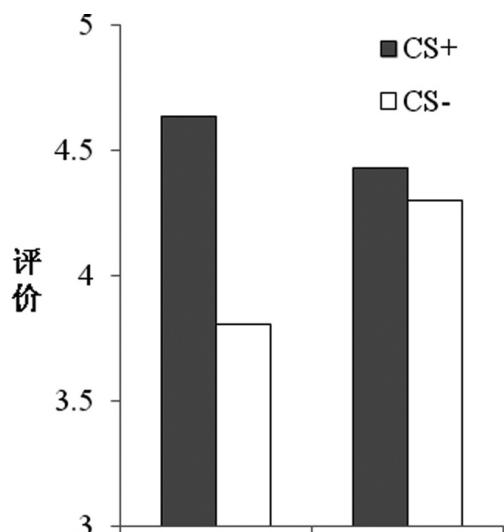
图 3　关联意识对 EC 效应的影响

3.3.3 CS 类型与关联意识的交互作用　简单效应分析发现，在强关联意识水平上，CS 类型效应显著，$M(CS+)=4.64$，$M(CS-)=3.81$，$F(1,111)=26.65$，$p<0.001$，$\eta_p^2=0.22$；在弱关联意识水平上，CS 类型效应不显著，$M(CS+)=4.43$，$M(CS-)=4.30$，$F(1,111)=0.05$，$p=0.827$，$\eta_p^2<0.01$（图 3）。说明只有被试意识并记住了大部分的 CS-US 配对关系，EC 效应才能产生。继而采用基于项目的统计分析，以提高分析关联意识的统计敏感性，同样发现 CS 类型与关联意识的交互作用显著，$F(3,898)=5.73$，$p=0.001$，$\eta_p^2=0.02$，但效应量有大幅下降。

3.3.4 效价强度与关联意识的交互作用　在对曾经与积极图片匹配过的商标进行评价时，强效价水平上，关联意识的效应不显著，$F(1,110)=0.36$，$p=0.552$，$\eta_p^2=0.01$；在弱效价水平上，关联意识的效应显著，$F(1,110)=4.62$，$p=0.034$，$\eta_p^2=0.05$；在对曾经与消极图片匹配过的商标进行评价时，强效价水平上，关联意识的效应显著，$F(1,110)=14.45$，$p<0.001$，$\eta_p^2=0.15$；在弱效价水平上，关联意识的效应不显著，$F(1,110)=0.33$，$p=0.566$，$\eta_p^2=0.01$。

3.4　四图再认测验的结果与关联意识

基于项目分析，对四图再认测验的结果进行统计。分别对四图再认四种选择下的 CS 评价做 t 检验，发现仅在正确选择项目中，发生 EC 效应，$M(CS+)=4.61$，$M(CS-)=3.88$，$t(519)=-5.72$，$p<0.001$，$Cohen's\ d=0.51$。其余 3 种选择下，EC 效应均不显著；对两类 CS 下，四种选择的差异进行检验，发现 CS 类型与四种选择的交互作用显著，$F(3,898)=5.73$，$p=0.001$，$\eta_p^2=0.02$。在 CS+ 与 CS- 条件下，正确选择项目的评价与相反效价项目的评价之间均差异显著，$M(CS+_{正确选择})=4.61$，$M(CS+_{相反效价})=4.29$，$t(374)=2.10$，$p=0.036$，$Cohen's\ d=0.24$；$M(CS-_{正确选择})=3.88$，$M(CS-_{相反效价})=4.50$，$t(352)=-3.38$，$p=0.001$，$Cohen's\ d=0.42$。如果被试基于需求意识做出评价，那么可以预测相反效价的关联记忆会导致被试反向评价配对过的 CS，虽然 $M(CS+_{相反效价})<M(CS-_{相反效价})$，但 $t(207)=1.09$，$p=0.277$，$Cohen's\ d=0.15$，不能说明被试作评价时是

表 1 基于呈现时长与效价强度分类的四图再认测验卡方分析

实验条件		关联意识				χ^2
		正确选择 %	相同效价 %	相反效价 %	中性 %	
US 呈现时长	短呈现	53.1	15.1	27.6	4.2	12.14**
	长呈现	63.1	13.3	18.8	4.8	
US 效价强度	弱效价	65.1	12.5	19.0	3.4	18.90***
	强效价	51.0	16.0	27.5	5.5	

注：*p <0.05; **p <0.01; ***p <0.001

基于需求意识。如果被试是基于推论意识来完成四图再认测验，那么可以预测，对那些由"消极条件反射"（非指 US 为消极刺激，而是指与 US+ 配对过的 CS 被被试评价为消极，与 US– 配对过的 CS 被被试评价为积极）产生的 CS，被试会报告与之匹配过的是相反效价的 US。因此，挑选出"消极条件反射"下的 CS，计数统计发现，正确选择 US 效价的次数为 127 次，相同效价为 34 次，相反效价为 67 次，中性选择 11 次，χ^2(3, N=239)=127.44, p<0.001，相反效价的选择只占总数 28%，不能说明被试基于推论意识来完成四图再认测验。进而基于呈现时长与效价强度对四图再认测验进行分类计数并进行卡方分析，发现呈现时长与效价强度的两个水平在与关联意识的四个选择上存在显著差异，结果见表 1。

3.5 呈现时长、效价强度和关联意识对 EC 效应的作用机制分析

为进一步检验本研究所涉及的因素对 EC 效应的影响，采用基于被试的统计分析，以呈现时长、效价强度、关联意识（正确选择的次数）为自变量，以两类 CS 评价的差值作为因变量，使用全部纳入法进行多元回归分析。结果发现，回归方程总体显著，F(3,112) =11.79, p<0 .001，确定系数 (R^2) 为 0.245。呈现时长、效价强度、关联意识程度可以显著预测回归方程 ($B_{呈现时长}$ =0.615, $Partial$ =0.265,t= 2.87，P<0.01；$B_{效价强度}$ =0.649, $Partial$ =0.273, t=2.97，P<0.01；$B_{关联意识}$ =0.199, Partial =0.382, t= 4.32，P<0.001)。

进而进一步探索纳入回归方程的三个变量影响 EC 效应的作用机制，以呈现时长和效价强度为自变量 (r=–0.009)，以关联意识为中介变量，以两类 CS 评价差值为因变量（所有变量都经过了中心

化），进行中介效应分析（温忠麟，张雷，候杰泰，刘红云，2004）。发现关联意识的中介效应不显著，$Z_{呈现时长 – 关联意识 –EC 评价}$ =0.354, p=0.362; $Beta_{效价强度}$ =0.163, p=0.072。

4 讨论

本研究中，研究者使用商标图案作为 CS，操纵 US 的呈现时长与效价强度，通过外显测量来考察这两个因素对 EC 效应的影响，并结合四图再认测验与基于项目分析，对关联意识的作用进行了进一步检验。实验在总体水平上发现了显著的 EC 效应，后续讨论均基于总体的 EC 效应展开。

4.1 呈现时长、效价强度对 EC 效应的影响

Jones 等人 (2009) 从归因角度提出了 EC 的解释模型——内隐错误归因机制。本研究对其实验结果及推论进行了验证和探索，结果不支持内隐错误归因机制。以往研究常常控制整个 CS–US 匹配过程中被试的意识和注意等影响因素，单独控制 US 呈现时长时也往往设置在阈下水平，缺乏对 US 阈上呈现时长的探讨。Hofmann De Houwer, Peragini, Baeyens 和 Crombez (2010) 进行的元分析发现，阈上呈现 US 相比于阈下呈现 US，会产生更大的 EC 效应。而本研究在 US120ms 呈现水平上并未发现 EC 效应，相比于阈上长呈现与以往研究的阈下呈现，EC 效应在阈上短呈现水平上出现了一个缺口，这表明被试的注意时长会影响到 EC 效应，EC 效应在时间维度上并不是一个连续存在的效应，阈上长呈现与阈下水平 EC 效应的发生机制可能不同。

在 US 效价强度这个问题上，目前包括本研究在

内的涉及效价强度的三个研究给出了三个不同的结果。Jones 等人 (2009) 让被试在刺激搜索任务中完成条件反射，在大样本的条件下发现了一个较弱的 EC 效应 (p=0.049)，弱效价的 US 相比于强效价 US 会产生更强的 EC 效应，且此结果只发生在无关联意识的被试上。他们的研究中 US 刺激既有文字又有图片，但效价强度的划分是基于被试对积极、消极图片进行效价分类的反应时而非对效价的直接评价得出的，US 刺激材料的筛选降低了其研究的效度。Bayens 等人 (1988) 将实验伪装成生理指标测量以此让被试完成条件反射，实验在探讨 EC 消退的主题下附带发现强、弱效价组有相近的 EC 效应。但其实验仅区分了消极 US 的效价强度，且使用了含有高恐怖唤起的刺激 (残缺的脸) 作为强效价 US，而使用这类包含血腥与恐怖成分的消极刺激，可能会带来混淆因素而影响实验结果。本研究使用相对简单与规范的范式，严格区分了 US 效价强度，发现强效价组，EC 效应显著，弱效价组 EC 效应不显著。结果显示，EC 效应的发生需要一定效价强度的 US，这一实验结果对效价强度如何影响 EC 效应提供了进一步的例证。

4.2 关联意识对 EC 效应的影响

关联意识是区分经典条件反射与 EC 的一个重要特征。早期研究认为，经典条件反射需要关联意识，而 EC 可以独立于关联意识发生。但本研究从基于被试和基于项目的分析中，都发现 EC 效应的产生需要基于关联意识。这与 Pleyers 等人 (2007, 2009) 的研究结果相契合，但与 Walther 和 Nagengast (2006) 的结果相反。这个结果不支持内隐错误归因机制，因为关联意识的存在不仅没有降低 "错误归因" 的可能性，反而是促使 EC 效应发生的重要因素。内隐错误归因机制还预测，在条件反射阶段中，如果 CS 与多个不同的 US 配对，随着不同 US 个数的增加，被试错误归因的可能性也会提高，错误累积会增大 EC 效应。但这也与 Stahl & Unkelbach(2009) 的研究结果不一致。多个相同效价的 US 会导致任务复杂程度的增加，被试的注意资源有限，无法形成 CS-US 关联意识。基于本研究结果，我们也推测，与同一个 CS 配对的

US 个数的增加会降低 EC 效应。

Stahl 等人 (2009) 认为，在前人研究中，即使被试没有形成 CS-US 关联意识，EC 效应也可以发生，其原因可能在于被试记住了更为有效且经济的 US 效价而非 CS-US 配对。Stahl 等人通过实验检验了其假设。但本研究发现，在选择了相同效价的项目上，EC 效应并不显著。但这可能也与同一个 CS 配对的 US 个数不同有关。本实验中，CS 只与唯一的 US 配对，而在 Stahl 等人的研究中，CS 与多个 US 配对。此时如果要发生 EC 效应且在关联意识必不可少的条件下，被试可能转而使用 US 效价作为 CS-US 关联意识的替代。此外，在本研究中，需求意识与推论策略未对 EC 效应有显著影响。

本研究也发现，基于项目分析与基于被试分析相比，关联意识虽然显著地作用于 EC 效应，但效应值有大幅下降，这与 Pleyers 等人 (2007) 认为基于项目分析有更强的检验力的观点相符，因此，对关联意识的作用进行推论时还需谨慎。同时，关联意识的中介效应不显著，没有发挥出命题性解释模型所赋予的更为核心的作用。以往研究缺乏对关联意识与其他变量间关系的探讨，对于关联意识在命题性模型中的核心作用也缺乏进一步的检验。本研究结果提示：可能因为存在其他中介机制而削弱了关联意识的中介作用：对于不同自变量，是共享一种中介机制还是存在多种中介机制，不同的中介机制背后可能潜藏着不同的解释模型？对于命题性模型而言，过度依赖于关联意识可能会降低其本身的解释力。

4.3 EC 效应与 APE 模型、EC 的解释模型

在态度的 APE 模型中，EC 被视为内隐态度改变的核心过程，并可以影响外显态度。APE 模型认为，EC 属于自下而上的低水平自动激活过程，不需要太多的注意资源或是目的导向 (Gawronski & Bodenhausen, 2006)。但本研究的结果与其他研究 (Corneille, Yzerbyt, Pleyers, & Mussweiler, 2009; Field & Moore, 2005; Pleyers, et al., 2009) 并不支持这种观点。Corneille 等人 (2009) 发现，EC 效应会受到加工目标

的影响，他们分别让两组被试先完成一个寻找两张图片知觉相似或知觉相异的任务，以使两组被试产生不同的加工目标，然后再进入条件反射阶段，结果发现知觉相似组被试的 EC 效应更大。本研究发现 EC 效应受到 US 呈现时长的影响。以上结果都表明，EC 并不完全是一个自动激活的联想评价过程 (Moors & De Houwer, 2006)。

本研究也发现 EC 效应受到关联意识的积极影响，但 APE 模型认为 EC 会受到关联意识的消极影响，被试会因为关联意识的存在无法信任联想评价过程中的效用，联想结构的改变无法体现在外显评价上，最终导致外显 EC 效应无法发生，本研究结果与此观点相悖。APE 模型同时认为，关联意识作用于命题加工的外显评价，不会作用于内隐测量的 EC 效应。但新近的研究不支持此观点。被试如果缺少关联意识，通过内隐测量也不能发现 EC 效应 (Dawson, et al., 2007; Stahl & Unkelbach, 2009; Stahl, et al., 2009)。不过，也有研究者指出，内隐测量的手段是否能完全避免命题加工的影响，命题加工是否会影响联想加工过程也需实证支持 (De Houwer, 2007)。目前对关联意识的研究仅仅通过事后再认测量对被试或反应项目进行分类，尚没能直接控制操纵关联意识，其中存在的问题还有待精细的实验研究 (Shanks, 2010)。同时，如果关联意识对 EC 效应的发生至关重要，那么无意识 EC 效应也有待深入研究。

较早提出的 EC 效应解释模型有概念分类解释 (the conceptual-categorization account)、整体性解释 (the holistic account)、参照性解释 (the referential account) (De Houwer et al., 2001, 2010)。不过，它们各自在刺激类型、US 再评价、感觉前条件化、无意识 EC 效应、关联意识、基于指导语发生的 EC 效应等特定问题中缺乏一定的解释力。最近提出的内隐错误归因机制也存在类似缺陷。Hofmann 等人 (2010) 的 EC 效应元分析结果也较少支持内隐错误归因机制。De Houwer 等人提出的命题性解释 (the propositional account) 认为，形成命题的过程是一个处在意识层面并需要注意资源的过程，因此命题性解释与其他几种不同，其预测 EC 效应依赖于对 CS-US 配对的意识，并且注意损耗会减弱 EC 效应，但此模型无法对一部分无意识 EC 效应给出解释。近年来，随着越来越多的研究发现 EC 效应基于关联意识发生，命题性解释模型也逐渐引起重视。不过，解释模型的提出并没有解决相关争论。其原因在于研究者们至今还没有明确相关因素对 EC 效应的影响方式。本研究结果部分支持命题性解释模型，也提示命题性解释模型仍需补充、修订。

De Houwer 等人 (2005, 2007) 为了统合争论，也指出另一种可能性，EC 效应可能在不同的条件或范式下，基于不同的过程发生：在某些条件下基于自动激活的过程，某些条件下基于命题评价过程。例如，当被试的注意资源被分散时，EC 效应的发生更可能基于自动激活过程。Jones olson 和 Fazio (2010) 也认为可能存在多种机制共同影响 EC 效应，并认为不同的实验范式可能存在不同的作用机制。但 Bar-Anan 等人 (2010) 采用 Jones 等人 (2009) 的实验范式，结果也支持命题性解释模型。而从本研究结果来看，假使被试在 US120ms 呈现水平上基于自动激活过程，那么 EC 效应也应该能够发生。虽然本研究结果提示阈上长呈现与阈下呈现的 EC 效应发生机制可能不同，但这种 EC 效应的"自动激活—命题评价"双过程观点也有待进一步论证。即使成立，两种过程间也可能存在复杂的相互影响，并且目前也尚未明确某一种过程的专属测量方法。本研究也存在一些局限性，例如基于本研究设计无法详细探讨 CS(+) 与 CS(-) 各自部分的 EC 效应及自变量的影响。因此，未来研究可以进一步改进研究设计 (如增加对 CS 的前后测量)，使用多种方法 (内隐测量、生理测量) 进一步检验本研究的结论；在本研究呈现时长变量的基础上进一步从时间维度上对 EC 效应进行探讨，探讨阈下呈现与阈上长呈现 EC 效应的发生机制，从而进一步检验双过程观点；探索影响 EC 效应的变量之间的关系，尤其是检验可能的关联意识与其他变量共同作用的中介或调节机制，或找出影响 EC 效应的其他中介变量；进一步设计操控关联意识的实验方法。

5 结论

①无条件刺激的呈现时长对评价性条件反射效应影响显著。在长呈现水平上，效应显著发生；在短呈现水平上，效应不发生。

②无条件刺激的效价强度对评价性条件反射效应影响显著。在强效价水平上，效应显著发生；在弱效价水平上，效应不发生。

③关联意识对评价性条件反射效应影响显著。本研究只在那些较好地记住了条件刺激与无条件刺激匹配关系的被试上发现了效应。

④关联意识在呈现时长（效价强度）与 EC 效应间的中介作用不显著。

参考文献

Baeyens, F., Crombez, G., Van den Bergh, O., & Eelen, P. (1988). Once in contact, always in contact: EC is resistant to extinction. *Advances in Behaviour Research and Therapy, 10*, 179–199.

Baeyens, F., Eelen, P., & Bergh, O. (1990). Contingency awareness in evaluative conditioning: A case for unaware affective–evaluative learning. *Cognition & Emotion, 4*(1), 3–18.

Bar–Anan, Y., De Houwer, J., & Nosek, B. A.(2010). Evaluative conditioning and conscious knowledge of contingencies: A correlational investigation with large samples. *The Quarterly Journal of Experimental Psychology, 63*(12), 2313–2335.

Chen, J., Liang, N. J., Yang, F. Y. (2009). A preliminary experimental study on the changeabiligy of implicit self–esteem. *Psychological Science, 32*(4):812–815.

[陈进，梁宁建，杨福义. (2009). 自尊内隐改变的初步实验研究. *心理科学, 32*(4), 812–815.]

Chua, H. F., Boland, J. E., & Nisbett, R. E. (2005). Cultural variation in eye movements during scene perception. *Proceedings of the National Academy of Sciences of the United States of America, 102*(35), 12629–12633.

Corneille, O., Yzerbyt, V. Y., Pleyers, G., & Mussweiler, T. (2009). Beyond awareness and resources: Evaluative conditioning may be sensitive to processing goals. *Journal of Experimental Social Psychology, 45*(1), 279–282.

Dawson, M., Rissling, A., Schell, A., & Wilcox, R. (2007). Under what conditions can human affective conditioning occur without contingency awareness? Test of the evaluative conditioning paradigm. *Emotion, 7*(4), 755–766.

De Houwer, J. (2007). A conceptual and theoretical analysis of evaluative conditioning. *Spanish Journal of Psychology, 10*, 230–241.

De Houwer, J. (2010). Evaluative conditioning: A review of procedure knowledge and mental process theories. In T. R. Schachtman & S. Reilly. Applications of learning and conditioning. Oxford, UK: Oxford University Press.

De Houwer, J., Baeyens, F., & Field, A. (2005). Associative learning of likes and dislikes: Some current controversies and possible ways forward. *Cognition & Emotion, 19*(2), 161–174.

De Houwer, J., Thomas, S., & Baeyens, F. (2001). Associative learning of likes and dislikes: A review of 25 years of research on human evaluative conditioning. *Psychological Bulletin, 127*(6), 853–869.

Dijksterhuis, A. (2004). I like myself but I don't know why: Enhancing implicit self–esteem by subliminal evaluative conditioning. *Journal of Personality and Social Psychology, 86*, 345–355.

Field, A., & Davey, G. (1999). Reevaluating evaluative conditioning: A nonassociative explanation of conditioning effects in the visual evaluative conditioning paradigm. *Journal of Experimental Psychology Animal Behavior Processes, 25*, 211–224.

Field, A. P., & Moore, A. C. (2005). Dissociating the effects of attention and contingency awareness on evaluative conditioning effects in the visual paradigm. *Cognition & Emotion, 19*, 217–243.

Fulcher, E., & Hammerl, M. (2001). When all is revealed: A dissociation between evaluative learning and contingency awareness. *Consciousness and Cognition, 10*(4), 524–549.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*, 692–731.

Gibson, B. (2008). Can evaluative conditioning change attitudes toward mature brands? New evidence from the implicit association test. *Journal of Consumer Research, 35*(1), 178–188.

Hammerl, M., & Fulcher, E. (2005). Reactance in affective–evaluative learning: Outside of conscious control? *Cognition & Emotion, 19*(2), 197–216.

Hofmann, W., De Houwer, J., Perugini, M., Baeyens, F., & Crombez, G. (2010). Evaluative conditioning in humans: A meta–analysis. *Psychological Bulletin. 136*(3), 390–421.

Jones, C. R., Fazio, R. H., & Olson, M. A. (2009). Implicit misattribution as a mechanism underlying evaluative conditioning. *Journal of Personality and Social Psychology, 96*(5), 933–948.

Jones, C. R., Olson, M. A., & Fazio, R. H. (2010). Evaluative conditioning: The "How" question. In M. P. Zanna & J. M. Olson (Eds.), *Advances in Experimental Social Psychology* (Vol. 43, pp. 205 - 255). San Diego: Academic Press.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2008). *International affective picture system (IAPS): Affective ratings of pictures and instruction manual.* Technical Report A–8.University of Florida, Gainesville, FL.

Lipp, O., & Purkis, H. (2005). No support for dual process accounts of human affective learning in simple Pavlovian conditioning. *Cognition & Emotion, 19*(2), 269–282.

Lovibond, P. F., & Shanks, D. R. (2002). The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. Journal of Experimental Psychology: *Animal Behavior Processes, 28*(1), 3–26.

Mitchell, C., De Houwer, J., & Lovibond, P. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences, 32*(02), 183–198.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin, 132*, 297–326.

Olson, M. A., & Fazio, R. H. (2001). Implicit attitude formation through classical conditioning. *Psychological Science, 12*, 413–417.

Olson, M. A., & Fazio, R. H. (2002). Implicit acquisition and manifestation of classically conditioned attitudes. *Social Cognition, 20*(2), 89–104.

Olson, M. A., & Fazio, R. H. (2006). Reducing automatically activated racial prejudice through implicit evaluative conditioning. *Personality and Social Psychology Bulletin, 32*(4), 421–433.

Olson, M. A., Kendrick, R. V., & Fazio, R. H. (2009). Implicit learning of evaluative vs. non–evaluative covariations: the role of dimension accessibility. Journal of Experimental Social Psychology, 45(2), 398–403.

Pleyers, G., Corneille, O., Luminet, O., & Yzerbyt, V. (2007). Aware and (dis) liking: Item–based analyses reveal that valence acquisition via evaluative conditioning emerges only when there is contingency awareness. *Journal of Experimental Psychology: Learning, Memory, and Cogniton, 33*(1), 130–144.

Pleyers, G., Corneille, O., Yzerbyt, V., & Luminet, O. (2009). Evaluative conditioning may incur attentional costs. *Journal of Experimental Psychology: Animal*

Behavior Processes, 35(2), 279–285.

Purkis, H. M., & Lipp, O. V. (2001). Does affective learning exist in the absence of contingency awareness? *Learning and Motivation, 32*(1), 84–99.

Shanks, D. R. (2010). Learning: From association to cognition. A*nnual Review of Psychology, 61*, 273–301.

Stahl, C., & Unkelbach, C. (2009). Evaluative learning with single versus multiple unconditioned stimuli: The role of contingency awareness. *Journal of Experimental Psychology: Animal Behavior Processes, 35*(2), 286–291.

Stahl, C., Unkelbach, C., & Corneille, O. (2009). On the respective contributions of awareness of US valence and US identity in attitude formation through evaluative conditioning. *Journal of Personality and Social Psychology, 97*(3), 404–420. .

Walther, E. (2002). Guilty by mere association: Evaluative conditioning and the spreading attitude effect. *Journal of Personality and Social Psychology, 82*(6), 919–934.

Walther, E., Gawronski, B., Blank, H., & Langer, T. (2009).

Changing likes and dislikes through the back door: The US–revaluation effect. *Cognition & Emotion, 23*(5), 889–917.

Walther, E., & Nagengast, B. (2006). Evaluative conditioning and the awareness issue: Assessing contingency awareness with the four–picture recognition test. *Journal of Experimental Psychology Animal Behavior Processes, 32*(4), 454–459.

Walther, E., Nagengast, B., & Trasselli, C. (2005). Evaluative conditioning in social psychology: Facts and speculations. *Cognitioin and Emotion, 19*, 175–196.

Wardle, S., Mitchell, C., & Lovibond, P. (2007). Flavor evaluative conditioning and contingency awareness. *Learning & Behavior, 35*(4), 233–241.

Wen, Z. L., Zhang, L., Hou, J.T., & Liu, H.Y. (2004). Testing and application of the mediating effects(in Chinese). *Acta Psychologica Sinica, 36*, 614– 620.

[温忠麟 , 张雷 , 侯杰泰 , 刘红云 .(2004). 中介效应检验程序及其应用 . *心理学报 , 36*, 614– 620.]

Evaluative conditioning: The role of display duration, valence Intensity and contingency awareness

ZHAO Xian [1,2] ; LI Ye [2] ; LIU Li [1] ; ZENG Hongling [2] ; ZHENG Jian [1]

([1] School of Psychology , Beijing Normal University ; Beijing Key Lab of Applied Experimental Psychology , Beijing 100875, China)

([2] School of Psychology , Central China Normal University ; Hubei Human Development and Mental Health Key Laboratory

Wuhan 430079, China)

Abstract   Evaluative conditioning (EC) refers to the change of an attitude toward an affectively neutral object (the conditioned stimulus, or CS), following that object's pairing with another positively or negatively valenced stimulus (the unconditioned stimulus, or US). EC is theoretically regarded as an associative learning process in associative-propositional evaluation (APE) model, but many controversies have arisen in empirical studies of EC. Some researchers found EC could not occurred without awareness of CS-US contingencies (supports propositional account), but others didn't. Same debate also exists in whether EC relies on much attention or not. Based on propositional account, the purpose of the present study was to investigate the effect of display duration of US, retest the effect of valence intensity, CS-US contingency awareness, and their integrated mechanism on EC, measured by explicit evaluative rating, combined with four-picture recognition test and item-based analyses. The hypotheses were tested in a sample of 122 college students (38 males). In a 2(display duration) × 2(valence intensity) × 2(CS type) mixed design, with CS type as a within-group factor (CS+ means that CS was paired with positive pictures and CS- means that CS was paired with negative pictures), participants were randomly assigned to four groups: US short display (120ms)-strong valence, US long display (1000ms)-strong valence, US short display-mild valence, US long display-mild valence. The CSs were real pictorial trademarks but unfamiliar to participants and USs were chosen partially from the International Affective Picture System and partially from the Internet. All pictures used in the experiment were selected based on pilot rating of valence. The experiment consisted of four sequential phases: the conditioning phase, evaluative rating phase, four-picture recognition phase and questionnaire phase. The results were analyzed by MANOVA on a global method of statistical analysis (participant-based), and on participants' responses to individual items (item-based) which was particularly for contingency awareness. Multiple regression analysis and chi-square test were also conducted. Results showed that evidence was obtained for EC effect on the global level, with more positive ratings of CS+ than of the CS-. For display duration, EC emerged only on long display level, M(CS+)=4.68，M (CS-) =3.86，F(1,111)=25.19, p<0.001,0.21. For valence intensity, EC was found only on strong valence level, M(CS+)=4.57，M(CS-)=3.93，F(1,111) = 14.57, p<0.001, 0.16. And only participants who were categorized as "contingency aware" showed significant EC effect. In addition, no effect of demand awareness and inference strategy was obtained. However, contingency awareness did not significantly mediate the relationship between display duration and EC effects, nor does

the relationship between valence intensity and EC effects. This pattern of results partly supported the hypotheses, and emphasized the role of attention and awareness on EC effect. EC relies on US display duration, US valence intensity and CS–US contingency awareness, which in line with propositional account. Results also raise doubts about the views in implicit misattribution mechanism and APE model. EC is unlikely to be an associative evaluation process which requires little cognitive capacity and contingency awareness. Nonetheless, the insignificant mediation effect also indicates that propositional account needs further modification.

Different effects of alcohol on automatic detection of colour, location and time change: A mismatch negativity study

Jinbo He [1], Yue Hu [1], Satu Pakarinen [2], Bingbing Li [1], Zongkui Zhou [1]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior of Ministry of Education, Central China Normal University, Wuhan, China)

([2] Finnish Institute of Occupational Health, Helsinki, Finland)

Abstract Effects of alcohol on pre−attentive processing have been well studied in the past decades. However, the majority of studies focus on pre−attentive auditory processing and only a few studies have examined the effects of alcohol on visual processing. This study investigates the effects of alcohol on automatic pre−attentive processing of visual changes in colour, location and duration. We compared the mean amplitudes and the amplitude decline ratios of visual mismatch negativity (vMMN) responses elicited for small and large changes in image colour, location, and duration in 12 individuals under the influence of alcohol and in a placebo condition. The vMMN amplitudes for changes in location and duration significantly decreased in the alcohol condition as compared with the placebo condition, and the magnitude of decrease was not related to the magnitude of change in the deviant stimuli. However, the amplitude of colour vMMN, in the alcohol condition, did not change significantly compared to the placebo condition. These results show that pre−attentive visual processing is impaired by alcohol, and that this impairment may be feature−specific. In addition, this impairment was not related to the magnitude of stimuli change.

Keywords alcohol; event−related potential; visual mismatch negativity; multi−feature paradigm; pre−attentive change detection

1 Introduction

Automatic detection of changes in the environment, even when unattended, helps us to adapt to the continually changing environment (e.g. orienting reflex; Sokolov, 1960, 1963) and to conduct complicated tasks more efficiently (e.g. driving). Alcohol is well known to impair both the accuracy as well as speed of processing, affecting both perception and response execution. Whereas the impairing effects of alcohol on auditory sensory processing are well known (e.g. He et al., 2013; Jäskeläinen et al., 1995a; Kähkänen et al., 2005) the influence of alcohol on the visual system is considerably less studied.

The pre−attentive modelling of the auditory environment and the detection and prediction of potentially important changes in the auditory stream have long been studied with the mismatch negativity (MMN; for a review see Nätänen et al., 2007) of the auditory event−related potential (ERP). The

MMN is a change−specific ERP component elicited by any discriminable change in stimulation, with its amplitude and latency correlating both with the magnitude of change as well as with the behavioural measures of sound change detection, such as accuracy and speed (e.g. Amenedo and Escera, 2000; Pakarinen et al., 2007). The visual analogue of the MMN is termed as visual mistnatch negativitg (vMMN; e.g. Czigler et al., 2002; for review see Czigler, 2007; Heslenfeld, 2003). Studies on the effects of alcohol on pre−attentive processing have mainly focused on auditory modality, and in those studies it has been found that alcohol impairs the automatic detection of changes in different sound features (He et al. 2013; Jäskeläinen et al., 1995a,1995b, 1996; Kähkänen et al., 2005; Kenemans et al., 2010).

Only recently, Kenemans et al. (2010), using reversed oddball procedure, showed that even moderate doses (0.54 mL/kg) of alcohol disrupt the automatic detection of spatial frequency change in a visual grating, as indicated by reduced rareness−related negativity (RRN，Kenemans et al., 2003) response. Importantly, this occurred despite the elicitation of an spatial−frequency dependent difference at 80 ms (Kenemans et al., 2000) response, which indicates intact modelling of the visual features per se. This pattern of evidence, together with the earlier findings on auditory ERPs (e.g. Jäskeläinen et al., 1995a; Kähkönen et al., 2005 He et al., 2013); suggest that alcohol decreases the sensitivity of the sensory system to detect unattended but potentially significant changes outside the focus of attention. However as Kenemans et al (2010) used only one type of changes in visual gratings, it is unclear whether this disruption is more general in nature, or specific to certain aspects of vision such as colour, location and duration. It is important for people to automatically detect changes of these visual features. For example, drivers need to automatically detect changes traffic lights. It is theoretically important for prevention in traffic accidents caused by drunk driving to investigate whether the automatic detection of colour, location, or lighting duration changes is impaired as that of spatial frequency changes (Kenemans et al., 2010) and whether alcohol impairs the automatic detection of different features equally.

Although vMMN and RRN are both posterior negativity elicited by deviant stimuli compared to standard stimuli, they actually reflect qualitatively different processing. In a study which used both traditional oddball sequence and equiprobable oddball sequence, it was found that two negativities between 100−150 ms and between 200−250ms were elicited by deviant stimuli compared to standard stimuli and only a negativity between 200−250 ms was elicited by deviant stimuli compared to control stimuli. Therefore, the early negativity between 100−150ms (e.g. RRN) reflects refractory effect, and the late posterior negativity between 150−200ms (vMMN) reflects memory−comparison based change detection (Kimura et al., 2009). The present study used a multi−feature vMMN paradigm in which memory−comparison−based vMMN but not refractory based RRN was elicited by deviant stimuli during 150−300 ms post−stimulus onset (e.g. Qian et al., 2014 Shi et al., 2013;). Thus we could investigate whether alcohol impairs memory comparison−based change in detection of colour, location and duration.

In the auditory modality it has been shown, that the processing of various auditory attributes is differentially affected by alcohol (He et al., 2013). Our aim here is to examine whether processing of different visual features is differentially affected by alcohol by recording the visual MMN to changes in item colour, location and duration and, further, to assess the effect of deviation magnitude under the influence of alcohol as compared to a placebo condition.

2 Method and material

2.1 Participants

Twelve participants (right−handed; 21−25 years old;

mean age 22.75; two females) with normal (or corrected) vision participated in this experiment. All of the participants were healthy normal social drinkers (3–12 standard alcohol portions per month during the previous year) with no chronic alcoholism or family history of alcoholism and other mental illnesses. They were asked to abstain from caffeine, alcohol, nicotine, and other psychoactive substances for 24h prior to the experiment. A written informed consent from was signed and participants committed to obey the aforementioned request. All of the participants were paid for their involvement in this study. This study was approved by the institutional ethical committee of the National Key Laboratory of Cognitive Neuroscience and Learning of China.

2.2 Treatment

All of the participants attended two experimental sessions (alcohol and placebo conditions) separated by two weeks. The sequence of alcohol and placebo sessions was counterbalanced between the participants so that a half of them participated in the alcohol condition first and the other half paricipatedin the placebo condition first. For the alcohol condition, the participants received a dose of 0.65g/kg of alcohol (53% v/v white wine) and were provided with little food (e.g. peanuts). They were given 10min to finish the drink. Once the participants finished drinking, the blood alcohol concentration (BAC) level was tested every 5min until a constant level was reached. The mean time interval between the finish of drinking and the beginning of the experiment was 18.5min. The BAC level was also tested immediately after the experiment. The average BAC level was $0.06 \pm 0.005\%$ before the start of the task and $0.048 \pm 0.006\%$ after the experiment. For the placebo condition, the participants received a dose of 0.02g/kg alcohol (53% v/v white wine mixed with distilled water) and were provided with a little food. The same BAC level test procedure as in the alcohol condition was performed after they finished drinking. The time interval

between the finish of drinking and the beginning of the experiment was 15min. The BAC level of the participants in the placebo condition was 0 before the start of the task and after the experiment. The mean BAC level in the alcohol condition was significantly higher than that in the placebo condition [$t_{(11)}$=58.17; $p<0.001$]. In addition to BAC level measurement, intoxication and confidence in the participants' performance in the experiment were reported on a five–point Likert–type scale before and after the experiment in both sessions. The subjective report of intoxication was 3.08 (± 1.38) before the start of the task and 2.81 (± 1.04) after the experiment in the alcohol condition. By contrast, the subjective report of intoxication was 1.00 (± 0.29) before the start of the task and 1.11 (± 0.23) after the experiment in the placebo condition. The subjective report of intoxication in the alcohol condition was significantly higher than that in the placebo condition [$t_{(11)} = 3.73$; $p<0.01$]. The subjective report of confidence on performance was 2.25 (± 1.03) before the start of the task and 1.83 (± 0.87) after the experiment in the alcohol condition. For the placebo condition the subjective report of confidence on performance was 2.58 (± 0.90) before the start of the task and 2.92 (± 0.79) after the experiment. The subjective report of confidence on the performance in the alcohol condition was significantly lower than that in the placebo condition [$t_{(11)}$=2.49, $p< 0.05$].

2.3 Stimulus design and task

The stimuli were shown against a silver background in the center of the screen 65cm from the participant. A black cross ($0.8^0 \times 0.7^0$) was constantly shown in the center of the screen during the experiment. Circular colour stimuli (2.6^0diameter) were presented on the left and right sides of the cross (2.8^0 away). The size of the cross was occasionally increased ($1.0^0 \times 0.9^0$) or decreased ($0.4^0 \times 0.3^0$) during stimulus presentation (randomly occurred 26 times in each block). The participants were asked to fixate to the cross during the experiment. Keys in
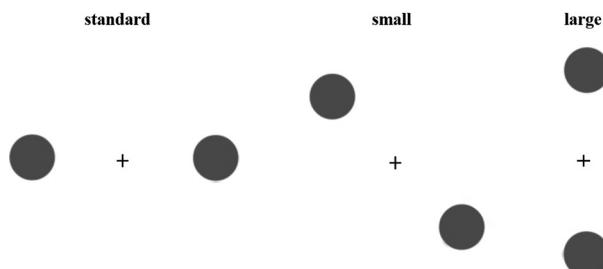
Figure. 1. The illustration for locations of standard stimuli and location deuiant stimuli. Note: Half of the small location deviant stimuli were also located at the opposite orientation to the standard stimulus as the small location deviant stimulus in this figure.

the keyboard were assigned for the large or the small cross (the corresponding keystones to the large and the small cross were counterbalanced among the participants in both sessions). The task of the participants was to press the corresponding key when they detect changes in the size of the cross. The participants were also informed that the stimuli beside the cross were not related to the experiment and should be ignored.

The standard stimuli were sea−green cycles (25.63cd/m^2, L, a, b(CIE Lab 1976 colour speces): 51, −38, 19) that were presented for 20ms. The deviant stimuli differed from the standard stimuli only in terms of colour, location, and duration. The colour deviant stimuli were green for the small deviants stimuli (23.61cd/m^2, L, a, b: 54, −53, 55) and red for the large deviants stimuli (22.92cd/m^2, L, a, b: 54, 81, 70). The location deviant stimuli were located 45° (small deviant stimuli) and 90°(large deviant stimuli) with respect to the cross, at the same distance as the standard stimuli (see Figure 1 for a description of the locations of standard and deviant stimuli). The small duration deviant stimuli were shown for 60ms and the large duration deviants stimuli for 100 ms. Thus, there were three types of stimuli in total, each with two deviant levels.

Participants were seated in a soundproof electromagnetically shielded room. The stimuli were presented with the E−prime 2.0. The stimulus sequence was a visual analogue of the recently developed fast multi−feature auditory MMN paradigm (Nätänen et al., 2004; Pakarinen et al., 2007).

In the stimulus sequence, two standards were presented between each two successive deviant stimuli. Each deviant stimulus type (colour, location, and duration) was presented once in an array of three successive deviant stimuli, and two successive deviant stimuli were always of different types. The order in which the different levels of deviation appeared in the sequence was pseudo−random. Furthermore, each type of deviant stimuli was presented equiprobably in each stimulus sequence. There were four 5.5−min sequences (820 stimuli per sequence (550 standard stimuli and 270 deviants), all sequences beginning with 10 successive standard stimuli), with the presentation order of the sequences varying randomly between the subjects. The inter−stimulus intervals (ISIS) were randomized between 300ms and 400ms.

2.4 ERP recording

The electroencephalogram (EEG) was recorded (0.05−40Hz at the sampling rate of 500Hz) with the NeuroScan system. Subjects wore an electrode cap, in which 64 Ag/AgCl electrodes were inserted. The electrodes were referred to the tip of the nose. Two electrodes were placed below and above the center of left eye to record vertical electro−oculograms (Eoas) and two were placed on the canthi of the eyes to record horizontal (Eoas). EOG artifacts were corrected using a linear regression estimate method (Semlitsch et al., 1986). Subjects were told to be relaxed and focus on the task and to minimize eye movement to reduce ocular−artifact. The EEG was filtered offline with a pass band of 0.05−30Hz. Epochs of 450ms (including a 50 ms pre−stimulus period served as a baseline for the amplitude measurement) were separately averaged for stimuli of different types and levels. Epochs elicited by the first 10 stimuli of each sequence and those with the voltage exceeding ±75μV were omitted from the analysis.

2.5 Data analysis

Hit rate and reaction time were calculated as a percentage of correct responses within 1s from the

stimulus. Responses occurring after 1s from the change were classified as omissions. Responses for no change within two stimulus intervals were classified as false alarms and were excluded from the analysis. Paired t–test was conducted separately on reaction times and hit rates under placebo and alcohol conditions.

vMMNs were derived by subtracting the response to the standard stimuli from the response to the deviant stimuli of each type and level. The grand averaged mean amplitude of the left parietal–occipital electrode cluster (P1, P3, P5, PO3, PO5) and the right parietal–occipital electrode cluster (P2, P4, P6, PO4, PO6) was used for analysis of variance (ANOVA). Based on previous studies on vMMN (e.g. Grimm Czigler and Sulykos, 2010 et al., 2009;) and observation of the ground averaged waveforms, mean amplitudes between 150–210ms, 150–230 ms, and 150–280 ms were quantified as the average amplitudes for colour vMMN, location vMMN, and duration vMMN, respectively. Three–way repeated–measures ANOVA (condition: placebo vs alcohol; deviant stimuluts magnitude: small vs large; hemisphere: left vs. right) were conducted separately for different vMMN types. Greenhouse–Geisser corrections were used in ANOVA when appropriate (However, uncorrected freedom with corrected p value was reported in the results part).

3 Results

Table 1 shows the hit rates and reaction times for correct responses under placebo and alcohol conditions. The paired t–test on reaction times and hit rates in

Table 1. Hit rates (in percentage) and reaction times (in ms) under placebo and alcohol conditions. Standard deviants are shown in the brackets.

	Placebo	Alcohol
Hit rate	97 (± 2)	96 (± 3)
Reaction time	531.24 (± 38.77)	553.36 (± 28.41)*

* $p < 0.05$
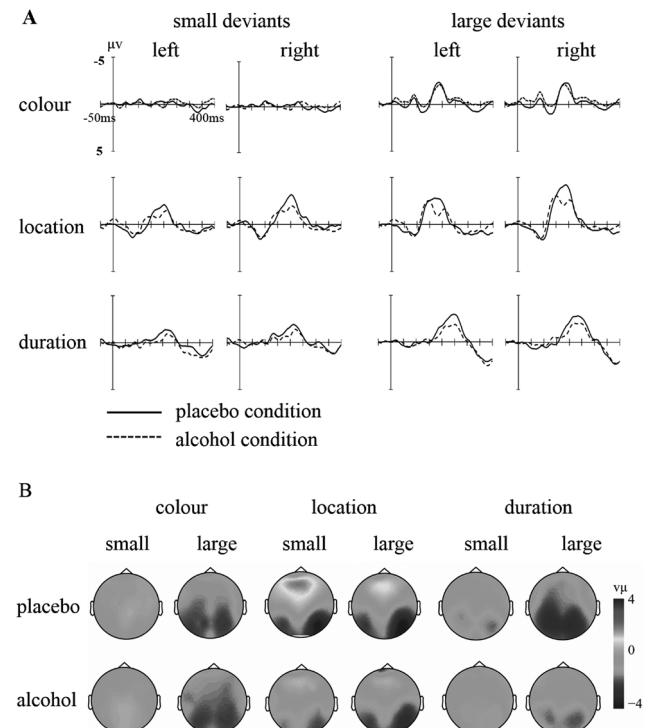
Figure. 2. (a) Grand–average visual mismatch negativity (vMMN) for each deviant stimulus type and deviant stimulus magnitude (column left are waveforms of the left parietal–occipital electrode cluster and column right are waveforms of the right parietal–occipital electrode cluster) in placebo and alcohol conditions. (b) Corresponding topographic maps of vMMNs of each type and deviance magnitude under placebo and alcohol conditions. The latencies for the topographic maps were as follows: 172–194 ms for colour vMMN, 196–218 ms for location vMMN, and 220–242 ms for duration vMMN. Column small represents the maps for small deviant stimuli and column large for large deviant stimuli.

placebo and alcohol conditions indicated that there was no significant effect of alcohol on hit rates, whereas there was a significant effect of alcohol on reaction times [$t_{(11)}=2.37$, $p<0.05$]. The reaction time for the alcohol condition was significantly longer than that for the placebo condition.

In the present study, although we did not include equiprobable sequences, the difference waveform between deviant–ERP and standard–ERP was clear at the time window of 150–250 ms post stimulus onset and hence the present deviant–related negativity is (real) vMMN but not RRN. The waveforms of vMMNs of each type and deviant level are shown in Figure2 a and the corresponding topographic maps are shown in Figure2 b. All deviant stimuli elicited significant vMMNs under both placebo and

alcohol conditions except for the the small colour deviant stimulus (see table 2). A three-way repeated-measures ANOVA (condition: placebo vs alcohol; deviant stimulus magnitude: small vs large; hemisphere: left vs right) was conducted separately for each type of vMMNs.

Neither three- nor two-way interactive effects were observed for vMMNs of the three types. The main effect of the condition was only significant for the location and duration vMMNs (location: $F_{(1,11)}=6.95$, $p<0.05$; duration: $F_{(1,11)}=5.28$, $p<0.05$). The amplitudes of location and duration vMMNs decreased under the alcohol condition compared with the placebo condition. By contrast, the amplitude of the colour vMMN was not affected by alcohol. The main effect of deviant stimulus magnitude was significant for vMMNs of the three types (colour: $F_{(1,11)}=19.69$, $p< 0.01$; location: $F_{(1,11)} =7.66$, $p<0.05$; duration: $F_{(1,11)} =11.13$, $p<0.01$), indicating that the amplitudes for the large deviant stimuli were higher than those for the small deviant stimuli for vMMNs of the three types. The main effect of hemisphere was only significant for the location vMMN($F_{(1,11)} =11.34$, $p<0.01$). The amplitudes of the right hemisphere of the location vMMN were higher than those of the left hemisphere.

To compare the different effects of alcohol on the three types of vMMNs, we calculated the decline ratios of vMMNs (calculated as

$$\frac{(amplitude_{placebo}-amplitude_{amplitudealcohol\ condition})}{amplitude_{placebo\ condition}}$$

for each type and each deviant level of vMMNs (Figure 3). Two-way repeated-measures ANOVA (type, colour,

location, and duration) × deviant magnitude (small and large) was conducted for the amplitude decrease ratios of vMMNs of the three types, indicating that only the main effect of type was significant. Comparison based on Bonferroni correction indicated that the amplitude decline ratio of the colour vMMN was significantly lower than that of the location and duration vMMNs ($p< 0.001$). No differences were observed between the amplitude decline ratios of the location and duration vMMNs ($p> 0.1$).

4 Discussion

In this study, the effect of alcohol on pre-attentive automatic detection of changes in image colour, location, and duration was studied with a new visual multi-feature MMN paradigm. Significant vMMNs were recorded for all three deviant stimulus types and both sizes of deviations (small and large), except for the small colour changes. Moreover, the amplitudes of location and duration vMMNs significantly decreased in the alcohol condition compared with the placebo condition, but the amplitude of the colour vMMN did not change. Further, this observed alcohol-induced impairment on pre-attentive visual discrimination was of comparable magnitude, i.e. approximately a reduction of one-third of a vMMN amplitude, for both the small and the large visual changes.

The behavioural detection hit rates of the distraction task were not affected by alcohol, but the reaction times were significantly delayed as compared with placebo condition. This is in line with previous results showing that for alcohol doses less than 0.5 g/kg only reaction time

Table 2. Mean amplitudes of visual mismatch negativity (vMMN) of each deviant stimulus (superscript sumbol indicates whether the amplitudes of vMMN were significantly different from zero).

vMMN type	Placebo		Alcohol	
	Small	Large	Small	Large
Colour	0.06(0.99)[ns]	−2.03(1.64) *	0.23(0.86) ns	−2.23(1.63) **
Location	−2.54(1.64)***	−3.54(2.18)***	−1.50(1.43)*	−2.37(2.04)**
Duration	−1.22(0.62)***	−2.43(0,85)***	−0.64(1.13)*	−1.50(1.83) *

ns non significant , * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure3. Alcohol—induced decline ratios of visual mismatch negativity (vMMN) amplitudes for each type and each level. Error bar represents one standard error.

was affected (Kenemans et al., 2010). The delay in the reaction time is explained, at least in part, by the alcohol (already at BAC levels of 0.05%; Tzambazis and Stough, 2000) slowing down the motor system which is producing the behavioural response i.e. pressing the button. It should also be noted that in this study the behavioural task was to discriminate the size of the cross in the center of the screen, and not the same stimuli as for which the vMMNs were recoded. On the other hand, the close to perfect discrimination accuracy in this behavioural task also suggest that the task difficulty was very low resulting in a low sensitivity to detect effects of alcohol especially in such moderate doses. Further studies are needed to directly explore the relationship between the vMMN and the behavioural discrimination of the same visual features under the influence of alcohol.

The small colour deviant stimuli probably failed to elicit the vMMN because of a relatively small contrast between the standard and the deviant stimuli. The vMMN has been found smaller for within category contrasts (here sea—green standard vs green deviant stimuli) as compared to between category (here sea—green standard vs red deviant stimuli) contrasts (Clifford et al., 2010 Czigler et al., 2002). It thus appears that a relatively high level of change is required to elicit the colour vMMN (Pazo—Alvarez et al., 2003; Czigler et al., 2002).

The amplitudes of the location and duration vMMNs

decreased under the alcohol condition compared with the placebo condition. These results supported the view that alcohol impaired visualpre—attentive change detection functions and extended this to location and duration features. Kenemans et al. (2010) posited two ideas to explain the mechanism by which alcohol affects the automatic detection of changes outside attention focus: (a) The attention of subjects became narrowed after alcohol consumption; (b) The sensitivity of subjects to changes outside the attention focus decreased after alcohol consumption. Our results support the second explanation, given that the reaction times of the subjects were delayed under the alcohol condition, which is against the view that alcohol increased the attention focused on a relevant task.

By contrast, alcohol did not affect the vMMN amplitude for the colour change. This probably reflects differences in the processing of colour from other visual features, starting from the retina by the distinct visual pathways and all the way to the visual cortex. Colour representation is separated already from the retina by a separate neural, parvocellular pathway projecting to the temporal cortical regions related to the "what" system of the visual cortex (Maunsell, 1995). The effects on location and timing (duration) of a visual stimulus may be more similar with each other since they both are mainly processed in the parvocellular inter—blob pathway and project to the parietal regions related to the 'where' system of the cortex (Haxby et al., 1991; Lewis and Miall, 2003). However, it is worth noting that the non—significant effect of alcohol on colour vMMN may also be related to a low alcohol dose. Whether alcohol impairs colour vMMN under a higher alcohol dose needs to be further investigated. Another possibility is that effect of alcohol on automatic detection of colour change might be confounded by the effect of alcohol on colour processing.

What is notable is that the null effect of alcohol on

colour vMMN might be a result of contamination of differences between the physical stimulus properties of standard stimuli and colour deviant stimuli on the colour vMMN. since we did not use a reversed oddball paradigm, in which the three colours serve as standard and deviants stimuli alternatively in different blocks, in our experiment,the estimated colour deviance effect (vMMN) reflects both the colour effect and the true deviance effect. Alcohol could enhance the colour effect but reduce the true deviance effect, resulting in a net zero effect. Further studies are needed to investigate whether moderate alcohol affects true memory comparison–based detection of colour change using a reversed oddball paradigm to exclude the contamination of physical stimulus properties.

Interestingly, the extent by which alcohol impaired the vMMN did not change significantly as the magnitude of deviant stimuli increased in this study. Alcohol thus seems to affect visual change detection in a linear fashion. However, studies on the effects of alcohol on the auditory MMN have shown that the impairing effect of alcohol on the auditory MMN decreases as the magnitude of deviance increases (He et al., 2013; Jäskeläinen et al., 1995b). This suggests actual differences between these two sensory systems, for instance in the way the deviation magnitude is represented (e.g. linearly, logarithmically). The same studies suggest also similarities between the processing principles of the visual and auditory systems. As for the visual system studies here, also the sensitivity to the alcohol has been shown to be modulated by the auditory feature studied. The auditory MMNs for several auditory attributes such as frequency, duration, location and intensity decrease have been found to decrease under the influence of alcohol but to a different extent, with the frequency processing being most affected (He et al., 2013). These results together suggest that the effects of alcohol are seen widely throughout the sensory systems, but not at all uniformly.

5 Conclusion

This study, applying multi–feature vMMN paradigm, investigated the effects of alcohol on pre–attentive automatic detection of changes of colour, location, and duration. The effects of alcohol on automatic detection of visual changes differed between the different visual features. In particular, the amplitudes of the location and duration vMMNs significantly decreased in the alcohol condition compared with the placebo condition, but the amplitude of the colour vMMN did not change significantly compared with the placebo condition. Moreover, the impairment of alcohol on pre–attentive automatic detection of visual changes was not related to the magnitude of stimulus deviation.

References

Amenedo, E., & Escera, C. (2000). The accuracy of sound duration representation in the human brain determines the accuracy of behavioural perception. *European Journal of Neuroscience, 12*, 2570–2574.

Clifford, A., Holmes, A., Davies, I. R., et al (2010) Color categories affect pre–attentive color perception. *Biological psychology, 85*, 275–282.

Czigler, I. (2007). Visual mismatch negativity: Violation of nonattended environmental regularities. *Journal of Psychophysiology, 21*, 224–230.

Czigler, I., & Sulykos, I. (2010). Visual Mismatch Negativity to Irrelevant Changes Is Sensitive to Task–Relevant Changes. *Neuropsychologia, 48*,1277–1282.

Czigler, I., Balázs, L. & Winkler, I. (2002). Memory–based detection of task–irrelevant visual changes. *Psychophysiology, 39*, 869–873.

Grimm, S., Bendixen, A., Deouell, L.Y., & Schr?ger,

E. (2009). Distraction in a Visual Multi-Deviant Paradigm: Behavioral and Event-Related Potential Effects. *International Journal of Psychophysiology, 72*, 260–266.

Haxby, J. V., Grady, C. L., Horwitz, B., et al (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences, 88*, 1621–1625.

He, J., Li, B., Guo, Y., et al (2013). Effects of alcohol on auditory pre-attentive processing of four sound features: *evidence from mismatch negativity. Psychopharmacology, 225*, 353–360.

Heslenfeld, D. J. (2003). Visual mismatch negativity. In: Polich J (Ed) *Detection of change: Event-related potential and fMRI fidings*. Boston: Kluver Academic Press, 41–59.

Jääskeläinen, I. P., Lehtokoski, A., Alho, K., et al (1995a). Low dose of ethanol suppresses mismatch negativity of auditory event-related potentials. *Alcoholism: Clinical and Experimental Research, 19*, 607–610.

Jääskeläinen, I. P., Pekkonen, E., Alho, K., et al (1995b). Dose-related effect of alcohol on mismatch negativity and reaction time performance. *Alcohol, 12*, 491–495.

Jääskeläinen, I. P., Pekkonen, E., Hirvonen, J., et al (1996). Mismatch negativity subcomponents and ethyl alcohol. *Biological psychology, 43*, 13–25.

Kähkönen, S., Rossi, E. M. & Yamashita, H. (2005). Alcohol impairs auditory processing of frequency changes and novel sounds: a combined MEG and EEG study. *Psychopharmacology, 177*, 366–372.

Kenemans, J. L., Hebly, W., Vanden, Heuvel, E. H. M., et al (2010). Moderate alcohol disrupts a mechanism for detection of rare events in human visual cortex. *Journal of Psychopharmacology, 24*, 839–845.

Kenemans, J., Jong, T. & Verbaten, M. (2003). Detection of visual change: mismatch or rareness? *NeuroReport, 14*, 1239–1242.

Kimura, M., Katayama, J.I., Ohira, H., & Schr?ger, E. (2009). Visual Mismatch Negativity: New Evidence from the Equiprobable Paradigm. *Psychophysiology, 46*, 402–409.

Lewis, P. A. & Miall, R. C. (2003). Distinct systems for automatic and cognitively controlled time measurement: evidence from neuroimaging. *Current Opinion in Neurobiology, 13*, 250–255.

Maunsell, J. H. (1995). The brain's visual world: Representation of visual targets in cerebral cortex. *Science, 270*, 764–768.

Näätänen, R., Paavilainen, P., Rinne, T., et al (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology, 118*, 2544–2590.

Näätänen, R., Pakarinen, S., Rinne, T., et al (2004) The mismatch negativity (MMN): towards the optimal paradigm. *Clinical Neurophysiology, 115*, 140–144.

Pakarinen, S., Takegata, R., Rinne, T., et al (2007). Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event-related potential (ERP). *Clinical Neurophysiology, 118*, 177–185.

Pazo-Alvarez, P., Cadaveira, F. & Amenedo, E. (2003). MMN in the visual modality: a review. *Biological psychology, 63*, 199–236.

Qian, X., Liu, Y., Xiao, B., et al (2014). The visual mismatch negativity (vMMN): Towards the optimal paradigm. *International Journal of Psychophysiology, 93*, 311–315.

Semlitsch, H. V., Anderer, P., Schuster, P., et al (1986). A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology, 23*, 695–703.

Shi, L., Wu, J., Sun, G., Dang, L., & Zhao, L. (2013). Visual Mismatch Negativity in the" Optimal" Multi-Feature Paradigm. *Journal of integrative neuroscience,*

12, 247–258.

Sokolov, E. N. (1960). Neuronal models and the orienting reflex. In: Brazier M A B (Ed), *The central nervous system and behavior.* New York: Macey, pp. 187–276.

Sokolov, E. N. (1963). Higher nervous functions: the OR. *Annual Review of Physiology, 25*, 545–580.

Tzambazis, K. & Stough, C. (2000). Alcohol impairs speed of information processing and simple and choice reaction time and differentially impairs higher–order cognitive abilities. *Alcohol and Alcoholism, 35*, 197–201.

Effects of alcohol on auditory pre-attentive processing of four sound features: Evidence from mismatch negativity

Jinbo He[1,2], Bingbing Li[2], Yongyu Guo[2], Risto Näätänen[3,4,5], Satu Pakarinen[6], Yuejia Luo[1]

([1] State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing 100875, China)

([2] Key Laboratory of Adolescent Cyberpsychology and Behavior of Ministry of Education, School of Psychology, Central China Normal University, Wuhan, China)

([3] Department of Psychology, University of Tartu, Tartu, Estonia)

([4] Center of Functionally Integrative Neuroscience (CFIN), University of Aarhus, Aarhus, Denmark)

([5] Institute of Behavioral Sciences, University of Helsinki, Helsinki, Finland)

([6] Finnish Institute of Occupational Health, Topeliuksenkatu 41 a, 00250 Helsinki, Finland)

Abstracts *Rational* Studies have shown that alcohol could impair automatic pre−attentive change detection. However, several earlier studies which investigated alcohol−induced effects on single auditory feature independently were different from each other on the results. Meanwhile, only few auditory features have been investigated yet. Therefore, it is meaningful to investigate effects of alcohol on multiple auditory features in one experiment. Objectives This study investigates the effects of alcohol on automatic pre−attentive change detection of four kinds of auditory features (frequency, intensity, location, and duration) in one experiment. *Methods* This study, using multi−feature oddball paradigm, compares and analyzes mismatch negativity (MMN) elicited by four kinds of auditory features (frequency, intensity, location, and duration), of 12 participants, under alcohol (0.65 g/kg) and non−alcohol condition. *Results* Compared to non−alcohol condition, amplitudes of all the four MMN types significantly declined under alcohol condition, and their amplitude decline ratios decreased as deviant magnitude became larger. Latencies of frequency and intensity MMN were delayed while latencies of location and duration MMN were not delayed significantly. *Conclusion* Alcohol impaired automatic pre−attentive change detection of all the four auditory features (frequency, intensity, location, and duration). However, the alcoholinduced impairment magnitude on automatic pre−attentive detection of the four auditory features was different from each other. According to analysis of amplitude, frequency seems to be affected most among the four auditory features. According to analysis of latency, only frequency and intensity were affected.

Keywords alcohol; event−related potential; mismatch negativity; multi−feature paradigm; pre−attentive change detection

1 Introduction

The damage in attention function caused by alcohol may cause a variety of dangerous accidents, such as car accidents (Brewer and Sandow 1980; Näätänen and Summala 1976). Many studies have already reported the negative effects of alcohol on active attention, but there are only a few studies on pre–attentive processing in audition. Individuals not only cope with environmental stimuli by engaging their attentive processes but also through their pre–attentive processing. There processes determine whether an object is worthy of attention (Wei and Yan, 2008). Furthermore, pre–attentive processes are of great evolutionary value because of their capability to cope with large amounts of information automatically with no attentive resources.

Mismatch negativity (MMN, for review, see Näätänen et al. 2007) is a change–specific event–related potential (ERP) component elicited by any discriminable change in auditory stimulation. It is a marker for pre–attentive deviance detection (Grimm et al. 2006; Näätänen et al. 1978,2004). Therefore, MNN provides an objective index to study the effects of alcohol on pre–attentive processing, which have already been studied by several researchers (for review, see Ahveninen et al. 2000; Jääskeläinen et al. 1996b).

Jääskeläinen et al. (1995a) used the traditional oddball paradigm and found that 0.50 g/kg (alcohol/body weight) of alcohol significantly decreases the amplitude of MMN elicited by a change in frequency and significantly delays its peak latency. Thereafter, some other researchers also studied the effects of alcohol on frequency MMN (e.g., 0.55 g/kg, Jääskeläinen et al. 1995b); 0.80 g/kg (Kähkönen et al. 2005); and 0.54 mL/kg (Kenemans et al. 2010), using a similar paradigm, and their results were were similar to those of Jäskeläinen et al. (1995a). In addition,

Jääskeläinen et al. (1995b) found that frequency MMN elicited by smaller frequency changes is significantly impaired by a dose of 0.55g/kg but not 0.35g/kg alcohol, whereas MMN elicited by a larger frequency change is not impaired by both doses. Aside from studies on the effects of alcohol on frequency MMN, Jääskeläinen et al. (1996a) also examined the effects of alcohol on MMN elicited by duration change. At an interstimulus interval (ISI) of 800ms, they found that the amplitude and peak latency of duration MMN in subjects given a dose of 0.55 or 0.85 g/kg alcohol do not change significantly. However, at an ISI of 2,400 ms, the frontal MMN amplitude decreases significantly only under the larger dose, whereas the peak latency does not significantly change under either dose. These results suggest the following: (a) both frequency and duration MMN are impaired by alcohol, (b) the effects of alcohol on the amplitude of MMN differ from those on the peak latency, (c) the effects of alcohol on MMNs of different sound features differ from each other, and (d) alcohol impairment on MMNs is more obvious when the intake exceeds a certain dose. However, further studies should still be conducted to gain a better understanding regarding the effects of alcohol on auditory pre–attentive processing. First, earlier studies only tested the effects of alcohol on frequency and duration MMN. The effects of alcohol on MMNs elicited by other sound features should also be studied. Second, all earlier studies used the traditional single–feature oddball paradigm with pure sinusoidal tones. As a result, the external validity of these studies is affected because different sound features always appear at the same time and elicit mixed effects in real–life situations. Therefore, we believe that using the multi–feature MMN paradigm would allow the comparison of the effects of alcohol on the pre–attentive processing of different sound features under the same dose. In addition, it would also improve external validity by examining the effects of alcohol on MMNs elicited by different sound

features in one experiment.

Accordingly, the present study examined the effects of alcohol on the MMNs of four different sound features (frequency, intensity, location, and duration) by applying the multi−feature MMN paradigm promoted by Näätänen et al. (2004) and Pakarinen et al. (2007) to test whether alcohol affects the pre−attentive processing of different sound features differently. In addition, three deviant tones were used to test the effect of deviance magnitude on the effects of alcohol on MMN. In this study, we hypothesized the following: (a) alcohol impairs all MMNs elicited by the four sound features, (b) the effects of alcohol on MMN decrease as the magnitude of deviation of stimuli becomes larger, and (c) the effects of alcohol on MMNs elicited by the four sound features may differ from each other in terms of effect magnitude.

2 Materials and methods

2.1 Subjects

Twelve participants (right−handed, aged 19 to 26 years, one female) with normal hearing took part in this experiment. All of them were older than 18 and thus legitimate to consume alcohol in China. All participants were healthy normal social drinkers (3 to 12 standard alcohol per month during the past year) who had no chronic alcoholism or family history of alcoholism and other mental illnesses. They were asked to abstain from caffeine, alcohol, nicotine, and other psychoactive substances for 24 h prior to the experiment. They signed an informed consent form and committed to obey the above request. All participants were paid for taking part in this study. This study was approved by the institutional ethical committee of National Key Laboratory of Cognitive Neuroscience and Learning of China.

2.2 Treatment

All participants attended two experimental sessions (alcohol and placebo conditions) separated by 2 weeks. The order of alcohol and placebo sessions was counterbalanced across all participants. A single−blind procedure was employed in both sessions. Participants in the alcohol condition received a dose of 0.65 g/kg alcohol (53% v/v white wine) and were provided with little food (e.g., peanuts). They were given 10min to finish the drink. The breath alcohol concentration (BrAC) level was tested every 5 min after they finished drinking until the level was steady. The mean time interval between the finish of drinking and the beginning of the experiment was 18.75 minutes. The BrAC level was also tested immediately after the experiment finished. The average BrAC level was 0.25 (± 0.05) mg/L before and 0.22 (± 0.06) mg/L after the experiment. Participants in the placebo condition received a dose of 0.02 g/kg alcohol (53% v/v white wine mixed with distilled water) and were provided with little food. The same BrAC level test procedure as in the alcohol condition was performed after drink intake was finished. The time interval between drink intake and the experiment was 15min. The BrAC level was zero both before and after the experiment in the placebo condition. The mean BrAC level in the alcohol condition was significantly higher than that in the placebo condition ($t_{(11)}$ = 15.70, $p<0.01$). In addition to BrAC level measurement, the participants were asked to report their intoxication on a five−point Likert−type scale right before and after the experiment in both conditions. The subjective report of intoxication was 2.67 (± 0.52) before and 2.50 (± 0.37) after the experiment in the alcohol condition, whereas it was 1.00 (± 0.00) before and 1.08 (± 0.28) after the experiment in the placebo condition. The subjective report of intoxication in the alcohol condition was significantly larger than that in the placebo condition ($t_{(11)}$ = 6.37, $p < 0.01$).

2.3 Stimulus design and task

The stimuli and experimental procedure used were similar to Pakarinen et al. (2007). The standard tones

were harmonic tones of 75 (± 5) ms composed of three sinusoidal partials (523, 1046, and 1569 Hz), with the second and third partials at 3 and 6 dB lower in intensity, respectively. They were binaurally presented via headphones at an intensity of 70 dB. The deviant tones, with the magnitude of the deviation varying across the three levels, differed from the standards in terms of frequency, intensity, duration, or perceived sound−source location. The frequency deviants differed from the standards by 3/8, 10/8, and 21/8 semitones in the Western musical scale (fundamental frequencies 512, 487, and 450 Hz). The intensity deviants were softer than the standards by steps of 5 dB (65, 60, and 55 dB). The location deviants were tones perceived 10°, 40°, or 90° to the left or right of the participant. The duration deviants were shorter than the standards by steps of 16 ms (59, 43, and 27 ms).

The tones were presented in 5.5 min sequences (6 sequences in total, 628 tones per sequence, each of the 12 deviants was presented 156 times, and all sequences beginning with 4 successive standards), with the presentation order of the sequences randomly varying across the subjects. The stimulus−onset asynchrony was 500 ms. During the period of stimuli presentation (33 min), the subjects were asked to watch a silent video film and ignore the auditory stimuli.

2.4 Data acquisition

The electroencephalogram (EEG) was recorded (0 to 40 Hz, sampling rate of 500 Hz) by the NeuroScan system using a 64−channel Ag/AgCl electrode cap. An electrode was placed on the tip of the nose to serve as a reference channel. Both bipolar horizontal and vertical electrooculograms were recorded between electrodes placed at 1 cm from the canthi of the eyes. The EEG was filtered offline (pass band of 1 Hz to 30 Hz). Eyemovement artifacts were removed using the correlation method. Epochs of 600 ms (including a 100−ms pre−stimulus period served as a baseline for the amplitude measurement) were separately averaged for tones of different types and levels. Epochs of EEG elicited by the first eight tones of each sequence and exceeding ± 75 μVwere omitted from averaging.

The response to the standard tones was subtracted from the response to each type and level of deviant tones to derive MMNs. The most negative peak occurring at 100 to 250 ms after stimulus onset of the Fz channel was selected as the MMN peak amplitude and peak latency. We noticed two peaks during the MMN timewindow of intensity MMN. With reference to an earlier study (Jacobsen and Schröger 2001) and our observation of the wave maps of intensity deviant tones, the first peak was caused by the difference of N1 between intensity deviant tones and standard tones, which reflected the refreshing of neural cells. The second peak was the true MMN caused by the changes of intensity deviant tones compared with the standard tones, which reflected the automatic processing of stimulus change based on sensory memory. Therefore, we selected the amplitude of the second peak as the amplitude of intensity MMN.

2.5 Data analysis

The distribution of dependent measures of our experiment were not significantly different from normal distribution according to Kolmogorov−Smirnova test on the dependent measures($P>0.05$).To examine the effects of alcohol on the MMNs of different types and levels of deviant tones, a two−way repeated measures analysis of variance (ANOVA) was conducted separately for MMNs of different types, with the factors being condition (placebo and alcohol) and magnitude of deviation (small, medium, and large magnitudes).

To compare the different effects of alcohol on MMNs among the four types, a two−way repeated ANOVA was conducted separately for the MMN amplitude decline ratio, which is calculated as ($amplitude_{placebo\ condition}$ −$amplitude_{alcohol\ condition}$)/$amplitude_{placebo\ condition}$, and the peak latency delay ratio, calculated as ($peak\ latency_{alcohol\ condition}$ − $peak\ latency_{placebo\ condition}$)/ $peak\ latency_{placebo\ condition}$ for

Table 1. MMN amplitudes and latencies in placebo and alcohol condition (mean ± SD)

Feature	Deviant level	Amplitude (μv)		Latency (ms)	
		Placebo	Alcohol	Placebo	Alcohol
Frequency	Small	−0.96(.32)	−0.57(.39) *	183(19)	202(12) **
	medium	−1.62(.33)	−1.28(.34) *	162(15)	184(14) *
	large	−2.39(.26)	−1.87(.24)	149(17)	160(26)
Intensity	Small	−0.80(.29)	−0.44(.30) **	184(33)	198(35) **
	medium	−1.35(.33)	−1.07(.34) *	178(22)	188(25)
	large	−1.68(.28)	−1.36(.24)	173(35)	186(36)
Location	Small	−1.02(.54)	−0.76(.45) *	202(26)	200(37)
	medium	−1.45(.30)	−1.17(.29) *	150(22)	152(15)
	large	−1.52(.36)	−1.31(.29)	152(19)	148(25)
Duration	Small	−0.91(.33)	−0.65(.30) **	184(16)	183(21)
	medium	−2.01(.35)	−1.57(.32) **	174(22)	173(25)
	large	−2.58(.24)	−2.49(.27)	154(17)	156(20)

* $P<0.05$, ** $P<0.01$

MMNs of different types and levels. The factors used were the type (frequency, intensity, location, and duration) and magnitude (small, medium, and large magnitudes) of deviation. Greenhouse–Geisser correction was used in ANOVA when appropriate, and Bonferroni correction tests were carried out as post hoc analysis.

3 Results

3.1 Amplitude

The results of two–way repeated ANOVA showed no significant interaction effects between the condition (placebo and alcohol) and magnitude of deviation (small, medium, and large magnitudes) for frequency, intensity, and location MMNs. The main effect of condition was significant for MMNs of all these three sound features (frequency, $F_{1,11}=8.027$, $p<0.05$; intensity, $F_{1,11}=9.745$, $P<0.05$; location, $F_{1,11}=12.388$, $p<0.01$). The amplitude of MMNs for each magnitude level was larger in the placebo condition than in the alcohol condition for all the three types of MMNs. Meanwhile, the main effect of magnitude of deviation was also significant for MMNs of all these

Figure. 1. Grand average MMN for electrode site Fz of different types and levels. From left to right are frequency, intensity, location, and duration MMMs. *Rows 1, 2, and 3* show MMNs of small, medium, and large magnitudes of deviation, respectively.
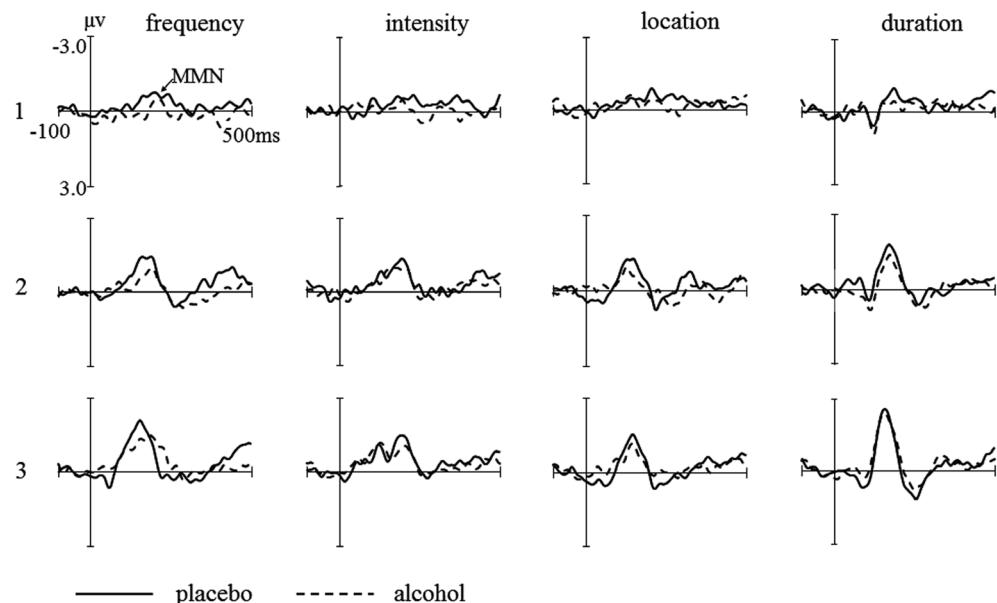
three types (frequency, $F_{1,11}$=26.066, p<0.01; intensity, $F_{1,11}$=34.180, p<0.01; location, $F_{1,11}$=18.704, p<0.01). The amplitude of MMNs became larger as the magnitude of deviation became larger for all the three types of MMNs. However, the interaction effect between condition and magnitude of deviation was significant for duration MMN ($F_{1,11}$=3.586,p<0.05). Analysis of the simple effect of condition showed that duration MMNs elicited by deviant stimuli of small and medium magnitude levels were significantly larger in the placebo condition than in the alcohol condition (small magnitude, $F_{1,11}$=14.19, p<0.01; medium magnitude, $F_{1,11}$=10.11, p<0.01). However, the amplitude of duration MMN elicited by large magnitude level was not significantly changed in the alcohol condition compared with the placebo condition (p>0.1; See Fig1). The absolute value of amplitude of MMNs was illustrated in Table 1(significant level in it was based on paired t test).

3.2 Peak latency

The results of two−way repeated ANOVA showed no significant interaction effects between condition (placebo and alcohol) and magnitude of deviation (small, medium, and large magnitudes) for all the four types of MMNs (p>0.1). The main effect of condition was significant for both frequency and intensity MMNs (frequency, $F_{1,11}$=20.203, p<0.01; intensity, $F_{1,11}$=8.488, p<0.05). The peak latency of MMNs was delayed in the alcohol condition compared with the placebo condition for both frequency and intensity MMNs. Meanwhile, the main effect of magnitude of deviation was also significant for MMNs of both sound features (frequency, $F_{1,11}$=20.437, p<0.01; intensity, $F_{1,11}$=16.462, p<0.01). The peak latency of MMNs of both types became shorter as the magnitude of deviation became larger. The main effect of condition was not significant for both location and duration MMNs (p>0.1), whereas the main effect of the magnitude of deviation was significant (location, $F_{1,11}$=15.903, p<0.01; duration, $F_{1,11}$=20.396, p<0.01). The peak latency of MMNs of both
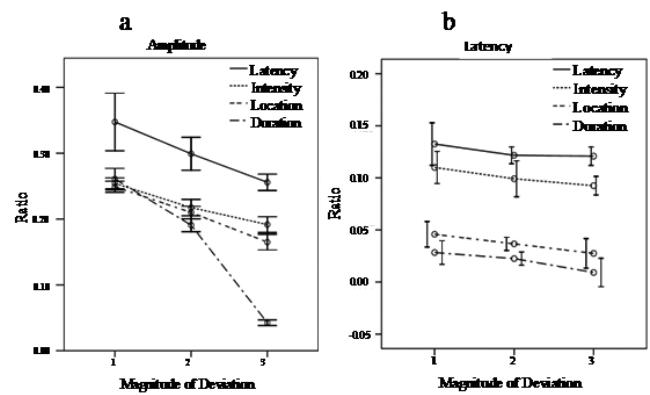
Figure. 2. (a) Amplitude decline ratios of MMNs of different types and levels. (b) Peak latency delay ratios of MMNs of different types and levels. *Numbers 1, 2,* and *3* in the *x*−axis each represent small, medium, and large magnitudes of deviation, respectively.

types also became shorter as the magnitude of deviation became larger （See Fig1）. The absolute value of latency of MMNs was illustrated in table 1(significant level in it was based on paired t test).

3.3 Amplitude decline ratio

The interaction effect of type (frequency, intensity, location, and duration) and magnitude of deviation (small, medium, and large magnitudes) was significant ($F_{6,66}$=7.49, p<0.01). Based on simple effect analysis, the comparison of the different types showed that the amplitude decline ratio was significantly different from one another for MMNs of the four types under medium and large magnitudes of deviation (medium, $F_{3,33}$=4.89, p<0.01; large, $F_{3,33}$= 19.96, p<0.01). The amplitude decline ratio of frequency MMN was larger than all the other three types of MMNs under both medium and large magnitudes of deviation. No differences were observed among the amplitude decline ratios of intensity, location, and duration MMNs under medium magnitude of deviation. However, the amplitude decline ratios of intensity and location MMNs were larger than duration MMNs under large magnitude of deviation, whereas no differences were observed between the amplitude decline ratios of MMNs elicited by both sound features.

The comparison within types showed that the effects of magnitude of deviation were significant for intensity,

location, and duration MMNs (intensity, $F_{2,22}$=7.26, $p<$ 0.01; location, $F_{2,22}$=4.79, $p<0.05$; duration, $F_{2,22}$= 19.07, $p<0.01$), whereas the effects of magnitude of deviation were approaching significance (p=0.068) for the amplitude decline ratios of frequency MMN. The amplitude decline ratio decreased as the magnitude of deviation became larger (See Fig. 2).

3.4 Peak latency delay ratio

The results of two-way repeated ANOVA showed that neither the interaction effects between type (frequency, intensity, location, and duration) and magnitude of deviation (small, medium, and large magnitudes) nor the main effects of magnitude of deviation were significant for peak latency delay ratio ($p>0.1$). However, the main effect of type was significant ($F_{3,33}$=4.73, $p<0.01$). Post hoc comparison based on Bonferroni correction showed that the peak latency decay ratios of frequency and intensity MMNs were larger than those of location and duration MMNs ($p<0.05$). Neither the differences between the peak latency delay ratios of frequency and intensity MMNs nor the differences between the peak latency delay ratios of location and duration MMNs were significant ($p>0.1$；See Fig. 2）.

4 Discussion

The results from the analysis of amplitude and peak latency of MMNs indicated that alcohol elicited different effects on the amplitude and peak latency of MMNs. The main effect of condition (alcohol and placebo) was significant for MMNs of all the four sound features. This result suggested that the amplitude of MMNs was larger in placebo condition than in the alcohol condition for MMNs of all types and levels, except for the duration MMNs elicited by tones of large magnitude of deviant levels. Moreover, the ability of pre-attentive change to detect changes in sound features (frequency, intensity, location, and duration) was significantly impaired by a

dose of 0.65 g/kg alcohol. This result is consistent with our first hypothesis. In addition, the effects of alcohol on frequency and duration MMNs in this study are similar to those reported in earlier studies (Jääskeläinen et al. 1996a，1995a,b; Kähkönen et al. 2005; Kenemans et al. 2010). This study confirmed the results of earlier studies. In addition, it also extended the results by proving that alcohol not only impairs the pre-attentive detection of frequency and duration auditory changes but also impairs that of intensity and location auditory changes.

However, the impairment of peak latency was not as steady as that of amplitude. The main effect of condition (alcohol and placebo) was significant for frequency and intensity MMNs but not for location and duration MMNs. Earlier studies also found that alcohol elicits different effects on amplitude and peak latency. Jääskeläinen et al (1996a) found that the amplitude of duration MMN was significantly decreased after alcohol drinking, whereas the peak latency of duration MMN did not significantly change. These results indicated that the amplitude of MMN was more sensitive to alcohol than peak latency.

Amplitude decline ratio decreased as the magnitude of deviation became larger. This result suggested that alcohol impairment of pre-attentive change detection decreased as the magnitude of deviation became larger. Alcohol did not affect simple change detection but only the more complicated and subtle sensory perceptive processing of stimuli. Jääskeläinen et al (1995b) also found that the frequency MMN of more widely deviant stimuli did not change significantly, whereas the MMN of less deviant stimuli decreased significantly after alcohol ingestion. They attributed this result to the higher threshold for the pre-attentive detection of acoustic deviations after alcohol ingestion, which meant that the pre-attentive change detection of subtle changes was impaired. Earlier studies (Jääskeläinen et al. 1996a，b) also found that N1 was not impaired by alcohol, whereas MMNs were significantly

impaired. Kenemans et al. (2010) found that the amplitude of visual SFD80 (spatial frequency−dependent difference at 80 ms; Kenemans et al. 2000) did not change after alcohol ingestion, whereas visual mismatch negativity (for reviews, see Czigler 2007; Pazo−Alvarez et al. 2003) decreased significantly. N1 and SFD80 are ERP components that reflect sensory processing, whereas MMN reflects memory−dependent processing.

The main effect of type (frequency, intensity, location, and duration) for the amplitude decline ratio was significant under medium and large magnitudes of deviation level. The decline ratios of frequency MMNs were significantly larger than all the other three MMN types under both levels. The decline ratios of intensity and location MMNs were larger than duration MMNs under both small and medium magnitudes, whereas no significant differences were found between them under both levels. This is somehow unclear. These results indicated that alcohol impaired the frequency MMN the most, intensity and location MMN less, and duration MMN the least of all the four MMN types. The amplitude decline ratio of duration MMNs elicited by tones of large magnitude of deviation level was significantly less than that of the small and medium levels and that of the other MMN types under the same deviant level. No differences were found between the amplitude of duration MMNs elicited by the large magnitude of deviation level in the placebo and alcohol conditions (Fig.1)，which indicated that alcohol elicited no effect on the pre−attentive change detection of temporal features of tones, when the magnitude of deviation was excessively large compared with the standard tones that pre−attentive processing was not deep enough.

The effects of feature and magnitude of deviation level on peak latency delay ratio were different from those on amplitude decline ratio. No significant interaction effects were found between the type (frequency, intensity, location, and duration) and magnitude (small, medium, and large magnitudes) of deviation. The main effect of magnitude of deviance level was not significant for any of the four MMN types. Different from the amplitude decline ratio, the peak latency delay ratio did not decrease as the magnitude of deviance level became larger. This result implied that peak latency was not as sensitive to the effects of alcohol as amplitude. In addition, the differences in peak latency delay ratio between the different sound features were also different from those in amplitude decline ratio. Similar peak latency delay ratios were found between frequency and intensity MMNs as well as between location and duration MMNs under the same magnitude of deviation level. However, the peak latency delay ratios of frequency and intensity MMNs were larger than those of location and duration MMNs. Further studies should be conducted to determine why the effects of alcohol on amplitude and peak latency were different. However, earlier studies reported that peak latency is not as steady as amplitude to be used as an index of the effects of alcohol on MMN. Therefore, further studies are necessary to verify the results based on peak latency.

The results showed that effects of alcohol on frequency and intensity MMNs as well as on location and duration MMNs were similar. Moreover, the effects of alcohol on the former two MMN types were different from those of the latter two MMN types. This result indicated that the effects of alcohol on the pre−attentive change detection of spectrum auditory information (frequency and intensity) were different from those of temporal information (duration). This result may be attributed to the fact that pre−attentive change detection of spectrum information is different from that of temporal information. Whether acoustic features are processed independently or pre−attentively integrated has been under debate (e.g. Giard et al. 1995; Schairer et al. 2001; Winkler et al. 1996). Recent studies have supported the independent view. Grimm et al. (2006) found a right hemisphere preponderance for frequency MMN

but not for duration MMN. Molholm et al. (2005) found that anatomically distinct networks of auditory cortices are activated by different acoustic features (frequency and duration) using functional magnetic resonance imaging technology. Changes in duration activated both the left and right frontal cortices, whereas changes in frequency only activated the right frontal cortex. Frequency and intensity features reflected spectrum information of tones, whereas location (the different initial time between two ears in this study) and duration features reflected temporal information of tones. Therefore, the results of this study not only supported the view that the pre-attentive processing of acoustic features occurred independently, but they also suggested that pre-attentive change detection of spectrum information was more sensitive to alcohol than temporal information.

This study used the multi-feature MMN paradigm to examine the effects of alcohol on MMNs of four sound features (frequency, intensity, location, and duration). In conclusion, this study showed that: (a) The MMNs of all the four sound features were impaired by alcohol, and the impairment decreased as the magnitude of deviation level became larger. In addition, the amplitude of MMN was more sensitive to alcohol than peak latency; (b) The trends of the effects of alcohol on the MMNs of different sound features differed from each other. In terms of the amplitude decline ratio, frequency MMN was impaired the most, intensity and location MMN were less impaired, and duration MMN was impaired the least. Meanwhile, the peak latency delay ratios of frequency and intensity MMNs were larger than those of location and duration MMNs, whereas no differences were observed between the latter two MMNs.

References

Ahveninen, J., Escera, C., Polo, M. D., Grau, C., & Jääskeläinen, I. P. (2000). Acute and chronic effects of alcohol on preattentive auditory processing as reflected by mismatch negativity. *Audiol Neuro-Otol, 5*, 303 - 311.

Brewer, N., & Sandow, B. (1980). Alcohol effects on driver performance under conditions of divided attention. *Ergonomics, 23*, 185 - 190.

Czigler, I. (2007). Visual mismatch negativity: Violation of nonattended environmental regularities. *Journal of Psychophysiology, 21*, 224 - 230.

Giard, M., Lavikahen, J., Reinikainen, K., Perrin, F., Bertrand, O., Pernier, J., & N??t?nen, R. (1995). Separate representation of stimulus frequency, intensity, and duration in auditory sensory memory: An event-related potential and dipole-model analysis. *Journal of Cognitive Neuroscience, 7*, 133 - 143.

Grimm, S., Roeber, U., Trujillo-Barreto, N. J., & Schroger, E. (2006). Mechanisms for detecting auditory temporal and spectral deviations operate over similar time windows but are divided differently between the two hemispheres. *Neuroimage, 32*, 275 - 282.

Jääskeläinen, I., Pekkonen, E., Hirvonen, J., Sillanaukee, P., & Näätänen, R. (1996a). Mismatch negativity subcomponents and ethyl alcohol. *Biological psychology, 43*, 13 - 25.

Jääskeläinen, I. P., Lehtokoski, A., Alho, K., Kujala, T., Pekkonen, E., Sinclair, J. D., Näätänen, R., & Sillanaukee, P. (1995a). Low dose of ethanol suppresses mismatch negativity of auditory event-related potentials. *Alcoholism: Clinical and Experimental Research ,19*, 607 - 610.

Jääskeläinen, I. P., Näätänen, R., & Sillanaukee, P. (1996b). Effect of acute ethanol on auditory and visual event-related potentials: A review and reinterpretation. *Biological psychiatry, 40*, 284 - 291.

Jääskeläinen, I. P., Pekkonen, E., Alho, K., Sinclair, J. D., Sillanaukee, P., & Näätänen, R. (1995b). Dose-related

effect of alcohol on mismatch negativity and reaction time performance. *Alcohol, 12*, 491－495.

Jacobsen, T., & Schr?ger, E. (2001) Is there pre－attentive memory－based comparison of pitch?. *Psychophysiology, 38*, 723－727.

Kähkönen, S., Rossi, E. M., & Yamashita, H. (2005). Alcohol impairs auditory processing of frequency changes and novel sounds: a combined MEG and EEG study. *Psychopharmacology, 177*, 366－372.

Kenemans, J., Baas, J., Mangun, G., Lijffijt, M., & Verbaten, M. (2000). On the processing of spatial frequencies as revealed by evoked－potential source modeling. *Clinical Neurophysiology, 111*, 1113－1123.

Kenemans, J. L., Hebly, W., vanden, Heuvel, E. H. M., Grent, T., & Jong, T. (2010). Moderate alcohol disrupts a mechanism for detection of rare events in human visual cortex. *J Psychopharmacol, 24*, 839－845.

Molholm, S., Martinez, A., Ritter, W., Javitt, D. C.,& Foxe, J. J. (2005). The neural circuitry of pre－attentive auditory change－detection: an fMRI study of pitch and duration mismatch negativity generators. *Cerebral Cortex, 15*, 545－551.

Näätänen, R., Gaillard, A. W. K., & M?ntysalo, S. (1978). Early selective－attention effect on evoked potential reinterpreted. *Acta psychologica, 42*, 313－329

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical Neurophysiology, 118*, 2544－2590.

Näätänen, R., Pakarinen, S., Rinne, T., & Takegata, R. (2004). The mismatch negativity (MMN): towards the optimal paradigm. *Clinical Neurophysiology, 115*, 140－144.

Näätänen, R., & Summala, H. (1976). Road－user behavior and traffic accidents. North－Holland Pub. Co.(Amsterdam and New York)

Pakarinen, S., Takegata, R., Rinne, T., Huotilainen, M., & Näätänen, R. (2007). Measurement of extensive auditory discrimination profiles using the mismatch negativity (MMN) of the auditory event－related potential (ERP). *Clinical Neurophysiology, 118*, 177－185.

Pazo－Alvarez, P., Cadaveira, F., & Amenedo, E. (2003). MMN in the visual modality: a review. *Biological psychology, 63*, 199－236.

Schairer, K. S., Gould, H. J., & Pousson, M. A. (2001). Source generators of mismatch negativity to multiple deviant stimulus types. *Brain topography, 14*, 117－130.

Wei, J. H., & Yan, K. L. (2008). *Fundametals of Cognitive Neuroscience.* People's Education Press, Beijing.

Winkler, I., Karmos, G., & Näätänen, R. (1996). Adaptive modeling of the unattended acoustic environment reflected in the mismatch negativity event－related potential. *Brain Research, 742*, 239－252.

Cultural differences in the representativeness heuristic: Expecting a correspondence in magnitude between cause and effect

Roy R. Spina[1], Lijun Ji[1], Tieyuan Guo[1], Zhiyong Zhang[2], Ye Li[3], Leandre Fabrigar[1]

([1] Queen's University, Kingston, Ontario, Canada)

([2] Beijing University, Beijing, China)

([3] Huazhong Normal University, Wuhan, China)

Abstract   Based on previous research on cultural differences in analytic and holistic reasoning, it was hypothesized in these studies that when explaining events, North Americans would be more likely than East Asians to expect causes to correspond in magnitude with those events (i.e., big events stem from big causes and small events stem from small causes). In a series of studies, Canadian and Chinese participants judged the likelihood that high- or low-magnitude events were caused by high- or low-magnitude causes. Overall, Canadians expected events and their causes to correspond in magnitude to a greater degree than did Chinese. Also, Canadians primed to reason holistically expected less cause – effect magnitude correspondence than did those primed to reason analytically.

Keywords   heuristics; representativeness; holism; attribution; culture and cognition

1 Introduction

In the 1960s, mathematician and meteorologist Edward Lorenz created a computer simulation of hydrodynamic flow (Lorenz, 1963). Allegedly, while using the program to model weather patterns, Lorenz entered the value .506 rather than the actual value .506127. He was surprised to find that the outcomes stemming from the two ostensibly similar initial values varied substantially. Upon publishing the findings, one meteorologist remarked that if the theory were correct, one flap of a seagull's wings could change global weather patterns forever. Over time, through some

---

form of cultural metamorphosis, the ungainly seagull transformed into a delicate butterfly and the term *butterfly effect* was born. Regardless of whether this amusing anecdote is true, Lorenz found that seemingly negligible variations in initial conditions led to dramatic divergences in outcome.

The butterfly effect has surprised many of us because it violates our expectations of cause‐effect magnitude correspondence. Why do we expect big causes to lead to big effects and small causes to lead to small effects? Is it simply because we seldom observe violations of this association and thus have developed a heuristic that serves us well most of the time? The present research indicates that the explanation is far more interesting and complex.

1.1 The representativeness heuristic

categorization. Kahneman and Tversky (1973) demonstrated that when categorizing targets into groups, people relied mostly on the degree of similarity between the target and a prototypical member of each group. For example, when judging a fictional student's academic discipline based on his personality, people relied on the personalities of typical students from each discipline. Furthermore, the authors demonstrated that this strong tendency to rely on the degree of similarity caused errors in judgment because people ignored the percentages of students enrolled in each discipline, the base rates. Kahneman and Tversky (1972) referred to this type of judgment as the *representativeness heuristic.*

causal judgments. Most of the research conducted on the representativeness heuristic has focused on judgments people make in categorization contexts (Gilovich & Savitsky, 2002). However, the representativeness heuristic is not limited to categorization. People also use this heuristic when making causal judgments, such as when searching for causes that are similar to an effect. Relatively little research has been conducted on the representativeness heuristic in the domain of causal

judgments. Nonetheless, across a variety of domains including medicine, pseudoscientific systems, and psychoanalysis, people employ the representativeness heuristic when making causal judgments by relying on similarities between causes and effects (Gilovich & Savitsky, 2002).

Einhorn and Hogarth (1986) divided such similarities into two categories: physical resemblance and congruity of strength. Beliefs that causes physically resemble effects are prevalent in traditional medicine, for example, when cures that resemble diseases are sought (Gilovich & Savitsky, 2002). The second category, the tendency to search for causes that resemble effects in congruity of strength, is especially interesting because strength has been identified as one of the fundamental meaningful dimensions people use when judging entities (Osgood, 1957, referred to as potency and included adjective pairs such as *large−small and strong−wea*k). Nisbett and Ross (1980) also proposed that people may seek causes that correspond in magnitude with events they are trying to explain. However, few studies have empirically tested people's expectations of a correspondence in magnitude between causes and effects.

1.2 Empirical research on expectations of cause−effect magnitude correspondence

In one of the first studies that indirectly provided supporting evidence that people expect cause‐effect magnitude correspondence, Shultz and Ravinsky (1977) demonstrated that French Canadian schoolchildren typically chose causes that were similar to effects. For example, in one scenario, they attributed a loud noise to a heavy rather than a delicate lever. However, the other scenarios used would fall under Einhorn and Hogarth's (1986) physical resemblance category, such as attributing physical retaliation to physical rather than verbal aggression.

McCauley and Jacques (1979) also provided indirect supporting evidence of an expectation of cause‐effect

magnitude correspondence. In their study, American participants read about a successful or unsuccessful assassination. Participants then estimated the probabilities that the assassinator was acting alone or acting as a member of a group. The authors found that participants attributed the more consequential effect, the successful assassination, to the group and the less consequential effect, the unsuccessful assassination, to the individual. However, this study does not provide strong evidence supporting an expectation of cause‐effect magnitude correspondence because participants could be attributing the more consequential effect to a conjunction of causes, a group, more than a single cause, a lone gunman.

McClure, Lalljee, and Jaspars (1991) examined whether people explained extreme and moderate effects by using a conjunction of causes or a single cause. For example, British participants read an extreme crime involving multiple murders and mutilation of bodies, and a moderate crime involving hitting people with a bottle at a football match. Most participants explained the extreme crime by generating single–cause explanations that tended to correspond in magnitude with the effects.

A recent set of studies by Ebel–Lam, Fabrigar, MacDonald, and Jones (2008) provided more direct support for an expectation of cause‐effect magnitude correspondence. Canadian participants read a scenario describing either a high– or moderate–magnitude effect. For example, participants read either that a plane crashed killing everybody onboard (high magnitude) or that with difficulty the pilot successfully landed the plane (moderate magnitude). Participants estimated the likelihood that a number of high– and moderate–magnitude causes had led to the effect. Participants attributed high–magnitude effects to high–magnitude causes and low–magnitude effects to low–magnitude causes.

Either directly or indirectly, the aforementioned studies provided consistent evidence that people (at least,

Americans, British, and Canadians) typically expect a correspondence in magnitude between an effect and its cause. Why is this so and is it true across cultures? Research from the cultural psychology literature may provide some clues.

1.3 Cultural differences in analytic and holistic reasoning

Contrary to a Washington newspaper article mocking Walter Reed's suggestion that yellow fever with all of its devastating effects was caused by a tiny mosquito (Nisbett & Ross, 1980), Asian folk wisdom states, "One tiny insect may be enough to destroy a nation." Research indicates that compared with North Americans (including Americans and Canadians), East Asians (including Chinese, Japanese, and Koreans) differ in their reasoning. In particular, East Asians tend to reason holistically, whereas North Americans tend to reason analytically (Nisbett, Peng, Choi, & Norenzayan, 2001).

Across a variety of domains, East Asians attend to situational factors or contexts more than North Americans do, and North Americans attend to focal people or objects more than East Asians do (Miller, 1984; Morris & Peng, 1994). For example, one set of studies found that when describing an underwater scene, the first element Americans typically mentioned was a focal element, such as a large fish in the center of the picture. In contrast, the first element Japanese typically mentioned was a contextual element, such as seaweed (Masuda & Nisbett, 2001). In addition, Japanese focused on the entire scene as a unit, and their ability to recall a focal object from the scene was impaired when the background was altered. Alternatively, Americans focused on the focal objects independent of the background, and changing the background had little or no effect on their ability to recall a focal object.

Focusing relatively different amounts of attention on focal objects and contexts has implications for other aspects of cognition, such as the ability to detect

covariation among stimuli. People who focus on focal objects more than contexts should be less likely to detect covariation between elements in a scene compared with people who attend to both focal objects and contexts. Ji, Peng, and Nisbett (2000) asked participants to estimate the degree of covariation between images on a computer screen and this is exactly what they found. That is, compared with American estimates, Chinese estimates were better calibrated with actual levels of covariation.

If Easterners attend more to contextual elements and notice a greater degree of covariation between elements than do North Americans, then they should explain effects differently. This pattern has been found when people make attributions. When people observe a person's behavior in a social situation, a number of elements are present other than the person. That is, other people are often involved and there is a surrounding context or situation in which the person is acting. If an observer focuses relatively more attention on the situation, then he or she should be more likely to attribute causality to that situation. Alternatively,if an observer focuses relatively more attention on the focal person, then he or she should attribute causality to that person. Consistent with this reasoning, past research has shown that Asians tend to make more situational attributions whereas Americans tend to make more dispositional attributions (e.g., Miller, 1984, with Indian participants; Morris & Peng, 1994, with Chinese participants).

Choi, Dalal, Kim-Prieto, and Park (2003) further explicated the cultural differences in reasoning by providing evidence that Easterners may have more complex causal theories and therefore consider more causal factors in their attributions than do Westerners, who may have relatively simple causal theories. For example, in Choi et al., American and Korean participants read that a graduate student killed his or her advisor, along with 97 pieces of information related to the student or advisor. When asked to select items that were pertinent to establishing a motive for the murder, Koreans considered a greater number of items as relevant than did Americans. Additionally, Choi et al. developed and included a 10-item measure of holistic tendency. The measure included statements such as, "Any phenomenon has numerous numbers of causes, although some of the causes are not known." They found that Koreans were more holistic than were Americans, and within each culture, the higher a participant's holistic tendency, the greater the number of items considered relevant to establishing a motive.

1.4 The present research

To summarize the cross-cultural literature, East Asians reason holistically whereas North Americans reason analytically. It is important to note that holism does not appear to be a simple uniform construct determined by a single cognitive mechanism. For example, researchers have focused on at least four factors under the umbrella term holism: causality (Choi et al., 2003), attitude toward contradiction (e.g., Peng & Nisbett, 1999), perception of change (Ji, Nisbett, & Su, 2001), and locus of attention (e.g., Masuda & Nisbett, 2001; see also Nisbett et al., 2001). The present research focused on the causality factor. With respect to causality, East Asians reason holistically by focusing on many causes, whereas North Americans reason analytically by focusing on relatively fewer causes (Choi et al., 2003).

If an observer tends to focus on one or a few causes only when explaining an effect, then the observer should expect a greater correspondence in magnitude between cause and effect. For example, imagine judging how likely it is that each of two buildings sold for a high price. The two buildings look very similar except that one is larger than the other. If all else were equal, one would expect the larger building to sell for more because size is one factor that determines the value of a building. However, there are a number of other factors that also determine the value of a building, such as location. If an observer tends to focus

on only one cause, such as the size of the building in this case, when trying to understand a high-magnitude effect, the high selling price, then that person should expect a high-magnitude cause, the large building, to be far more likely than a low-magnitude cause, the small building.

Alternatively, if an observer tends to focus on many factors when explaining an effect, then the observer should expect a lesser correspondence in magnitude between cause and effect. Considering the previous example, if an observer tends to focus on numerous factors, including size, location of the building, and other factors, when trying to understand a high-magnitude effect, such as a high selling price, then a high-magnitude cause, the large building, is not as necessary. Although this multiple-cause observer may also reason that the large building would likely sell for more than the small building, this effect would tend to be less extreme when compared with the observer with a single-cause focus. The lesser correspondence between the magnitudes of cause and effect makes sense because one of the other factors, such as location, could be working against the larger building and in favor of the smaller building when it comes to selling price.

Therefore, based on previous research on cultural differences in analytic and holistic reasoning, specifically regarding differences in perceptions of causal complexity, we hypothesized that expectations of a correspondence in magnitude between effects and their causes would be stronger among North Americans than among East Asians. Specifically, we hypothesized that when explaining an effect, North Americans would tend to look for causes that correspond in magnitude with the effect. In contrast, East Asians would be less likely than North Americans to expect such a correspondence in magnitude between effects and their causes. Lastly, we predicted that the cultural differences in the tendency to look for causes that correspond in magnitude with effects would be explained by differences in analytic and holistic reasoning. We

conducted a series of studies to test these hypotheses.

2 Study 1

The purpose of Study 1 was to test the hypothesis that Canadians would expect a greater correspondence in magnitude between effects and their causes than would Chinese. Participants read hypothetical scenarios with consequences of high or low magnitude and indicated how likely the effect was due to causes that were high or low in magnitude. We used two different versions of scenarios with two different samples, one in Study 1a and the other in Study 1b.

2.1 Study 1a

2.1.1 Method

Participants. Fifty-nine European-Canadians (45 women) were recruited from Queen's University, and 60 Chinese (30 women) were recruited from Beijing University. In all of the studies in this article, Canadian participants were Caucasians of European descent and Chinese were Chinese nationals, mostly of Han descent. Canadian participants received course credit or $5 for their participation, and Chinese participants received a small gift.

Materials and procedure. Participants read a questionnaire describing a disease outbreak that either killed some people (high-magnitude effect) or hospitalized them (low-magnitude effect). The effect was followed by two potential causes: a highly infectious strain of bacteria (high-magnitude cause) or a standard strain of bacteria (low-magnitude cause; see the appendix). Participants rated the likelihood that each of the two causes had led to the effect on a 9-point scale (1=*not likely at all*, 9 =*extremely likely*). In summary, Study 1a had a 2 (culture: Canadians vs. Chinese) × 2 (effect magnitude: high vs. low) × 2 (cause magnitude: high vs. low) design. The cause magnitude factor varied within participants, and the other factors varied between participants.

Cause and effect magnitude. To operationalize the magnitude of effects, we selected single events, such as a disease outbreak, and varied the severity of the consequences associated with that event, for example, few or many deaths. We then selected causes that people would intuitively associate with that event, such as a virus. To operationalize the magnitude of these causes, we chose causes such that if all else were equal, a more extreme version, such as a treatment−resistant strain of the virus, would be associated with the more extreme effect to a greater extent than would a less extreme version, such as a standard strain of the virus. This same procedure for generating study materials was followed in all of the studies reported in this article.

The study materials were generated by two Canadian and two Chinese researchers to ensure they were familiar and realistic to both cultures. Pilot testing indicated that Canadians and Chinese perceived the magnitudes and independent likelihoods of effects and causes to be equivalent. This same procedure for generating and pilot testing study materials was followed in all of the studies reported in this article.

Translation. All materials were first developed in English and then translated into Chinese. The Chinese and English versions were then compared by two Chinese researchers who have lived in North America for at least 4 years. Additionally, a back−translation procedure was used to check consistency of meaning, and finally the translations were checked by at least three Chinese researchers in China to ensure they were free of error and that they sounded natural. The same procedure for translating study materials was followed in all of the studies reported in this article.

2.1.2 Results and discussion

Preliminary analyses indicated no significant gender effects, Fs >1, as was true for all studies reported in this article. Thus, gender will not be discussed further.

A 2 (culture) × 2 (effect magnitude) × 2 (cause magnitude) mixed−model ANOVA revealed a significant main effect of effect magnitude, $F(1, 115)=13.55$, $p<.001$, such that participants in the high−magnitude effect condition gave higher likelihood ratings for the causes ($M=5.89$, $SD=0.91$) in comparison with those in the low−magnitude effect condition ($M=5.27$, $SD=0.92$). This pattern indicated a stronger reaction to the high−magnitude effect compared with the low−magnitude effect. The Effect Magnitude × Cause Magnitude interaction was significant, $F(1, 115)=160.18$, $p<0.001$, $\eta_p^2=0.58$. Over−
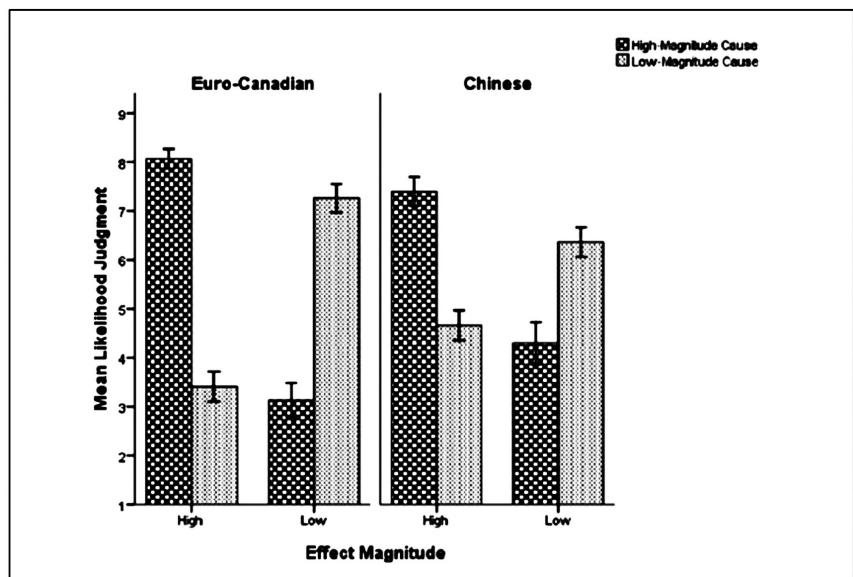
Figure. 1. Canadian and Chinese likelihood estimates (+SE) of high− and low−magnitude causes leading to high− and low−magnitude effects (Study 1a)

all, participants tended to associate the high-magnitude effect with the high-magnitude cause (*M*=7.73, *SD*=1.40) more than with the low-magnitude cause (*M*=4.05, *SD*=1.77), and the low-magnitude effect with the low-magnitude cause (*M*=6.82, *SD*=1.67) more than with the high-magnitude cause (*M*=3.72, *SD*=2.21). More importantly, the Culture × Effect Magnitude × Cause Magnitude interaction was significant, $F(1, 115) =13.80$, $p<.001$, $\eta_p^2=0.11$. As hypothesized, Canadians tended to exhibit a stronger cause–effect magnitude correspondence than Chinese (see Figure 1).

Follow-up, independent-sample *t* tests on the interaction indicated that to account for the high-magnitude event, Canadians rated the high-magnitude cause as more likely than did Chinese and rated the low-magnitude cause as less likely than did Chinese. For the low-magnitude event, Canadians rated the high-magnitude cause as less likely than did Chinese and rated the low-magnitude cause as more likely than did Chinese, *t*s>2.10, *p*s<0.04.

2.2 Study 1b

2.2.1 Method

Participants. Eighty-four European Canadians (57 women) were recruited from Queen's University, and 60 Chinese (33 women) were recruited from Beijing University. Canadian participants received course credit or $5 for their participation, and Chinese participants received a small gift.

Materials and procedure. Study 1b followed a similar design and procedure to Study 1a but had a different scenario. The scenario described either a long negotiation delay (high-magnitude effect) or a brief negotiation delay (low-magnitude effect). The causes were a major disagreement (high-magnitude cause) or a minor one (low-magnitude cause; see the appendix). In summary, Study 1b followed a 2(culture: Canadians vs. Chinese) × 2 (effect magnitude: high vs. low) × 2 (cause magnitude: high vs. low) design, with the cause magnitude varying within participants and the other factors varying between participants.

2.2.2 Results and discussion

The results were similar to those obtained in Study 1a. A 2 (culture) × 2 (effect magnitude) × 2 (cause magnitude) mixed-model ANOVA revealed a significant main effect of effect magnitude, $F(1,140)=4.12, p=0.04$, such that the likelihood ratings for causes were higher in the
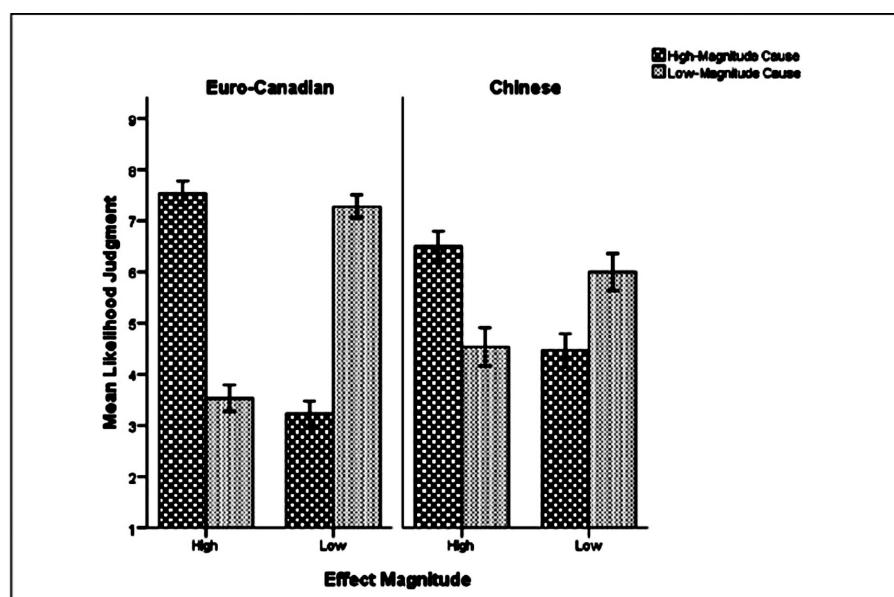
Figure. 2. Canadian and Chinese likelihood estimates (+SE) of high- and low-magnitude causes leading to high- and low-magnitude effects (Study 1b)

high-magnitude effect condition(M=5.53, SD=0.81)than in the low-magnitude effect condition (M=5.25, SD=0.82). The Effect Magnitude × Cause Magnitude interaction was significant, F(1, 140)=133.36, p<0.001, η_p^2=0.49. Overall, participants tended to associate high-magnitude effects with high-magnitude causes (M=7.10, SD=1.63) more than with low-magnitude causes (M=3.96, SD=1.87), and low-magnitude effects with low-magnitude causes (M=6.75, SD=1.81) more than with high-magnitude causes (M=3.74, SD=1.77). More importantly, the Culture × Effect Magnitude × Cause Magnitude interaction was significant, F(1, 140) =20.68, p<.001, η_p^2=0.13. As hypothesized, Canadians tended to exhibit a stronger cause–effect magnitude correspondence than Chinese (see Figure 2).

Follow-up, independent-sample t tests on the interaction indicated that to account for the high-magnitude event, Canadians rated the high-magnitude cause as more likely than did Chinese and rated the low-magnitude cause as less likely than did Chinese. For the low-magnitude event, Canadians rated the high-magnitude cause as less likely than did Chinese and rated the low-magnitude cause as more likely than did Chinese, ts > 2.28, ps < .03

Thus, for both the disease (Study 1a) and the negotiation (Study 1b) scenarios, Canadians associated high-magnitude effects with high-magnitude causes more than with low-magnitude causes and low-magnitude effects with low-magnitude causes more than with high-magnitude causes. And in both scenarios, Chinese exhibited this pattern to a significantly lesser degree.

3 Study 2

The purpose of Study 2 was to replicate the results from Study 1 using a different format. Study 1 manipulated the magnitudes of causes and effects in detailed scenarios.

In Study 2, we used simple scenarios that were described using pictures of common effects. Again, participants read hypothetical scenarios with consequences of high or low magnitude and indicated how likely the effect was due to causes that were high or low in magnitude.

3.1 Method

Participants. Seventy-eight European Canadians (46women) were recruited from Queen's University, and 60 Chinese (27 women) were recruited from Beijing University. Canadian participants received course credit or $5 for their participation, and Chinese participants received a small gift.

Materials and procedure. Participants were randomly assigned to a high- or low-effect magnitude condition. In each condition, they were presented first with a picture of two basketball players, one tall (high-magnitude cause) and one short (low-magnitude cause), and indicated the likelihood that each of the two players had scored the most points (high-magnitude effect) or the least points (low-magnitude effect) in a game. Next, participants were presented with a picture of two tornadoes that had traveled through a city, one wide (high-magnitude cause) and one narrow (low-magnitude cause), and indicated the likelihood that each tornado had caused extensive damage (high-magnitude effect) or no damage (low-magnitude effect). The likelihood judgments were made on a 9-point scale (1=*not likely at all*, 9=*extremely likely*). In summary, Study 2 had a 2 (culture: Canadians vs. Chinese) × 2 (effect magnitude: high vs. low) × 2 (cause magnitude: high vs. low) design. The cause magnitude factor varied within participants, and the other factors varied between participants.

3.2 Results and discussion

Likelihood estimate computation. Each participant completed both scenarios (basketball game and tornadoes). For each scenario, participants gave two likelihood estimates, one for a high-magnitude cause and one for a

low-magnitude cause. The pattern of the results was the same for each of the two scenarios. Therefore, we combined the scenarios by averaging the two likelihood estimates for the high-magnitude causes and by averaging the two likelihood estimates for the low-magnitude causes. These two averages were then treated as repeated measures variables.

Test of cause-effect magnitude correspondence. A 2(culture) × 2 (effect magnitude) × 2 (cause magnitude) mixed-model ANOVA revealed a significant main effect of effect magnitude, $F(1, 134)=23.72$, $p<.001$, such that participants in the high-magnitude effect condition gave higher likelihood ratings for the causes ($M=5.72$, $SD=0.67$) in comparison with those in the low-magnitude effect condition ($M=5.22$, $SD=0.45$). The Effect Magnitude × Cause Magnitude interaction was significant, $F(1, 134)=50.20$, $p<.001$, $\eta_p^2=0.27$. Over-all, participants tended to associate high-magnitude effects with high-magnitude causes ($M=6.80$, $SD=1.28$) more than with low-magnitude causes ($M=4.63$, $SD=1.30$), and low-magnitude effects with low-magnitude causes ($M=5.91$, $SD=1.71$) more than high-magnitude causes ($M=4.54$, $SD=1.73$). Replicating the results from Study 1, the Culture × Effect Magnitude × Cause Magnitude interaction was significant, $F(1, 134)=7.48$, $p=0.007$, $\eta_p^2=0.05$. As hypothesized, Canadians exhibited a stronger cause-effect magnitude correspondence than Chinese (see Figure 3).

Follow-up, independent-sample t tests on the interaction indicated that to account for the high-magnitude event, Canadians rated the high-magnitude cause as more likely than did Chinese and rated the low-magnitude cause as less likely than did Chinese. For the low-magnitude event, Canadians rated the high-magnitude cause as less likely than did Chinese and rated the low-magnitude cause as more likely than did Chinese, ts>1.96, ps<.05.

Thus, in two detailed scenarios represented by words

in Study 1 and two simple scenarios represented by pictures in Study 2, Canadians associated high-magnitude effects with high-magnitude causes more than with low-magnitude causes and low-magnitude effects with low-magnitude causes more than with high-magnitude causes. Furthermore, for all four scenarios, Chinese exhibited this pattern to a significantly lesser degree.

4 Study 3

Studies 1 and 2 found that compared with Chinese, Canadians expect effects and their causes to correspond in magnitude to a greater degree. We argue that the pattern of results is caused by cultural differences in holistic reasoning. However, a simple alternative explanation exists: The results from Studies 1 and 2 could be explained by a stronger preference to choose midpoints on scales by Chinese than by Canadians. In both studies, participants rated their likelihood judgments on a Likert-type scale. Chen, Lee, and Stevenson (1995) found evidence that East Asians typically prefer points closer to the midpoints of such scales, even though the degree of midpoint-response bias in their study was weak, especially comparing Chinese and Canadians. Furthermore, other studies have found no such tendency (e.g., Ji, Schwarz, & Nisbett, 2000).

Nonetheless, we designed Study 3 to rule out the possibility that Chinese were engaging in such a moderacy-response bias. Instead of rating the likelihoods of causes on a scale, participants chose the cause they perceived to have most likely led to the effect. If the same pattern of results as in Studies 1 and 2 emerged regarding participants'choices, then these results would provide strong evidence against the alter-native explanation.

4.1 Method

Participants. Sixty-three European Canadians (48 women) and 63 Chinese nationals living in Canada (44 women) were recruited from Queen's University.

Participants received course credit or \$5 for their participation. At the time of the study, the Chinese nationals had lived in Canada for an average of 28.83 months (*SD*=13.54).

Materials and procedure. Participants were randomly assigned to either the high or the low effect magnitude condition. Within each condition, participants read three scenarios.One scenario was taken from Study 1 (disease outbreak described in words) and two scenarios were taken from Study 2 (basketball game and tornadoes depicted in pictures). In the high−magnitude effect condition, all three scenarios described high−magnitude effects and likewise in the low−magnitude effect condition, all three scenarios described low−magnitude effects. For each scenario, participants chose the more likely cause of the effect between the high− and the low−magnitude causes. Additionally, participants indicated their confidence in each choice on an 8−point scale (1=*not at all confident,8 =extremely confident*). In summary, Study 3 had a 2 (culture: Canadians vs. Chinese) × 2 (effect magnitude: high vs. low) × 2 (cause magnitude: high vs. low) design. The cause magnitude factor varied within participants, and the other factors varied between participants.

4.2 Results and discussion

Test of cause−effect magnitude correspondence. Each participant chose three causes, one for each scenario. Each choice was analyzed separately by conducting a 2 (culture) × 2 (effect magnitude) × 2 (cause magnitude) log−linear analysis. For all three scenarios, the analyses revealed significant interactions between effect magnitude and cause magnitude, all G^2s >25.72, all *df*=1, all *p*s<.001, indicating that for each scenario, participants were more likely to associate high−magnitude effects with high−magnitude causes than with low−magnitude causes, and low−magnitude effects with low−magnitude causes than with high−magnitude causes. More importantly, for all three scenarios, the analyses revealed significant Culture × Effect Magnitude × Cause Magnitude interactions, all G^2s >13.58, all *df*=1, all *p*s<.001. For all three scenarios, Canadians were more likely to exhibit the magnitude−matching pattern than were Chinese (see Table 1 for frequency counts). Thus, we successfully replicated the results from Studies 1 and 2 using a method that did not rely on a Likert−type scale, suggesting that a midpoint preference by Chinese was unlikely to account for the results obtained in Studies 1 and 2.
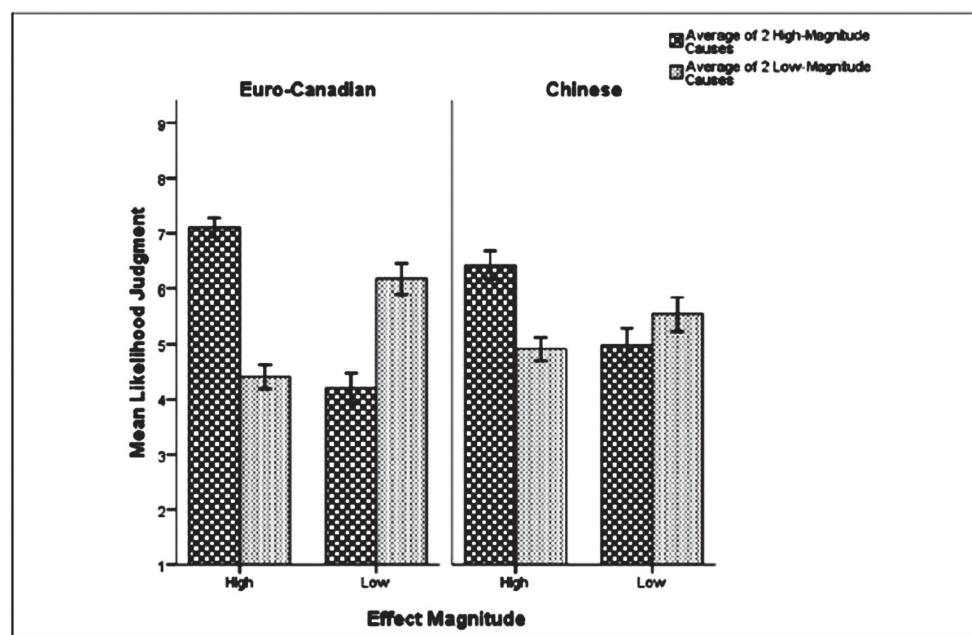

Figure 3. Canadian and Chinese likelihood estimates (+ SE) of high−and low−magnitude causes leading to high− and low−magnitude effects (Study 2)

Confidence ratings. A 2 (culture) × 2 (effect magnitude) × 3 (scenario) mixed-model ANOVA indicated no significant interactions, all ps >.79. The effect magnitude main effect was significant, $F(1, 122) = 5.94$, $p = 0.02$, $\eta_p^2 = 0.05$, indicating that participants were more confident in their choices in the low-magnitude condition ($M = 6.25$, $SD = 0.99$) than in the high-magnitude condition ($M = 5.80$, $SD = 1.15$). In addition, the culture main effect was significant, $F(1, 122) = 11.20$, $p < .001$, $\eta_p^2 = 0.08$, such that Chinese were more confident of their choices overall ($M = 6.33$, $SD = 0.82$) than were Canadians ($M = 5.71$, $SD = 1.24$). Therefore, on this scale, Chinese were not more likely than Canadians to prefer responses near the midpoint of the scale, providing further evidence that a midpoint preference by Chinese was unlikely to account for the results obtained in Studies 1 and 2.

In summary, Study 3 replicated the results of Studies 1 and 2 using a different method and ruled out the potential alternative explanation that Chinese were engaging in a moderacy-response bias. In fact, Chinese responded significantly further from the midpoint than did Canadians on confidence ratings of their likelihood judgments.

5 Study 4

The purpose of Study 4 was to directly test the hypothesis that analytic and holistic reasoning was responsible for the cultural differences in the tendency to expect a correspondence in magnitude between cause and effect. We developed an exercise to prime either analytic or holistic reasoning in Canadian participants. Participants completed the exercise, then read a scenario describing a high- or low-magnitude effect, and rated the likelihood that high- or low-magnitude causes had led to that effect. We used two different versions of scenarios with two separate samples, one in Study 4a and the other in Study 4b.

5.1 Study 4a

5.1.1 Method

Participants. Sixty-seven European Canadians (49 women) were recruited from Queen's University. Participants received course credit or $5 for their participation.

Analytic versus holistic prime. We primed analytic and holistic thinking by focusing participants' attention on either a simple or a complex causal field, respectively. For the prime, participants completed an exercise ostensibly unrelated to the rest of the study. All participants read:

Getting into a competitive university such as Queen's University is a major achievement. The majority of high school students do not make it into any university at all, and a large number of applicants to Queen's are turned away every year.

In the analytic prime condition, participants then listed the most significant event in their life that had enabled them to get into Queen's and described how it had done so. Last, they completed a diagram consisting of two ellipses, one labeled *Event* and the other *Getting into Queen's,* by writing the significant event in the

Table 1. Frequencies of Canadian and Chinese Cause Magnitude Choices (and Percentages Within Each Culture): Study 3

Effect	Culture	Disease outbreak causes		Basketball game causes		Tornado damage causes	
		High	Low	High	Low	High	Low
High	Canadian	29	3	30	2	28	4
		(90.6%)	(9.4%)	(9.38%)	(6.2%)	(87.5%)	(12.5%)
	Chinese	18	14	20	12	19	13
		(56.2%)	(43.8%)	(62.5%)	(37.5%)	(59.4%)	(40.6%)
Low	Canadian	4	27	3	28	4	27
		(12.9%)	(87.1%)	(9.7%)	(90.3%)	(12.9%)	(87.1%)
	Chinese	14	17	13	18	13	18
		(45.2%)	(54.8%)	(41.9%)	(58.1%)	(41.9%)	(58.1%)

event ellipse and by drawing an arrow between it and the getting into Queen's ellipse. The exercise was designed to focus each participant's attention on a single cause that had led to a major event in his or her life. The holistic prime was nearly identical to the analytic prime except that participants listed the three most significant events. Holistic reasoning involves not only focusing on numerous causes but also focusing on the interactions between such causes. Therefore, participants also described how the three events had influenced each other.The diagram consisted of four ellipses, three on the periphery labeled *Event* and one in the center labeled *Getting into Queen's*. After writing the three events in the event ellipses, participants drew arrows from each one to the Queen's ellipse. Lastly, they drew arrows connecting the three events to describe how these events had influenced or interacted with each other. The holistic prime was designed to focus participants' attention on a larger causal field and on the connectedness of causes within that field.

Magnitude manipulations. After the prime, participants read the disease scenario from Study 1 and rated the likelihood that each of the two causes had led to the effect on a 9-point scale (1 = *not likely at all*,9 = *extremely likely*). In summary, Study 4a had a 2 (prime:

analytic vs. holistic) × 2 (effect magnitude: high vs. low) × 2 (cause magnitude: high vs. low) design. The causal magnitude factor varied within participants, and the other factors varied between participants.

5.1.2 Results and discussion

Test of cause‑effect magnitude correspondence. A 2 (prime: analytic vs. holistic) × 2 (effect magnitude) × 2 (cause magnitude) mixed-model ANOVA revealed a significant main effect of effect magnitude, $F(1, 63)=12.47$, $p<.001$, such that participants in the high-magnitude effect condition gave higher likelihood ratings for the causes ($M=5.97$, $SD=0.94$) in comparison with those in the low-magnitude effect condition ($M=5.35$, $SD=0.92$). The Effect Magnitude × Cause Magnitude interaction was significant, $F(1, 63)=78.92$, $p<.001$, $\eta_p^2=0.56$. Overall, participants tended to associate high-magnitude effects with high-magnitude causes ($M=7.17$, $SD=1.56$) more than with low-magnitude causes ($M=4.51$, $SD=1.98$), and low-magnitude effects with low-magnitude causes ($M=7.00$, $SD=1.30$) more than with high-magnitude causes ($M=4.44$, $SD=1.76$). More importantly, the Prime × Effect Magnitude × Cause Magnitude interaction was significant, $F(1, 63)=12.02$, $p<.001$, $\eta_p^2=0.16$. As hypothesized, Canadians primed to reason analytically exhibited a
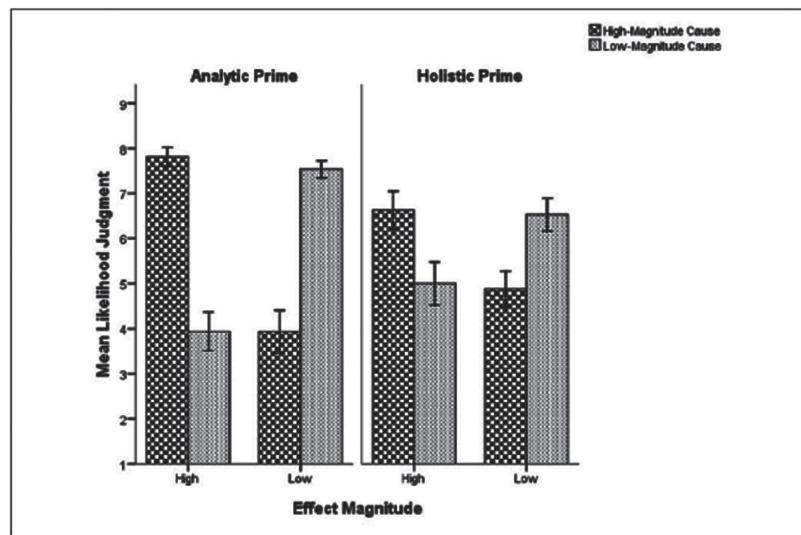
Figure 4. Analytically primed and holistically primed Canadian likelihood estimates (+SE) of high- and low-magnitude causes leading to high- and low-magnitude effects (Study 4a)

stronger cause–effect magnitude correspondence than Canadians primed to reason holistically (see Figure 4).

Follow–up, independent–sample t tests on the interaction indicated that to account for the high–magnitude event, analytically primed Canadians rated the high–magnitude cause as more likely than did holistically primed Canadians and rated the low–magnitude cause as less likely than did holistically primed Canadians. For the low–magnitude event, analytically primed Canadians rated the high–magnitude cause as less likely than did holistically primed Canadians and rated the low–magnitude cause as more likely than did holistically primed Canadians, ts> 1.96, ps<.05.

Comparison with nonprimed participants from Study 1a. Participants in Study 1a completed the identical disease scenario but without any prime. Therefore, we compared participants from Study 4a with those from Study 1a to determine more specifically what effect the analytic and holistic primes had on participants. First, we compared analytically primed Canadians with those who received no prime. The Prime × Effect Magnitude × Cause Magnitude interaction was not significant, $F(1, 86)$ =1.36, p=0.25, revealing that Canadians primed to reason analytically did not differ from those who received no prime. Next, we compared holistically primed Canadians with those who received no prime. The Prime × Effect Magnitude × Cause Magnitude interaction was significant, $F(1, 91)$ =21.39, p<.001, η_p^2=0.19, revealing that Canadians primed to reason holistically expected less cause–effect magnitude correspondence than did Canadians who received no prime. Lastly, we compared holistically primed Canadians with Chinese from Study 1a who received no prime. The Culture × Effect Magnitude × Cause Magnitude interaction was not significant, $F(1, 91)$ =1.47, p=0.23, revealing that Canadians primed to reason holistically did not differ from Chinese who received no prime.

5.2 Study 4b

5.2.1 Method

Participants. One hundred twenty–one European Canadians (75 women) were recruited from Queen's University. Participants received course credit or $5 for their participation.

Materials and procedure. The procedure was nearly identical to Study 4a except that participants read a money scenario after the prime. We also included a control condition in which participants did not receive any prime. The money scenario described a Canadian individual who had either accumulated greater than average savings (high–magnitude effect) or lesser than average savings (low–magnitude effect). The scenario was followed by two potential causes, one of high magnitude (the individual had a higher than average income) and one of low magnitude (the individual had a lower than average income). In summary, Study 4b had a 3 (prime: none, analytic, or holistic) × 2 (effect magnitude: high vs. low) × 2 (cause magnitude: high vs. low) design. The causal magnitude factor varied within participants, and the other factors varied between participants.

5.2.2 Results and discussion

A 3 (prime) × 2 (effect magnitude) × 2 (cause magnitude) mixed–model ANOVA revealed a significant main effect of effect magnitude, $F(1, 115)$= 8.72, p=0.004, such that participants in the high–magnitude effect condition gave higher likelihood ratings for the causes (M=5.52, SD=0.97) in comparison with those in the low–magnitude effect condition (M=4.77, SD=0.98). The Effect Magnitude × Cause Magnitude interaction was significant, $F(1, 115)$=239.66, p<.001, η_p^2= .68. Overall, participants tended to associate high–magnitude effects with high–magnitude causes (M=6.97, SD=1.53) more than with low–magnitude causes (M=3.43, SD=2.01), and low–magnitude effects with low–magnitude causes (M=6.57, SD=1.51) more than with high–magnitude causes

(M=3.07, SD=1.64). More importantly, the Prime × Effect Magnitude × Cause Magnitude interaction was significant, $F(2, 115)$=6.81, p=0.002, η_p^2=0.11. We conducted follow-up 2 (prime) × 2 (effect magnitude) × 2 (cause magnitude) mixed-model ANOVA to determine more specifically the effects of the primes. As hypothesized, Canadians primed to reason analytically exhibited a stronger cause－effect magnitude correspondence than Canadians primed to reason holistically, $F(1, 61)$=10.16, p=0.002, η_p^2=0.14. Canadians primed to reason analytically did not differ in the tendency to make this association compared with Canadians who received no prime, $F(1, 87)$ = .22, p=0.64. Lastly, Canadians primed to reason holistically tended to exhibit a weaker cause－effect magnitude correspondence than did Canadians who received no prime, $F(1, 82)$= 10.26, p=0.002, η_p^2=0.11 (see Figure 5).

Analytically primed Canadians did not differ from those receiving no prime, and thus we collapsed the two groups together, referred to them as analytic-Canadians, and compared them with holistically primed Canadians. Follow-up, independent-sample t tests on the interaction indicated that to account for the high-magnitude event, analytic-Canadians rated the high-magnitude cause as more likely than did holistically primed Canadians and rated the low-magnitude cause as less likely than did holistically primed Canadians, ts > 2.10, ps < .04. For the low-magnitude event, analytic-Canadians rated the high-magnitude cause as less likely (M=2.78, SD=1.16) than did holistically primed Canadians (M=3.50, SD=1.28), t=1.96, p<.05, and rated the low-magnitude cause as marginally more likely than did holistically primed Canadians, t=1.91, p=0.07.

In summary, for both the disease (Study 4a) and the negotiation (Study 4b) scenarios, analytically primed Canadians associated high-magnitude effects with high-magnitude causes more than with low-magnitude causes,

and low-magnitude effects with low-magnitude causes more than with high-magnitude causes. And in both scenarios, holistically primed Canadians exhibited this pattern to a significantly lesser degree. Additionally, comparing participants from Study 4a with participants from Study 1a who completed the same materials without any prime revealed that the analytic prime had no effect. Instead, the holistic prime was the one that caused Canadians to expect less cause－effect magnitude correspondence. This pattern of results was replicated in Study 4b by comparing the participants who received either the analytic prime or the holistic prime with participants in a control condition who received no prime. Lastly, comparing holistically primed Canadians in Study 4a with Chinese in Study 1a who received no prime revealed no cultural differences in the tendency to expect a cause－effect magnitude correspondence.

6 General discussion

Across three studies, Canadians expected a greater correspondence in magnitude between effects and their causes than did Chinese. In a fourth study, both analytically primed Canadians and Canadians who received no prime were more likely to expect the correspondence in magnitude between cause and effect than were holistically primed Canadians, whereas the holistically primed Canadians showed similar responses to Chinese participants. The results not only demonstrated cultural differences in the extent to which people expect a correspondence in magnitude between cause and effect, but they also demonstrated the underlying factor responsible for such cultural differences, namely, the causal complexity factor of analytic or holistic reasoning.

6.1 Alternative explanations

One potential alternative explanation for the results is that Chinese may be less familiar with the scenarios than

are Canadians. If people are reasoning about the cause of an effect in a context where causes and effects tend to correspond in magnitude in the real world, then the tendency to predict such an association would be prudent. Therefore, if Canadians are more familiar than Chinese are with a scenario, and thus the underlying causes of the effect in that scenario, then Canadians should expect a greater degree of correspondence between cause and effect.

This potential alternative explanation has at least two problems. First, when choosing and designing our scenarios, we were very careful to generate scenarios that were familiar and understandable to both cultural groups. Indeed, a number of scenarios that were initially proposed by either the Canadian or Chinese researchers were rejected because they did not meet these criteria. Second, even if a cause‐effect magnitude correspondence existed for all of the scenarios we used, and all of our efforts at ensuring equivalent levels of familiarity with the scenarios failed, the familiarity explanation would not fit with the pattern of results in Studies 4a and 4b. It would be difficult to explain, based on familiarity with the scenarios, why priming Canadians to think holistically would cause them to respond similarly to Chinese, namely, by expecting a

lesser degree of cause‐effect magnitude correspondence.

6.2 Theoretical contributions

Culture and attributions. Gilovich and Savitsky (2002, p. 618) defined the representativeness heuristic as the tendency to process information "on the basis of one overarching rule: 'Like goes with like.'" Most of the research on this heuristic has been conducted in the context of categorization, whereas little research has explored the representativeness heuristic in the context of attributions. Although some have speculated on the tendency to expect cause‐effect magnitude correspondence when making attributions (Einhorn & Hogarth, 1986; Nisbett & Ross, 1980), little research has systematically investigated this speculation. Our studies contribute to the attribution literature by providing strong empirical evidence in support of this proposal.

In addition, we demonstrated that the degree to which people expect cause‐effect magnitude correspondence differs across cultures. Throughout our studies, both cultures associated causes and effects on the basis of magnitude similarity, and this pattern of results indicates that both Canadians and Chinese exhibited judgments consistent with the representativeness heuristic. However, the fact that Canadians expected a greater degree of
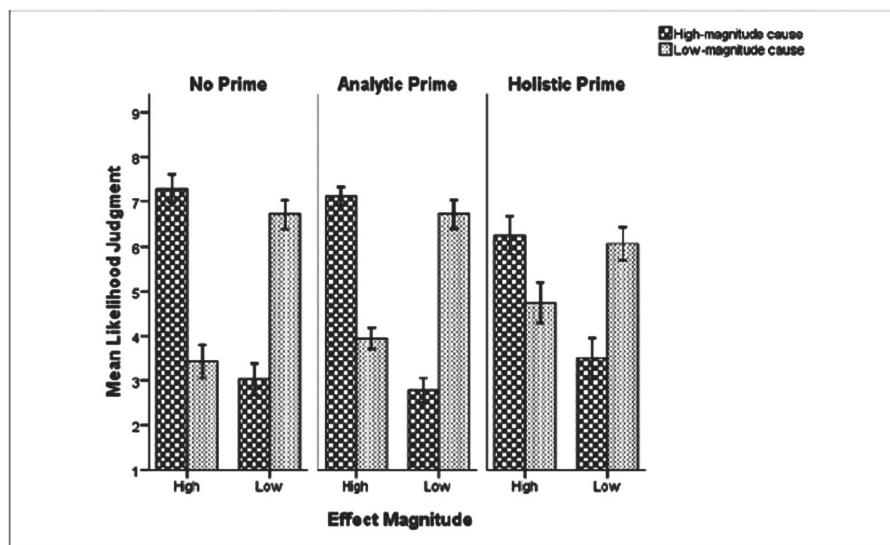
Figure 5. No prime, analytically primed, and holistically primed Canadian likelihood estimates (+ SE) of high‐ and low‐magnitude causes leading to high‐ and low‐magnitude effects (Study 4b)

correspondence in magnitude when making causal judgments indicates that they exhibit a more extreme form of the representativeness heuristic in this context, compared with Chinese. Our study is the first we are aware of to demonstrate cultural differences in the degree to which people employ the representativeness heuristic in the context of attributions.

In addition, in examining the literature on the representativeness heuristic, we are aware of no studies that have demonstrated any causal mechanisms. Our finding that the degree to which people expect cause–effect magnitude correspondence is in part determined by whether they reason analytically or holistically enhances our understanding of the cognitive underpinnings of the representativeness heuristic in the context of attributions. Furthermore, the fact that this tendency differs across cultures signals the need to investigate whether cultural differences would emerge for other heuristics.

An important caveat here is that we demonstrated cultural differences in the representativeness heuristic in the context of causal judgments based on a similarity in magnitude. We have provided no evidence regarding whether the same pattern of results would emerge with respect to the representativeness heuristic in other contexts, such as attributing causes based on their physical resemblance to effects.

Cultural universals. Notably, within the domain of our scenarios and designs, participants from both cultures expected a cause–effect magnitude correspondence. What differed across cultures was the degree to which people expected this correspondence. Norenzayan and Heine (2005) outlined a taxonomy of cultural universals, defined as core mental attributes shared by people everywhere. The taxonomy consists of four levels of cultural universality: accessibility, functional, existential, and nonuniversality. Accessibility universals, the most stringent level of universality, are psychological processes that are available

to all people, used for the same function, and accessed to the same degree. Functional universals, the second most stringent level of universality, are cognitively available to all people, used for the same function, but accessed to different degrees. Existential universals, the third most stringent level of universality, are cognitively available to all people, but they may be used in markedly different ways and are accessed to different degrees. Nonuniversals are those processes that are not cognitively available to all people.

Although we only sampled people from two cultures, our pattern of results demonstrating that participants from both cultures expected cause–effect magnitude correspondence, but to different degrees, would potentially qualify as a functional universal. Our studies provide the first evidence we are aware of regarding the universality and cultural variability of the representativeness heuristic. What appears on the surface to be a simple heuristic, the tendency to associate high–magnitude effects with high–magnitude causes and low–magnitude effects with low–magnitude causes, is applied to different degrees depending on how our host culture has shaped our minds to process information.

6.3 Practical implications

The pattern of results could also have important practical implications. One potential application is in the domain of behavioral decision making. According to the U.S. Federal Reserve, per capita personal debt in the United States has increased by at least a factor of 10 between 1945 and 2005, reaching unprecedented levels (Massey, 2008). Canadian debt levels are following a similar trend according to the Bank of Canada. The causes of this trend are surely complex and we do not intend to oversimplify them or claim that our research has solved this problem. However, our results could potentially contribute to better understanding and ameliorating the situation. For example, North Americans may be more likely to believe

that to accumulate wealth or to resolve their debt problems, they need to focus on cutting back on major costs and purchases. In doing so, they may pay little attention to the financial impact of minor routine expenses. Consistent with this possibility, financial experts are advising people to cut back on minor expenses, such as the purchase of a daily cup of coffee at a trendy café, because such purchases add up over time. Thus, for people who find themselves in debt, realizing the importance of reducing minor routine expenses could enable them to more quickly and effectively ameliorate their financial hardship. In addition, the results from Study 4, demonstrating that the tendency to expect a magnitude correspondence is reduced when primed to think holistically, provides hope that people can reduce the degree to which they overlook minor expenses in such situations. We are investigating some of these implications in our ongoing research.

Meanwhile, this research could have mirror–image implications for East Asians. For example, East Asians may underestimate the importance of major health–related behaviors because they believe that other minor health–related behaviors will compensate. For example, when told by a medical expert that ceasing a pernicious behavior, such as smoking, is essential for dealing with a major illness, such as lung cancer, Easterners may be more likely to believe that their other health–related behaviors, such as eating well, will compensate for the detrimental effects of smoking.

The present research also has practical implications for problem solving and negotiation. The human world is becoming increasingly interconnected and many nations are becoming more ethnically diverse. People from different cultures are realizing the growing need to work together to solve problems, such as those currently undermining political and financial stability, as well as those plaguing our natural environment. When investigating the cause of a high–magnitude effect, Westerners' tendency

to expect high–magnitude causes may lead them to overlook relatively lower magnitude causes that might have played a key role. On the other hand, Chinese might over– emphasize factors that played a minor role at the expense of those that played a major role. These different tendencies to emphasize high– versus low–magnitude causes could lead to disagreement or international conflict. In a world in which cross–cultural interaction is crucial, further understanding cultural differences in reasoning about cause and effect relationships could prove important both for finding solutions to the problems and for negotiating diplomatic resolutions to the inevitable disagreements that will arise between nations.

People associate big causes with big effects and small causes with small effects. However, the degree to which we make this association is at least partially determined by the reasoning processes we are imbued with from the culture in which our minds developed. Some physicists believe that our universe, created by the Big Bang, will end in a big crunch. Perhaps a more holistic interpretation of this scenario would result in the universe that began from the Big Bang ending in a small crunch.

Appendix

Scenarios Used in Studies 1a, 3, and 4a

Effects

● 21 people at a major downtown company became ill. They were stricken with symptoms of nausea and vomiting. Within 3 days, 11 of these individuals had experienced a rapid but horrific death and the other 10 were still in hospital. (*High magnitude*)

● 7 people at a major downtown company became ill. They were stricken with symptoms of nausea and vomiting. Within 3 days, 3 of these individuals had recovered and the other 4 were still experiencing

minor symptoms. (*Low magnitude*)

Causes

● An employee came into contact with a highly infectious type of super−bacteria while on a business trip. (*High magnitude*)

● An employee came into contact with a standard type of bacteria while on a business trip. (*Low magnitude*)

Scenarios Used in Study 1b

Effects

● Two parties are negotiating a new contract. This negotiation took over 31 weeks to reach an agreement, which is much longer than usual compared with other similar negotiations. (*High magnitude*)

● Two parties are negotiating a new contract. This negotiation took only 2 weeks to reach an agreement, which is typical compared with other similar negotiations. (*Low magnitude*)

Causes

● Two negotiating parties could not agree on one major point, which amounted to 41% of the total contract value. (*High magnitude*)

● Two negotiating parties could not agree on one minor point, which amounted to 10% of the total contract value. (*Low magnitude*)

References

Chen, C., Lee, S. Y., & Stevenson, H. W. (1995). Response style and cross−cultural comparisons of rating scales among East Asian and North American students, *Psychological Science, 6*, 170–175.

Choi, I., Dalal, R., Kim−Prieto, C., & Park, H. (2003). Culture and judgment of causal relevance. *Journal of Personality and Social Psychology, 84*, 46–59.

Ebel−Lam, A. P., Fabrigar, L. R., MacDonald, T. K., & Jones, S. (2008). *Balancing causes and consequences:The proportionality principle in explanations for complex social events*. Unpublished manuscript, Queen's University, Kingston, Ontario, Canada.

Einhorn, H. J., & Hogarth, R. M. (1986). Judging probable cause. *Psychological Bulletin, 99*, 3–19.

Gilovich, T., & Savitsky, K. (2002). Like goes with like: The role of representativeness in erroneous and pseudo−scientific beliefs. In T. Gilovich, D. Griffin, & D. Kahnemann (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 617–624). Cambridge, UK: Cambridge University Press.

Ji, L. J., Nisbett, R. E., & Su, Y. (2001). Culture, change, and prediction. *Psychological Science, 12*, 450–456.

Ji, L. J., Peng, K., & Nisbett, R. E. (2000). Culture, control, and perception of relationships in the environment. *Journal of Personality and Social Psychology, 78*, 943–955.

Ji, L. J., Schwarz, N., & Nisbett, R. E. (2000). Culture, autobiographical memory, and behavioral frequency reports: Measurement issues in cross−cultural studies. *Personality and Social Psychology Bulletin, 26*, 585–593.

Kahneman,D.,&Tversky,A.(1972).Subjective probability:A judgment of representativeness. *Cognitive Psychology, 3*, 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review, 80*, 237–251.

Lorenz, E. N. (1963). Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences, 20*, 130–141.

Massey, D. S. (2008). Globalization and inequality: Explaining American exceptionalism. *European Sociological Review*, 1–15.

Masuda, T., & Nisbett, R. E. (2001). Attending holistically versus analytically: Comparing the context sensitivity of Japanese and Americans. *Journal of Personality and Social Psychology, 81*, 922–934.

McCauley, C., & Jacques, S. (1979). The popularity of conspir–acy theories of presidential assassination: A Bayesian analysis. Journal of Personality and Social Psychology, 37, 637–644.

McClure, J., Lalljee, M., & Jaspars, J. (1991). Explanations of extreme and moderate events. *Journal of Research in Personality, 25*, 146–166.

Miller, J. G. (1984). Culture and the development of everyday social explanation. *Journal of Personality and Social Psychology, 46*, 961–978.

Morris, M. W., & Peng, K. (1994). Culture and cause: American and Chinese attributions for social and physical events. *Journal of Personality and Social Psychology, 67*, 949–971.

Nisbett, R. E., Peng, K., Choi, I., & Norenzayan, A. (2001). Culture and systems of thought: Holistic versus analytic cognition. *Psychological Review, 108*, 291–310.

Nisbett, R. E., & Ross, L. (1980). *Human inference: Strategies and shortcomings of social judgment.* Upper Saddle River, NJ: Prentice Hall.

Norenzayan, A., & Heine, S. J. (2005). Psychological universals: What are they and how can we know? *Psychological Bulletin, 131*, 763–784.

Osgood, C. E. (1957). *Measurement of meaning.* Chicago: University of Illinois Press.

Peng, K., & Nisbett, R. A. (1999). Culture, dialectics, and reasoning about contradiction, *American Psychologist, 54*, 741–754.

Shultz, T. R., & Ravinsky, F. B. (1977). Similarity as a principle of causal inference. *Child Development, 48*, 1552–1558.

Goal contents and goal contexts: Experiments with chinese students

Ze Wang [1], Xiaoyong Hu [2], Yongyu Guo [2]

([1] University of Missouri)

([2] Central China Normal University, China)

Abstract　Using samples of Chinese middle school students, the 2 experimental studies presented here examined the effects of goal content and goal context on test performance, free-choice engagement, and test anxiety within the framework of self-determination theory. Students' learning goals were induced as intrinsic or extrinsic with the learning contexts of either autonomy-supportive or controlling. Results suggested that as the more recent extensions of self-determination theory, goal content and goal context effects existed among our samples of Chinese middle school students. However, there was some inconsistency between the authors' findings and previous findings in Western culture.

1 Introduction

At educational settings students engage in learning activities for various reasons. Self-determination theory (SDT) examines different orientations of engaging in a task (i.e., what type of motivation; Deci & Ryan, 1985; Ryan & Deci, 2000). SDT researchers focus on how social and cultural factors facilitate or undermine people's sense of volition and initiative. Basic psychological needs such as autonomy, competence, and relatedness are considered as requirements for the healthy development and functioning of human beings. The satisfaction of these needs serves as the foundation in understanding what social and cultural factors affect human motivation and engagement in activities (Deci & Ryan, 2008). Early SDT research differentiated between intrinsic motivation and extrinsic motivation; later research refined extrinsic motivation to reflect different degrees of self-regulation and internalization of an action, and the key differentiation within SDT shifted to a focus on autonomous-versus-controlled motivation (Deci & Ryan, 2008). More recently, goal contents (intrinsic vs. extrinsic), goal framing, and goal contexts have been studied and experimented within the framework of SDT as newer extensions (T. Kasser & Ryan, 1993, 1996; Vansteenkiste, Simons, Lens, Sheldon, & Deci, 2004; Vansteenkiste, Simons, Soenens, & Lens,

Address correspondence to Xiao Yong Hu, School of Psychology, Central China Normal University, 152 Luoyu Road, Wuhan, Hubei 430079, China. E-mail: huxiaoyong8123@yahoo.com.cn

2004). Goal contents reflect the outcomes that people are pursuing, whether for autonomous or controlled reasons (T. Kasser & Ryan, 1996).

The original SDT and its earlier extensions have been studied across cultures (e.g., Levesque, Zuehlke, Stanek, & Ryan, 2004) and contextual settings such as physical education (e.g., Haerens, Kirk, Cardon, De Bourdeaudhuij, & Vansteenkiste, 2010; Standage, Duda, & Ntoumanis, 2005), religious activities (e.g., Neyrinck, Vansteenkiste, Lens, Duriez, & Hutsebaut, 2006), health and medicine (e.g., Halvari & Halvari, 2006) politics (e.g., Losier & Koestner, 1999) and organizations (e.g., Deci, Connell, & Ryan, 1989). However, to date, the more recent extensions of experimental SDT research have mostly been studied in developed countries and in individualistic cultures (e.g., Vansteenkiste, Simons, Soenens, et al., 2004). As the largest population in the world, the Chinese differ in many ways from populations in developed countries in regards to cultural values, educational systems, and interpersonal relationship norms. Therefore, it is necessary to test the applicability of the recent extensions of SDT to the Chinese population. In this paper, we designed studies to examine goal content and goal context effects among Chinese middle school students in mainland China. We believe that testing those effects in a different culture would contribute to the advancement and universality of SDT.

1.1 Goal contents

People's goals vary. Research by Kasser and Ryan (1996) showed that people's long-term goals can be categorized as intrinsic or extrinsic. Intrinsic goals reflect people's inherent growth tendencies and yield an inward-oriented focus (e.g., self-development, health and physical fitness, community contribution, and affiliation), whereas extrinsic goals reflect people's desire to impress others by acquiring outward signs of worth (e.g., financial success, power, status, and physical attractiveness) and are characterized by an outward-oriented frame-of-

reference for viewing the world (T. Kasser & Ryan, 1996; Vansteenkiste, Lens, & Deci, 2006; Williams, Hedberg, Cox, & Deci, 2000). As a subtheory of SDT, goal contents theory distinguishes those two types of goals and examines their effect on behaviors and well-being (T. Kasser, 2002; T. Kasser & Ryan, 1996; Vansteenkiste, Lens, & Deci, 2006). Extrinsic and intrinsic goals are thought to relate differently to basic need satisfaction and therefore produce different psychological outcomes (Grouzet et al., 2005; T. Kasser & Ryan, 1996; V. G. Kasser & Ryan, 1999; Kim, Kasser, & Lee, 2003; Ryan et al., 1999; Schmuck, Kasser, & Ryan, 2000). Along this line, correlational studies have found that the relative importance of extrinsic goals relates negatively to adjustment outcomes and the relative importance of intrinsic goals relates positively to adjustment outcomes (Duriez, Vansteenkiste, Soenens, & De Witte, 2007; T. Kasser & Ryan, 2001; Williams et al., 2000). This basic pattern is called goal content effect and has been found in different cultures, including the United States (T. Kasser & Ryan, 1996), Germany (Schmuck et al., 2000), Spain (Romero, Gómez-Fraguela,& Villar, 2011), Russia (Ryan et al., 1999), and South Korea (Kim et al., 2003).

While long-term goal contents may reflect what people value, they should not be interpreted as goal motives, i.e., either as autonomous motivation that is volitional or as controlled motivation that involves the experience of being pressured or coerced. Vansteenkiste, Lens, and Deci (2006) used an example to demonstrate the difference between those two concepts:"… students could have an after-school job to earn money (extrinsic goal content) because they feel pressured by their parents (controlled motive) or because they value going to college and will need the money (autonomous motive)." Past research has shown that goal content and goal motives have unique effects on well-being and psychological adjustment (Sheldon, Ryan, Deci, & Kasser, 2004).

Nonetheless, short-term goals of students' learning activities may be manipulated and experimental manipulations represent the newer direction of goal contents theory research. Vansteenkiste and colleagues conducted such manipulations in order to induce different goal contents. Learning activities were framed by indicating their instrumentality for attaining future goals (intrinsic vs. extrinsic). It was found that goal contents could be manipulated to some extent and that manipulated goal contents affected behaviors and learning outcomes in much the same way that non-manipulated life goal contents affected well-being and adjustment outcomes (Vansteenkiste, Simons, Lens, Soenens, & Matos, 2005; Vansteenkiste, Simons, Soenens et al., 2004). Although in general, people may establish relatively stable intrinsic or extrinsic goals, goal content manipulations/influences are not rare in daily life. For instance, teachers may orient students' attention towards external signs of success, such as being well-known and admired (i.e., status), or being wealthy and rich (i.e., financial success); alternatively, they may encourage their students to develop their talents and skills (i.e., self-development), or to help people in need (i.e., community contribution). Placing different emphasis on these goal contents may result in different short or long-term outcomes. Findings from past research suggest that intrinsic goals are more conducive to individual functioning and adjustment than extrinsic goals (see Gollwitzer & Moskowitz, 1996; Ryan, Sheldon, Kasser, & Deci, 1996; Vansteenkiste, Timmermans, Lens, Soenens, & Van den Broeck, 2008) and that extrinsic goal framing (manipulation) undermined learning and persistence, regardless of the initial value placed on extrinsic goals (Vansteenkiste, Duriez, Simons, & Soenens, 2006).

1.2 Goal contexts

SDT is an organismic dialectical approach. It emphasizes the interactions between the active organism (the self) and the social context. SDT researchers have explored how social contexts can promote adjustment outcomes (e.g., Assor, Kaplan, & Roth, 2002; Deci, Eghrari, Patrick, & Leone, 1994; Deci & Ryan, 2000; Guay, Ratelle, & Chanal, 2008). Autonomy-supportive contexts tend to facilitate autonomous motivation, and controlling contexts tend to facilitate controlled motivation (Black & Deci, 2000; Sheldon & Krieger, 2007; Williams & Deci, 1996). SDT holds that, in contrast to controlling contexts, an autonomy-supportive environment is associated with more desirable effects (Assor, Roth, & Deci, 2004; Chirkov & Ryan, 2001; Grolnick, Kurowski, Dunlap, & Hevey, 2000; Grolnick & Pomerantz, 2009; Soenens & Vansteenkiste, 2005).

A goal can be introduced and communicated in either type of context. If short-term goal contents are manipulated in different learning contexts, we could test goal content effect, goal context effect, as well as any interaction between them. The effects of manipulating goal contents within autonomy-supportive or controlling learning contexts have been studied by Vansteenkiste and colleagues in Western culture (e.g., Vansteenkiste, Simons, Lens et al., 2004; Vansteenkiste, Simons et al., 2005). They found that when intrinsic goals were pursued in an autonomy-supportive context the outcomes were most conducive to psychological well-being and that even in a controlling context, intrinsic goals were associated with more positive outcomes than extrinsic goals. To explain these findings, Vansteenkiste and colleagues argued that the autonomy-supportive context allowed people to experience the congruence of pursuing an intrinsic goal that was closely aligned with their basic psychological needs, whereas controlling contexts tended to thwart basic need satisfaction.

1.3 Self-determination theory in Chinese culture

China has its own unique history and culture. The pursuit of fortune, political power, and high social status has been traditionally valued and set as the ultimate

goal by Chinese scholars and intellectuals. In ancient times, particularly since the Tang Dynasty (circa 700 AD), scholarship attained by vast reading was deemed as the fairest and most efficient way to climb the social ladder. Those excelling in scholarship (and tested via official examinations) were respected by all, and often assigned government positions by the emperor. Once becoming a government official, as a result of the raised status, the scholar would have an official residence and a state-provided salary, along with other benefits (Su, 2002). Because of this, the primary goal of most ancient Chinese scholars was to obtain an official position. Even as a youngster, Chinese were taught that scholarship was superior to all the other types of work and was the surest route to a better personal life (Chen & Uttal, 1988).

In today's China, scholarship is not tied to securing a government job, but is still very highly valued as a way to reach success. Academic achievement and good performance on exams remain the primary path of upward mobility, partly because of urban-rural economic and developmental differences and migration control (Chen & Uttal, 1988). However, the emphasis on achievement as well as high standards imposed by Chinese parents may result in students' relying on extrinsic motivations such as grades to maintain their interest in school.

China's basic education is examination-oriented (Chen & Uttal, 1988). In terms of policies and educational practices in China, nine years of education (Grade 1 through Grade 9; elementary and middle school education) is mandated by law; however, students have to compete very hard to enter top high schools through "Zhongkao," the Senior Secondary Education Entrance Examination. Typical middle school classrooms in China have approximately 60 or more students. Teaching and learning is often done in groups but assessment is almost always based on individual performance. Teaching is usually teacher-centered with students following instructions and passively receiving information. The teaching styles are usually controlling, restrictive, and authoritarian (Chiu, 1986). High-stakes testing, high educational expectations from parents, traditional values, and teaching practices that make comparisons transparent (e.g., test scores or rankings public to all teachers and/or students) contribute to a competitive school environment in China, even in middle schools.

Cross-cultural research has suggested that Chinese students are different from their Western counterparts in terms of psychological constructs such as perceptions of competence, task orientation, anxiety about academics, attributions for failure and success, and human values (Chiu, 1986; Hardré et al., 2006; Hong, 2001; Schwartz & Bilsky, 1990; Stevenson, Chen, & Lee, 1993). It also has been shown that autonomy is less supported in Asian societies (Iyengar & Lepper, 1999; Jang, Reeve, Ryan, & Kim, 2009; Kitayama, Markus, & Kurokawa, 2000; Olsen et al., 2002; Quoss & Wen, 1995; Schwartz, 1992; Triandis, 2001). In Chinese culture, the support of autonomy is not a common socialization practice because of the prevailing Confucian values (e.g., filial piety, humaneness, and ritual). Instead, Chinese culture emphasizes conformity and family interdependence (Bao & Lam, 2008; Chao & Tseng, 2002), and maintaining social harmony and family support is often seen as a lifelong obligation (Tseng, 2004). Thus, some cross-cultural perspectives suggest that the pursuit of autonomy hampers the development of satisfying rela tionships, and such conflicts might be especially problematic for the adjustment outcomes of individuals in collectivistic societies (Iyengar & Lepper, 1999).

Nevertheless, several studies based on SDT have indicated that the East-West differences are not so dramatic in terms of the relationships between autonomy and academic or psychological outcomes (Pu, 2006; Vansteenkiste, Lens, Soenens, & Luyckx, 2006; Vansteenkiste, Zhou, Lens, & Soenens, 2005). For

example, Vansteenkiste, Zhou, Lens, and Soenens (2005) found that autonomy was positively related to adaptive learning attitudes, academic success, and personal well-being among Chinese learners. They also argued that within the SDT framework autonomy should be defined in terms of individual, phenomenological experience, rather than in terms of interpersonal, culturally bounded values. Based on this argument, SDT researchers in cross-cultural studies should focus on relationships between, rather than the amounts of, psychological constructs and outcomes. So far, there has been no experimental research examining the goal content and goal context effects among students in China, and the research here filled this gap in the literature.

1.4 Present research

Before this research, we conducted a survey study to identify the goal contents pursued by Chinese scholars. One hundred and three college students responded to the survey, and 75% of those students expressed learning goals as (a) being financially successful, (b) being rich, or (c) having many expensive possessions. We also surveyed 204 high school students and found that 81% of them had similar goals. From these surveys, we concluded that in general the learning goal contents of contemporary Chinese students are mostly extrinsic.

The present research included two experimental studies. In Study 1, we examined differences in learning outcomes when extrinsic or intrinsic goals were induced to Chinese middle school students. Because those two types of goals were thought to relate differently to basic need satisfaction and therefore produce different psychological outcomes, we hypothesized that Chinese students with extrinsic goals would report less positive and more negative learning outcomes than those with intrinsic goals. In Study 2, we examined differences in learning outcomes when extrinsic or intrinsic goals were induced to Chinese middle school students in an autonomy-supportive or controlling

learning context. We hypothesized that the goal content variable would interact with the context variable such that the autonomy-supportive learning context would offset the negative effects of extrinsic goals and enhance positive effects of intrinsic goals.

Dependent variables in Studies 1 and 2 were test performance, free-choice engagement, and test anxiety. Following the positive psychological tradition, we chose test performance and free choice engagement as dependent variables. Those two variables have been used in similar studies of different populations (e.g., Vansteenkiste, Simons, Lens, et al., 2004). More important, they are the key outcome variables in SDT. In addition, test anxiety was measured as an outcome variable because it, too, is a key factor in undermining student performance (Hembree, 1988). Moreover, it is common across a broad spectrum of educational settings (Griffin & Griffin, 1998).

2　Study 1

2.1 Method

2.1.1 Pilot study

Before Study 1, we conducted a pilot study to check the manipulation of goal contents with 133 middle school students from the same city as in Studies 1 and 2. This pilot study was carried out in a similar procedure as used in Pilot Study 1 of Vansteenkiste, Simons, Lens, Sheldon, and Deci (2004). After random assignment, 66 participants were in the intrinsic goal condition and 68 were in the extrinsic goal condition. They received the same instruction sheet in each condition as in Study 1 and read the same text as in Study 1 on improving creativity. All participants answered questions concerning the importance of intrinsic goals (three items on a 5-point Likert scale) and the importance of extrinsic goals (another three items on a 5-point Likert scale) of improving creativity. On the importance of intrinsic goals, the mean was 10.80 (SD=2.58) and 8.94 (SD=2.38)

for those in the intrinsic goal condition and the extrinsic goal condition, respectively. The effect size (Cohen's d) was 0.75 (close to large effect). On the importance of extrinsic goals, the mean was 7.42 (SD=2.30) and 8.37 (SD=2.35) for those in the intrinsic goal condition and the extrinsic goal condition, respectively. The effect size (Cohen's d) was 0.41 (close to medium effect). The differences between students in the two goal conditions suggested effective goal content manipulation.

2.1.2 Participants

Participants were 188 middle school students from a city in central China. This city is a traditional agricultural economic zone where people's average social and economic status is at the middle level within China. In addition, this city has a long history of attaching high importance to education. As in most other Chinese middle schools, teaching in sampled schools is structured and teacher−centered. Students spend the majority of their day and evening time on in−class learning. Of those 188 participants in this study, 88 (47%) were female, and 100 (53%) were male. The mean age of the participants was 14.88 years (SD=0.71).

2.1.3 Procedure

The experiment took place during regular classes and students learned text materials about improving creativity as a class activity. We chose the topic of improving creativity because there was a nationwide initiative of quality education when the present study was conducted and cultivating creativity was considered part of this initiative. The teachers were contacted and agreed to participate without knowing anything about what was being examined until the study was completed. The teachers distributed written instructions (in Chinese) to students explaining their task.

There were two types of instruction sheets put in rotation in the same stack for the different manipulations. Those instruction sheets were handed out to students one by one, following the seating in each class to ensure randomization of the experimental manipulations. There were 95 participants in the intrinsic goal condition and 93 in the extrinsic goal condition. All instruction sheets were of similar length and looked similar with different content to ensure fidelity of the experiment. After receiving the instruction sheets, students were told to read them without any discussion. Students then engaged in the target activity of reading a text about improving creativity.

Next, each student was asked to write his or her name on the instruction sheet and then to turn it in at the end of the sessions (along with other materials subsequently explained). The instruction for the intrinsic goal condition read: "Learning how to improve creativity is very important. Reading the text about it will help you to know better about yourself and your potentials and hence contribute to your personal development." The instruction for the extrinsic goal condition included: "Learning how to improve creativity is very important. Reading the text about it will help you to make more money and to buy things you want through applying your knowledge."

After reading the text, students answered questions about their comprehension of it and completed the test anxiety inventory. Subsequently, students were told that there were additional exercise materials about improving creativity that they could practice if they chose to. Last, 1 week later, these exercise materials were collected and graded by their teachers.

2.1.4 Measures

Test performance Students' test performance was measured by eight questions following their reading the text on improving creativity. The first question was a multiple−choice item worth 10 points, and the next seven questions were short essay questions with the first five worth 10 points each and the last two 20 points each. The total possible score was 100 points (we used this scale because it is the most commonly used and familiar scale

for classroom testing at schools in mainland China; we also would like to point out that there is little standardized testing in mainland China). Two teachers graded all eight questions. They were given the highest possible score for each question and sample answers for three or four scores were provided for each essay question. However, specific coding criteria were not provided for every possible score. Test items focused on conceptual rather than rote learning and those questions were similar to typical Chinese reading comprehension questions asked during a Chinese test of middle school students. An example essay question was "What does the author try to tell us by using the example of Alfred Nobel in the sixth paragraph?" The total possible score of this question was 10 points. A two-point sample answer was "The author tries to tell us that Alfred Nobel had a notebook." A six-point sample answer was "The author tries to tell us that Alfred Nobel had the habit of writing down spontaneous ideas in a notebook." A 10-point sample answer was "By using the example of Alfred Nobel, the author tries to tell us that it is important to keep track of spontaneous ideas by writing them down and that we can cultivate creativity this way." The two teachers grading test items were blind to students' conditions and did not know about the study design or purpose. The correlation between the two teachers' ratings was .92. Their ratings were averaged to form a performance score for each student.

Free-choice engagement　Students were offered additional exercise materials about improving creativity. There were seven problems in the exercise materials. Four of them asked students to describe or to summarize an invention or discovery, and the other three problems asked them to think creatively to provide an answer. An example exercise problem asked students to think creatively about possible uses of a building brick. Students were asked to record time spent on each problem and were given 1 week to choose to work on those exercise materials. Two teachers rated students' responses to each question on a 0–2 scale with possible scores of 0, 0.5, 1.0, 1.5, and 2.0. A score of 0 was given if the student spent very little time (less than 10 min) or did not write any meaningful sentences. A score of 0.5 was given if the student spent some time (30 min to 2 hr) and provided answers in the most usual way (e.g., writing about how building bricks could be used to build apartments). A score of 1.0 was given if the student spent some time (30 min to 2 hr) and provided more answers involving some creative thinking (e.g., writing about how uses of building bricks could be categorized based on the purposes of the buildings/projects). A score of 1.5 was given if the student spent more than 2 hr and provided unusual but sensible answers involving some creative thinking (e.g., writing about how different attributes of building bricks such as weight, shape, and price should be considered when choosing building bricks). A score of 2.0 was given if the student spent more than 2 hr and provided answers that reflected "thinking outside of the box" (e.g., anthropomorphizing a building brick and having it go on travels to find the meaning of its life). The free-choice engagement score was the total score on the seven exercises. Thecorrelation between the two teachers' ratings was high (r=0.91). The final score was the average of teachers' ratings.

Test anxiety　We used the Test Anxiety Inventory to measure participants' test anxiety. The inventory was developed by Spielberger (1980) and was later translated into Chinese by Ye and Rocklin (1988). The Chinese version that was used in this study has been shown to have good reliability (Cronbach's α =0.88) and to be applicable to Chinese students (Ling & Fan, 2008). In this study, students responded to 20 items on a 4-point scale about how much they experienced specific symptoms of anxiety during the test following the reading. The total score was used to measure test anxiety. Cronbach's α of this measure was 0.87 in this study.

2.2 Results

Previous experimental goal content studies did not examine gender differences, either because only one gender group participated (e.g., Vansteenkiste, Simons, Lens, et al., 2004) or because of reasons unexplained (e.g., Sebire, Standage, & Vansteenkiste, 2009). We included gender as an independent variable to see whether it was a significant predictor. Adjusted variable means and associated standard errors of scores for intrinsic and extrinsic goal conditions were calculated for the two gender groups separately (see Table 1). For the free−choice engagement measure, it turned out that the majority of the participants (90%) responded to only one exercise problem. One reason may be that each of these exercises required a longer time to finish than those for the test performance measure (the requirement of each exercise was to write an essay of approximately 200 Chinese characters), and it was unlikely for students to complete all the questions within 1 week (group−learning consumed most of the time in class and there was little free time after class).

To test whether there were significant differences between the intrinsic and extrinsic goal conditions and between gender groups on test performance, free−choice engagement and test anxiety, we conducted a multivariate analysis of variance using a 2 (goal condition) × 2 (gender) design with the three outcome variables as dependent variables. By Wilks' criterion, the combined dependent variables were significantly different by goal condition, $F(3, 182)=17.96$, $p<0.001$, $\eta_p^2=0.228$ (close to large effect), but not by gender, $F(3, 182)=0.76$, $p=0.54$, or by the interaction

between goal condition and gender, $F(3, 182)=2.20$, $p=0.09$.

The main effects of goal condition on each of the dependent variables suggested that students framed in the intrinsic goal condition had better test performance, $F(1, 184)=15.56$, $p<0.001$, $\eta_p^2=0.08$, higher free−choice engagement, $F(1, 184)=14.64$, $p<0.001$, $\eta_p^2=0.07$, and less test anxiety, $F(1, 184)=19.79$, $p<0.001$, $\eta_p^2=0.09$. All those effects were medium or close to medium (Cohen, 1988).

2.3 Discussion

Study 1 provided initial support of goal content effects among Chinese middle school students on three learning−related outcomes: test performance, free−choice engagement, and test anxiety. According to SDT, learning environments that emphasize intrinsic (or extrinsic) goal contents may have similar functional effects on learning and achievement as individuals' relatively stable pursuit of intrinsic (or extrinsic) goals. Vansteenkiste and colleagues have demonstrated effects of manipulated goal contents on deep processing of learning materials, academic achievement, and persistence among students and adults in Western culture (e.g., Vansteenkiste, Simons, Lens, et al., 2004). This study extended the research and has found that goal content (through experimental manipulations) matters among Chinese middle school students as well.

3 Study 2

3.1 Method

3.1.1 Participants

A total of 395 middle school students from the same

Table 1. Adjusted Means and Standard Errors of Scores for Intrinsic and Extrinsic Goal Conditions ($N = 188$)

Goal	Test performance		Free−choice engagement		Test anxiety	
	M	SE	M	SE	M	SE
Intrinsic goal						
Female	45.43	2.25	0.75	0.07	32.89	1.28
Male	46.69	2.23	0.59	0.07	34.73	1.26
Extrinsic goal						
Female	37.81	2.41	0.32	0.07	38.54	1.37
Male	36.50	2.14	0.51	0.07	40.48	1.21

city in China as in Study 1 participated in this study as a regular class activity. One hundred and ninety participants (48%) were female, and 205 (52%) were male. The mean age of the participants was 14.46 years (SD = 1.11).

3.1.2 Procedure

Participants' teachers were first contacted and agreed to participate without knowing what the study was about until the study was completed. The teachers distributed written instructions (in Chinese) to their students prepared by the researchers.

There were four types of instruction sheets that were randomly distributed within each class (cell sizes ranged from 96 to 100). The four types of instruction sheets represent the four manipulation conditions: intrinsic goal in an autonomy-supportive learning context, extrinsic goal in an autonomy-supportive learning context, intrinsic goal in a controlling learning context, and extrinsic goal in a controlling learning context. The students and their teachers were not aware that there were different sets of instructions, and all instruction sheets were of similar length and looked similar with different content to ensure fidelity of the experiment. After reading their instructions, students were then engaged in a target activity of reading a text about improving creativity. Each student was asked to write his or her name on the instruction sheet and to turn it in at the end of the sessions (along with other materials explained below). As in Study 1, instructions for participants in the intrinsic goal conditions stated that "Learning how to improve creativity is very important. Reading the text about it will help you to know better about yourself and your potentials and hence contribute to your personal development." The instruction for the extrinsic goal conditions included "Learning how to improve creativity is very important. Reading the text about it will help you to make more money and to buy things you want through applying your knowledge." The learning context was also manipulated. Specifically, in the

autonomy supportive conditions, the instructions included "If you are interested, you may want to learn more about it. The following text provides information on this topic. You can decide to learn more about creativity enhancing strategies." In the controlling conditions, the instructions included "You must learn more about it. You do not have a choice. If you do not finish this required learning task, you will hardly graduate." The experiment had four conditions through manipulation of two factors: goal content and goal context.

After reading the text, students answered the same reading comprehension questions and completed the same Test Anxiety Inventory as in Study 1. They were also told that there were additional exercises about improving creativity that they could practice if they chose. Those exercises were the same as in Study 1. One week later, exercise books were collected and graded by their teachers in the same way as in Study 1.

3.1.3 Measures

Test performance, free-choice engagement, and test anxiety were measured the same way as in Study 1. In this study, the interrater reliability was .94 and .90 for test performance and free-choice engagement, respectively. Cronbach's α for the test anxiety measure in this study was 0.83.

3.2 Results

The adjusted means and standard errors of the three outcome variables for each condition are presented in Table 2. We conducted a multivariate analysis of variance to study how goal content and goal context may affect students' test performance, free-choice engagement, and test anxiety.

Because in this study, male and female students did not differ significantly on any of the three outcomes, and the gender variable or its interaction with the other independent variables was not statistically significant in the multivariate analysis of variance, gender was dropped

Table 1. Adjusted Means and Standard Errors of Scores for Intrinsic and Extrinsic Goal Conditions, by Learning Context ($N = 395$)

	Test performance		Free−choice engagement		Test anxiety	
Goal condition	M	SE	M	SE	M	SE
Autonomy−supportive context						
Intrinsic goal	44.36[ad]	1.23	0.67[be]	0.06	34.68[cf]	0.88
Extrinsic goal	35.84[ag]	1.25	0.28[bh]	0.06	40.44[c]	0.89
Controlling context						
Intrinsic goal	25.80[d]	1.23	0.48[e]	0.06	41.42[f]	0.87
Extrinsic goal	24.14[g]	1.23	0.51[h]	0.06	42.45	0.87

Note. Means with same−letter superscripts are statistically significant at the .05 level.

from the analysis. Therefore, the final multivariate analysis of variance was a 2 (goal content) × 2 (goal context) design with test performance, free−choice engagement, and test anxiety as the dependent variables.

With the use of Wilks' criterion, the combined dependent variables were significantly different by goal content, $F(3, 389) = 14.04$, $p<0.001$, $\eta_p^2 = 0.10$ (medium effect), goal context, $F(3, 389) = 58.66$, $p<0.001$, $\eta_p^2 = 0.31$ (large effect), as well as the interaction between goal content and goal context, $F(3, 389) = 09.21$, $p<0.001$, $\eta_p^2 = 0.07$ (close to medium effect).

The main effects of goal content indicated that students in the intrinsic goal condition had better test performance, $F(1, 391) = 017.00$, $p<0.001$, $\eta_p^2 = 0.04$ (small effect), higher free−choice engagement, $F(1, 391) = 8.33$, $p=0.004$, $\eta_p^2 = 0.02$ (small effect), and less text anxiety, $F(1,391) = 14.91$, $p<0.001$, $\eta_p^2 = 0.04$ (small effect) than those in the extrinsic goal condition, averaging across the two types of goal contexts. The main effects of goal context indicated that students in the autonomy−supporting learning context had better test performance, $F(1, 391) = 150.31$, $p<0.001$, $\eta_p^2 = 0.28$ (large effect), and less test anxiety, $F(1, 391) = 24.79$, $p<0.001$, $\eta_p^2 = 0.06$ (small to medium effect), but did not differ in free−choice engagement than those in the controlling learning context, averaging across the two types of goal contents. The interaction between goal content and goal context was statistically significant for all three outcome variables: test performance, $F(1, 391) = 7.72$,

$p=0.006$, $\eta_p^2 = 0.02$ (small effect), free−choice engagement, $F(1, 391) = 11.28$, $p=0.001$, $\eta_p^2 = 0.03$ (small effect), and test anxiety, $F(1, 391) = 7.24$, $p=0.007$, $\eta_p^2 = 0.02$ (small effect).

Simple effects of goal content suggested that goal induction made a difference in the autonomy−supportive learning context, but not in the controlling learning context. In an autonomy−supportive learning context, students in the intrinsic goal condition had better test performance, higher free− choice engagement, and less test anxiety (see Table 2). Simple effects of goal context revealed that the autonomy−supportive learning context produced better test performance, higher free−choice engagement, and less test anxiety than the controlling learning context in the intrinsic goal conditions. In the extrinsic goal conditions, students in the autonomy−supportive learning context had better test performance than those in the controlling learning context. The goal context did not make a difference on test anxiety for those in the extrinsic goal conditions. In addition, the autonomy−supportive learning context was associated with lower free−choice engagement than the controlling learning context in the extrinsic goal conditions.

3.3 Discussion

Study 2 was designed to extend the findings of Study 1 and to examine whether goal content manipulations would work differently in an autonomy−supportive learning context versus a controlling learning context. In

the autonomy–supportive learning context, similar goal content effects were observed as in Study 1. However, in the controlling learning context, goal content effects were not statistically significant on the three outcome variables. It is worth nothing that the most positive outcomes were obtained when the task was associated with intrinsic goal induction and was introduced in an autonomy–supportive context. This suggests that intrinsic goals were more fully engaged and accepted by individuals when they were encountered in an autonomy supportive climate. This finding is consistent with findings in Western culture (Vansteenkiste, Simons, Lens, et al., 2004).

In the extrinsic goal conditions, the autonomy–supportive learning context resulted in better test performance than the controlling context; however the learning context did not make a statistically significant difference on test anxiety. It is surprising that the autonomy–supportive/extrinsic–goal combination resulted in lower free–choice engagement than the controlling/extrinsic–goal condition. This finding of negative effect of autonomy–supportive context in the extrinsic–goal condition did not appear in experimental research among Western samples and was possibly a cross–culture difference. This inconsistency may be the result of various components of psychological control, such as love withdrawal (Ho, 1986), shaming procedures, and threats of abandonment (Wu et al., 2002), which are more frequent in Eastern societies, and the fact that those components are better accepted as a means of regulating Chinese adolescents' behaviors (Chao, 1994; Chao & Tseng, 2002; Olsen et al., 2002). As a consequence, once the application of these controlling strategies disappears, the behaviors may likely withdraw. It is also worth noting that although in the extrinsic goal conditions, free–choice engagement was greater in the controlling context than in the autonomy–supportive context, the greatest free–choice engagement was related to the combination of an intrinsic goal induced

in an autonomy–supportive context.

Conclusions and limitations. This was the first experimental research of goal content and goal context effects with samples of Chinese middle school students. The present research included two studies. In Study 1, we aimed to examine the effects of manipulating goal contents on learning outcomes and hypothesized that Chinese students with extrinsic goals would report less positive and more negative learning outcomes than those with intrinsic goals. In Study 2, we aimed to examine the effects of goal content manipulations in two learning contexts and hypothesized that the goal content variable would interact with the context variable such that the autonomy–supportive learning context would offset the negative effects of extrinsic goals and enhance positive effects of intrinsic goals.

The two studies provided evidence for our primary hypotheses and study results are generally in line with SDT and goal contents theory (Deci & Ryan, 2008; Vansteenkiste, Lens, & Deci, 2006). Framing a learning activity in terms of an intrinsic or extrinsic goal attainment resulted in different short–term outcomes. Intrinsic goals were more conducive than extrinsic goals on test performance, free–choice engagement, and test anxiety. Future research is needed to investigate whether similar results would be found in different samples, such as those in different regions of China and whether interventions focused on intrinsic goal contents would generate long–term positive outcomes. We have also found that goal contexts and goal contents interacted with each other such that the negative effect of extrinsic goals could be offset in an autonomy–supportive environment for academic outcomes such as test performance and that the positive effect of intrinsic goals was enhanced in an autonomy–supportive learning environment.

From Study 2, the most positive outcomes were produced in the intrinsic goal/autonomy–supportive

condition. However, the controlling context seemed to facilitate free−choice engage mentwhen an extrinsic goal was induced. This may be a cross−cultural difference and further research is needed to examine whether it would be replicated. The controlling context in this study likely made the extrinsic goal contents more salient by stating a consequence for not completing the reading task ("you will hardly graduate"). Under this condition, students might have felt pressure to do whatever they could to alleviate the possibility of the negative consequence. The controlling context may have had a carryover effect on free−choice engagement. The controlling context might also have been unclear, confusing, or unbelievable to students, resulting in ineffective manipulation (unlike in Study 1, there was no pilot study in Study 2 to check goal context manipulation). In the current two studies, the manipulated goal contents might also have served as short−term standards for success that students internalized. In Study 2, when students were offered choice after the controlling context was removed, they probably felt a sense of autonomy and the higher free−choice engagement might be a rebound.

Chinese culture values help to ensure that children will work diligently, and Chinese philosophy has traditionally emphasized malleability and the importance of the environment in the shaping and expression of human potential (Chen & Uttal, 1988). From this perspective, Chinese students are more likely to be cultivated to carry the incremental view of intelligence (Dweck, 1999) and more likely to hold approach−orientated goals (Elliot & Harackiewicz, 1996; Elliot & McGregor, 2001). However, goal orientations are different from goal contents. Goal contents reflect the outcomes that people pursue (Deci & Ryan, 2008), whereas goal orientations focus on the purpose—why an individual engages in certain behaviors (Kaplan, Middleton, Urdan, & Midgley, 2002). While goal contents could be manipulated as demonstrated in this and past research,

studies on achievement goal theory tend to treat goal orientations as attribute−like characteristics.

As a subtheory of SDT, goal contents theory grows out of the distinctions between intrinsic and extrinsic goals and their effect on motivation and wellness. However, the distinctions between intrinsic and extrinsic goals should not be confused with the distinctions between intrinsic and extrinsic motivation, or with the distinctions between autonomous and controlled motivation. Past research has suggested that what goals people pursue (extrinsic goal contents vs. intrinsic goal contents) and why people pursue them (autonomous vs. controlled motivation) have independent contributions to psychological well−being (Sheldon et al., 2004).

There are several limitations of the present research. First, although we used different reading instructions to induce intrinsic or extrinsic goals, students were different in their long−term learning goals. Although our pilot survey study suggested that the majority of Chinese students(in middle schools and in college) had extrinsic life goals, goal induction/framing may work differently for people with intrinsic life goals. Despite studies supporting SDT and showing that promoting extrinsic goals undermines learning regardless of whether the individuals are extrinsically or intrinsically oriented, there is a hypothesis that suggests induced goal contents would yield better learning outcomes when they are consistent with individuals' more stable life goal orientations (Hidi & Harackiewicz, 2000; Sagiv & Schwartz, 2000).

Second, the manipulations of experimental conditions were limited. In both studies, we used altered wording on a set of instructions. The goal content manipulations seemed effective from the pilot study of Study 1. However, we did not check the effectiveness of manipulating goal contexts. In Study 2, the controlling context was framed by including in the instruction "You must learn more about it. You do not have a choice. If you do not finish this required

learning task, you will hardly graduate." While using the phrases "must" and "have to" might have resulted in reduced sense of autonomy, the possible consequence ("you will hardly graduate" might have been confusing or unbelievable to students. In addition, although significant results were observed, we do not believe that our manipulations would have long–term effects. For example, the controlling or autonomy–supportive context in Study 2 would likely wane and the general learning context in the classroom would resume.

Third, as a result of limited resources, our samples were from only one city in central China and the participants were all in junior high schools in the year before graduation. Chinese students are usually under great academic pressure during this last year of junior high school. The results may be not generalizable to the larger population of Chinese students.

Fourth, we did not explore the mechanism of goal content and goal context effects. According to SDT, basic psychological needs may be used to explain goal context effects (Deci & Ryan, 2000). Past mediational analyses have found that autonomous motivation mediated goal content and social context effects on learning–related outcomes (Vansteenkiste, Lens, & Deci, 2006; Vansteenkiste, Simons, Lens, et al., 2004). Sheldon, Ryan, Deci, and Kasser (2004) claimed that the negative effects of extrinsic goals could be attributable to personality traits such as high insecurity, low self–esteem, or low cooperation. Vansteenkiste, Neyrionck, Niemiec, Soenens, Witte, and Brock (2007) suspected that traits such as neuroticism could also explain goal content effects.

In sum, we have examined goal content and goal context effects among samples of Chinese middle school students, using experimental studies. Our findings indicate that goal content matters, particularly in autonomy–supportive learning contexts, and that goal context matters, par– ticularly with intrinsic goal contents. These findings may provide some implications for creating optimal learning environments in Chinese middle schools.

References

Assor, A., Kaplan, H., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy–enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. *British Journal of Educational Psychology, 72*, 261–278. doi: 10.1348/000709902158883

Assor, A., Roth, G., & Deci, E. L. (2004). The emotional costs of parents' conditional regard: A self–determination theory analysis. *Journal of Personality, 72*(1), 47–88. doi: 10.1111/j.0022–3506.2004.00256.x

Bao, X.–H., & Lam, S.–F. (2008). Who makes the choice? Rethinking the role of autonomy and relatedness in Chinese children's motivation. *Child Development, 79*, 269–283. doi: 10.1111/j.1467–8624.2007.01125.x

Black, A. E., & Deci, E. L. (2000). The effects of instructors' autonomy support and students' autonomous motiva– tion on learning organic chemistry: A self–determination theory perspective. *Science Education, 84*, 740–756. doi: 10.1002/1098–237x(200011)84:6<740::aid–sce4>3.0.co;2–3

Chao, R. K. (1994). Beyond parental control and authoritarian parenting style: Understanding Chinese parenting through the cultural notion of training. *Child Development, 65*, 1111–1119. doi: 10.1111/1467–8624.ep7252822

Chao, R. K., & Tseng, V. (2002). Asian and American parenting. In M. H. Bornstein (Ed.), *Handbook of*

parenting (2nd ed., Vol. 4, pp. 59 – 93). Mahwah, NJ: Erlbaum.

Chen, C., & Uttal, D. H. (1988). Cultural values, parents' beliefs, and children's achievement in the United States and China. *Human Development, 31*, 351 – 358. doi: 10.1159/000276334

Chirkov, V. I., & Ryan, R. M. (2001). Parent and teacher autonomy-support in Russian and U.S. adolescents: Com- mon effects on welling-being and academic motivation. *Journal of Cross-Cultural Psychology, 32*, 618 – 635. doi: 10.1177/0022022101032005006

Chiu, L.-H. (1986). Locus of control in intellectual situations in American and Chinese school children. *International Journal of Psychology, 21*, 167 – 176. doi: 10.1080/00207598608247582

Cohen, J. (1988). *Statistical power analysis for the bahavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Deci, E. L., Connell, J. P., & Ryan, R. M. (1989). Self-determination in a work organization. *Journal of Applied Psychology, 74*, 580 – 590. doi: 10.1037/0021-9010.74.4.580

Deci, E. L., Eghrari, H., Patrick, B. C., & Leone, D. R. (1994). Facilitating internalization: The self-determination theory perspective. *Journal of Personality, 62*(1), 119 – 142. doi: 10.1111/1467-6494.ep9406221281

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavio*r. New York, NY: Plenum.

Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*, 227 – 268. doi: 10.1207/S15327965PLI1104 01

Deci, E. L., & Ryan, R. M. (2008). Facilitating optimal motivation and psychological well-being across life's domains. *Canadian Psychology, 49*(1), 14 – 23. doi: 10.1037/0708-5591.49.1.14

Duriez, B., Vansteenkiste, M., Soenens, B., & De Witte, H. (2007). The social costs of extrinsic relative to intrinsic goal pursuits: Their relation with social dominance and racial and ethnic prejudice. *Journal of Personality, 75*, 757 – 782. doi: 10.1111/j.1467-6494.2007.00456.x

Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development.* Philadelphia, PA: The Psychology Press.

Elliot, A. J., & Harackiewicz, J. M. (1996). Approach and avoidance achievement goals and intrinsic motivation: A mediational analysis. *Journal of Personality and Social Psychology, 70*, 461 – 475. doi: 10.1037/0022-3514.70.3.461

Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology, 80*, 501 – 519. doi: 10.1037/0022-3514.80.3.501

Gollwitzer, P. M., & Moskowitz, G. B. (1996). Goal effects on action and cognition. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 361 – 399). New York, NY: Guilford Press.

Griffin, M. M., & Griffin, B. W. (1998). An investigation of the effects of reciprocal peer tutoring on achievement, self-efficacy, and test anxiety. *Contemporary Educational Psychology, 23*, 298 – 311. doi: 10.1006/ceps.1998.0971 Grolnick, W. S., Kurowski, C. O., Dunlap, K. G., & Hevey, C. (2000). Parental resources and the transition to junior high. *Journal of Research on Adolescence, 10,* 465 – 488. doi: 10.1207/SJRA1004 05.

Grolnick, W. S., & Pomerantz, E. M. (2009). Issues and challenges in studying parental control: Toward a new concep- tualization. *Child Development Perspectives, 3*, 165 – 170. doi: 10.1177/1477878509104321

Grouzet, F. M. E., Kasser, T., Ahuvia, A., Dols, J. M. F., Kim, Y., Lau, S., . . . (2005). The structure of goal

contents across 15 cultures. *Journal of Personality and Social Psychology, 89*, 800–816. doi: 10.1037/0022-3514.89.5.800

Guay, F., Ratelle, C. F., & Chanal, J. (2008). Optimal learning in optimal contexts: The role of self-determination in education. *Canadian Psychology, 94*(1), 233–240. doi: 10.1037/a0012758

Haerens, L., Kirk, D., Cardon, G., De Bourdeaudhuij, I., & Vansteenkiste, M. (2010). Motivational profiles for secondary school physical education and its relationship to the adoption of a physically active lifestyle among university students. *European Physical Education Review, 16*, 117–139. doi: 10.1177/1356336×10381304

Halvari, A., & Halvari, H. (2006). Motivational predictors of change in oral health: An experimental test of self-determination theory. *Motivation and Emotion, 30*, 294–305. doi: 10.1007/s11031-006-9035-8

Hardre′, P. L., Chen, C. H., Huang, S. H., Chiang, C. T., Jen, F. L., & Warden, L. (2006). Factors affecting high school students' academic motivation in Taiwan. *Asia Pacific Journal of Education, 26*, 189–207. doi: 10.1080/02188790600937326

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research, 58*(1), 47–77.

Hidi, S., & Harackiewicz, J. M. (2000). Motivating the academically unmotivated: A critical issue for the 21st century. *Review of Educational Research, 70*, 151–179. doi: 10.3102/00346543070002151

Ho, D. Y. F. (1986). Chinese patterns of socialization: A critical review *The psychology of the Chinese people* (pp. 1–37).New York, NY: Oxford University Press.

Hong, Y.–Y. (2001). Chinese students' and teachers' inferences of effort and ability. In F. Salili, C.–Y. Chiu & Y. Y. Hong (Eds.), *Student motivation: The culture and context of learning* (pp. 105–120). New York, NY:

Plenum.

Iyengar, S. S., & Lepper, M. R. (1999). Rethinking the value of choice: A cultural perspective on intrinsic motivation. *Journal of Personality and Social Psychology, 76*, 349–366. doi: 10.1037/0022-3514.76.3.349

Jang, H., Reeve, J., Ryan, R. M., & Kim, A. (2009). Can self-determination theory explain what underlies the productive, satisfying learning experiences of collectivistically oriented Korean students? *Journal of Educational Psychology, 101*, 644–661. doi: 10.1037/a0014241

Kaplan, A., Middleton, M. J., Urdan, T., & Midgley, C. (2002). Achievement goals and goal structures. In C. Midgley (Ed.), *Goals, goal structures, and patterns of adaptive learning* (pp. 21–53). Mahwah, NJ: Erlbaum.

Kasser, T. (2002). Sketches for a self-determination theory of values. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of self-determination research* (pp. 123–140). Rochester, NY: The University of Rochester Press.

Kasser, T., & Ryan, R. M. (1993). A dark side of the American dream: Correlates of financial success as a cen–tral life aspiration. *Journal of Personality and Social Psychology, 65*, 410–422. doi: 10.1037/0022-3514.65.2. 410

Kasser, T., & Ryan, R. M. (1996). Further examining the American dream: Differential correlates of intrinsic and extrinsic coals. *Personality and Social Psychology Bulletin, 22*, 280–287. doi: 10.1177/0146167296223006

Kasser, T., & Ryan, R. M. (2001). Be careful what you wish for: Optimal functioning and the relative attainment of intrinsic and extrinsic goals. In P. Schmuck & K. M. Sheldon (Eds.), *Life goals and well-being: Towards a positive psychology of human striving* (pp. 116–131). Ashland, OH: Hogrefe and Huber.

Kasser, V. G., & Ryan, R. M. (1999). The relation of

psychological needs for autonomy and relatedness to vitality, well–being, and mortality in a nursing home. *Journal of Applied Social Psychology, 29*, 935 – 954. doi: 10.1111/j.1559– 1816.1999.tb00133.x

Kim, Y., Kasser, T., & Lee, H. (2003). Self–concept, aspirations, and well–being in South Korea and the United States. *The Journal of Social Psychology, 143*, 277 – 290. doi: 10.1080/00224540309598445

Kitayama, S., Markus, H. R., & Kurokawa, M. (2000). Culture, emotion, and well–being: Good feelings in Japan and the United States. *Cognition and Emotion, 14*(1), 93 – 124. doi: 10.1080/026999300379003

Levesque, C., Zuehlke, A. N., Stanek, L. R., & Ryan, R. M. (2004). Autonomy and competence in German and American university students: A comparative study based on self–determination theory. *Journal of Educational Psychology, 96*(1), 68 – 84. doi: 10.1037/0022–0663.96.1.68

Ling, X., & Fan, X. L. (2008). Analysis of application of test anxiety inventory in Changsha middle school students. *Chinese Journal of Clinical Psychology, 2*, 146 – 150.

Losier, G. F., & Koestner, R. (1999). Intrinsic versus identified regulation in distinct political campaigns: The consequences of following politics for pleasure versus personal meaningfulness. *Personality and Social Psychology Bulletin, 25*, 287 – 298. doi: 10.1177/0146167299025003002

Neyrinck, B., Vansteenkiste, M., Lens, W., Duriez, B., & Hutsebaut, D. (2006). Cognitive, affective and behavioral correlates of internalization of regulations for religious activities. *Motivation and Emotion, 30*, 321 – 332. doi: 10.1007/s11031–006–9048–3

Olsen, S. F., Yang, C., Hart, C. H., Robinson, C. C., Wu, P., Nelson, D. A., . . . Wo, J. (2002). Maternal psychological control and preschool children's behavioral outcomes in China, Russia, and the United States. In B. K. Barber

(Ed.), *Intrusive parenting: How psychological control affects children and adolescents* (pp. 235 – 262). Washington, DC: American Psychological Association.

Pu, Z. (2006). *The experimental research on internalization of extrinsic learning motivation of students at junior high school*. Changchun, China: Unpublished doctoral dissertation, Northeast Normal University.

Quoss, B., & Wen, Z. (1995). Parenting styles and children's satisfaction with parenting in China and the United States. *Journal of Comparative Family Studies, 26*, 265 – 280.

Romero, E., Go'mez–Fraguela, J. A., & Villar, P. (2011). Life aspirations, personality traits and subjective well–being in a Spanish sample. *European Journal of Personality*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/per. 815/pdfdoi:10.1002/per.815

Ryan, R. M., Chirkov, V. I., Little, T. D., Sheldon, K. M., Timoshina, E., & Deci, E. L. (1999). The American dream in Russia: Extrinsic aspirations and well–being in two cultures. *Personality and Social Psychology Bulletin, 25*, 1509 – 1524. doi: 10.1177/01461672992510007

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contem– porary Educational Psychology, 25*(1), 54 – 67. doi: 10.1006/ceps.1999.1020

Ryan, R. M., Sheldon, K. M., Kasser, T., & Deci, E. L. (1996). All goals are not created equal: An organismic perspective on the nature of goals and their regulation. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition and motivation to behavior* (pp. 7 – 26). New York, NY: Guilford Press.

Sagiv, L., & Schwartz, S. H. (2000). Value priorities and subjective well–being: Direct relations and congruity ef– fects. *European Journal of Social Psychology, 30*, 177 – 198. doi: 10.1002/(sici)1099–0992(200003/04)30:2<177::aid– ejsp982>3.0.co;2–z

Schmuck, P., Kasser, T., & Ryan, R. M. (2000). Intrinsic and extrinsic goals: Their structure and relationship to well-being in German and U.S. college students. *Social Indicators Research, 50*, 225–241. doi: 10.1023/a:100708400-5278

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 25, pp. 1–65). New York, NY: Academic Press.

Schwartz, S. H., & Bilsky, W. (1990). Toward a theory of the universal content and structure of values: Extensions and cross-cultural replications. *Journal of Personality and Social Psychology, 58*, 878–891. doi: 10.1037/0022-3514.58.5.878

Sebire, S. J., Standage, M., & Vansteenkiste, M. (2009). Examining intrinsic versus extrinsic exercise goals: Cognitive, affective, and behavioral outcomes. *Journal of Sport and Exercise Psychology, 31*, 189–210.

Sheldon, K. M., & Krieger, L. S. (2007). Understanding the negative effects of legal education on law students: A longitudinal test of self-determination theory. *Personality and Social Psychology Bulletin, 33*, 883–897. doi: 10.1177/0146167207301014

Sheldon, K. M., Ryan, R. M., Deci, E. L., & Kasser, T. (2004). The independent effects of goal contents and motives on well-being: It's both what you pursue and why you pursue it. *Personality and Social Psychology Bulletin, 30*, 475–486. doi: 10.1177/0146167203261883

Soenens, B., & Vansteenkiste, M. (2005). Antecedents and outcomes of self-determination in 3 life domains: The role of parents' and teachers' autonomy support. *Journal of Youth and Adolescence, 34*, 589–604. doi: 10.1007/s10964-005-8948-y

Spielberger, C. D. (1980). *Test anxiety inventory: Preliminary professional manual*. Palo Alto, CA: Consulting Psychol-ogists Press.

Standage, M., Duda, J. L., & Ntoumanis, N. (2005). A test of self determination theory in school physical education. *British Journal of Educational Psychology, 75*, 411–433. doi: 10.1348/000709904 × 22359

Stevenson, H. W., Chen, C., & Lee, S. (1993). Motivation and achievement of gifted children in East Asia and the United States. *Journal for the Education of the Gifted, 16*, 223–250.

Su, X. (2002). *Education in China: Reforms and innovations.* Beijing, China: China Intercontinental Press.

Triandis, H. C. (2001). Individualism-collectivism and personality. *Journal of Personality, 69*, 907–924. doi: 10.1111/1467-6494.696169

Tseng, V. (2004). Family interdependence and academic adjustment in college: Youth from immigrant and U.S.-born families. *Child Development, 75*, 966–983. doi: 10.1111/j.1467-8624.2004.00717.x

Vansteenkiste, M., Duriez, B., Simons, J., & Soenens, B. (2006). Materialistic values and well-being among business students: Further evidence of their detrimental effect. *Journal of Applied Social Psychology, 36*, 2892–2908. doi: 10.1111/j.0021-9029.2006.00134.x

Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist, 41*(1), 19–31. doi: 10.1207/s15326985ep4101 4

Vansteenkiste, M., Lens, W., Soenens, B., & Luyckx, K. (2006). Autonomy and relatedness among Chinese sojourners and applicants: Conflictual or independent predictors of well-being and adjustment? *Motivation and Emotion, 30*, 273–282. doi: 10.1007/s11031-006-9041-x.

Vansteenkiste, M., Neyrinck, B., Niemiec, C. P., Soenens, B., De Witte, H., & Van Den Broeck, A. (2007). On the

relations among work value orientations, psychological need satisfaction and job outcomes: A self-determination theory approach. *Journal of Occupational and Organizational Psychology, 80,* 251–277. doi: 10.1348/096317906×111024

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology, 87*, 246–260. doi: 10.1037/0022-3514.87.2.246

Vansteenkiste, M., Simons, J., Lens, W., Soenens, B., & Matos, L. (2005). Examining the motivational impact of intrinsic versus extrinsic goal framing and autonomy-supportive versus internally controlling communication style on early adolescents' academic achievement. *Child Development, 76*, 483–501. doi: 10.1111/j.1467-8624.2005.00858.x

Vansteenkiste, M., Simons, J., Soenens, B., & Lens, W. (2004). How to become a persevering exerciser? Providing a clear, future intrinsic goal in an autonomy-supportive way. *Journal of Sport and Exercise Psychology, 26*, 232–249.

Vansteenkiste, M., Timmermans, T., Lens, W., Soenens, B., & Van den Broeck, A. (2008). Does extrinsic goal framing enhance extrinsic goal-oriented individuals' learning and performance? An experimental test of the match per-spective versus self-determination theory. *Journal of Educational Psychology, 100*, 387–397. doi: 10.1037/0022-0663.100.2.387

Vansteenkiste, M., Zhou, M., Lens, W., & Soenens, B. (2005). Experiences of autonomy and control among Chinese learn-ers: Vitalizing or immobilizing? *Journal of Educational Psychology, 97*, 468–483. doi: 10.1037/0022-0663.97.3.468.

Williams, G. C., & Deci, E. L. (1996). Internalization of biopsychosocial values by medical students: A test of self-determination theory. *Journal of Personality and Social Psychology, 70*, 767–779. doi: 10.1037/0022-3514.70.4.767

Williams, G. C., Hedberg, V. A., Cox, E. M., & Deci, E. L. (2000). Extrinsic life goals and health-risk behaviors in adolescents. *Journal of Applied Social Psychology, 30*, 1756–1771. doi: 10.1111/j.1559-1816.2000.tb02466.x.

Wu, P., Robinson, C. C., Yang, C., Hart, C. H., Olsen, S. F., Porter, C. L., . . . Wu, X. (2002). Similarities and differences in mothers' parenting of preschoolers in China and the United States. *International Journal of Behavioral Development, 26*, 481–491. doi: 10.1080/01650250143000436.

Ye, R., & Rocklin, T. (1988). A cross-cultural study of test anxiety. *Psychological Science, 3*, 25–29.

Red is romantic, but only for feminine females: Sexual dimorphism moderates red effect on sexual attraction

Fangfang Wen, Bin Zuo, Yang Wu, Shan Sun, Ke Liu

(School of Psychology, Central China Normal University, Wuhan, Hubei Province, PR China)

Abstract　Previous researchers have documented that the color red enhances one's sexual attraction to the opposite sex. The current study further examined the moderating role of sexual dimorphism in red effects. The results indicated that red enhanced men＇s sexual attraction to women with more feminine facial characteristics but had no effect on ratings of perceived general attractiveness. Red clothing also had a marginally significant effect on men's sexual attractiveness. In addition, regardless of sexual dimorphism cues, male participants rated women with red as warmer and more competent. The underlying mechanisms of the red effect, the limitations of the current study, and suggestions for future directions are discussed.

Keywords　red effect; masculine; feminine; sexual attraction; perceived attractiveness

1 Introduction

Color is an omnipresent and inseparable property of objects that constitute our perceptual world. Academic research has abounded on the physics, physiology, and aesthetics of color, but not until recently did psychologists formulate theories on its psychological meaning and effects (Elliot and Maier, 2014). Color–in–context theory, advanced by Elliot and colleagues, argues that color carries psychologically relevant meanings that are rooted either in biological or social learning processes (though they are not necessarily mutually exclusive). Viewing a color in a specific context may automatically evoke several psychological processes, ranging from affect and cognition to behavior intentions (Elliot and Maier, 2012).

Recent studies have documented a unique effect of red on interpersonal attraction (Elliot and Niesta, 2008). Most researchers have concentrated on how and to what extent red enhances men's attraction to women. For example, several studies from America and Europe have consistently shown that women paired with a red background or clad in red are perceived as more sexually attractive by men (Elliot and Niesta, 2008; Guéguen, 2012a; Pazda, Elliot, and Greitemeyer, 2012, 2014; Re, Whitehead, Xiao,

and Perrett, 2011; Roberts, Owen, and Havlicek, 2010; Schwarz and Singer, 2013; Stephen and McKeegan, 2010). However, perceived general attractiveness and other traits were affected less or not at all (Elliot et al., 2010; Elliot, Tracy, Pazda, and Beall, 2013; Schwarz and Singer, 2013). Moreover, research has demonstrated that the effect of red could also be extended to behaviors. For example, when dating a girl dressed in red, men tend to ask questions that are more intimate and show more sexual interest during conversation (Niesta-Kayser, Elliot, and Feltman, 2010). They also offer more help (Guéguen, 2012a) and are more likely to send a contact solicitation in an online dating context (Guéguen and Jacob, 2013). Men in bars send more solicitations to women wearing red lipsticks (Guéguen, 2012b), and male customers to a restaurant tend to tip waitresses wearing red lipstick or red clothes more generously (Guéguen and Jacob, 2012, 2014). On the other hand, a few studies also investigated the red effect on male attractiveness as perceived by women. Through a series of experiments, Elliot et al. (2010) indicated that women find men paired with a red background or dressed in red more sexually attractive and desirable, and further established that the perception of status mediated the red effect on male attractiveness.

Both social conditioning and evolutionary biology can explain the red effect on sexual attraction. In terms of social conditioning, the red effect may be construed as resulting from the repeated pairing of red and some particular concepts, information, and experiences. Over time, these pairings will form strong implicit associations such that the perception of red alone can activate culturally-conditioned psychological reactions (Elliot and Maier, 2012, 2014). Because red has been associated with romance, passion, lust, and fertility across nearly every long-standing civilization (Hutchings, 2004; Kaya and Epps, 2004; Lee, 2006; Neto, 2002), it could be argued that the effect of red in enhancing sexual attraction is only a reflection of the color's past associative history. Alternatively, biological evolution can also account for the salience of red in sexuality, as red serves to signal sexual preparedness related to the reproductive process (Deschner, Heistermann, Hodges, and Boesch, 2004; Lynn, McCord, and Halliwell, 2007). During the estrus phase, when heightened sexual receptivity is conducive to conception, the genitals of female chimpanzees swell and redden as a sexual cue to males (Buss, 2008). Therefore, it could also be argued that red may suggest higher sexual opportunity for males. Pazda et al. (2012) found parallel mechanisms in humans, showing that sexual receptivity indeed mediates the red effect on women's sexual attractiveness. With regard to males, red usually indicates higher status in many vertebrates (e.g., Setchell and Dixson, 2001; for a short review, see Elliot and Maier, 2012, p. 94), and females evolve to prefer males who are high in status in order to solve the adaptive problem of securing resources during pregnancy (Buss, 2008). Elliot et al. (2010) verified the mediating role of status in red's effect on male sexual attractiveness. Thus, the effect of red in boosting sexual attractiveness may stem from its concomitance with physiological and psychological processes that are closely related to reproduction. It is crucial to note that although the two sources of red's effects have distinct mechanisms, they are not mutually exclusive and may operate conjointly (Elliot and Maier, 2014). The biological influence might drive or reinforce the cultural association of red, and the cultural norm may generalize or extend its inherent meaning.

Consistent with a biological evolutionary perspective, recently Schwarz and Singer (2013) found a moderator to the red effect. Their study showed that relative to menopausal women, only the younger women were perceived as more sexually attractive when paired with a red rather than white background. If red signals a possibility of conception and fertility and post-menopausal

women are unable to conceive, then the color red would lose its status as a signal of reproductive value. Therefore, red would cease to enhance men's sexual attraction. This possibility suggests that sociocultural perspectives alone do not explain the effect of red. If social learning theory is correct, and it is the cultural meaning and the resulting mental association of red that enhances sexual attraction, then red should have an effect on women's attractiveness regardless of whether or not they are menopausal. Additionally, there might be other boundary conditions for the red effect. It could be inferred from above that the factors influencing the perceived reproductive value or future fertility of a target can moderate or change the intensity or directions of the red effect. Consistent with this reasoning, since facial cues of sexual dimorphism (masculine and feminine) are another factor reflecting one's reproductive value in addition to age, they may also moderate the red effect. Therefore, not all individuals would be equally affected by red, and its effect might differ in individuals with different facial traits of masculinity and femininity.

Previous work has already established that the facial cues of sexual dimorphism may be linked with female reproductive value and male status. For female faces, the femininity cues may reflect the level of hormonal secretion, especially estrogen (Law−Smith et al., 2005), implying fertility and an advantage in reproduction (Gesquiere, Wango, Alberts, and Altmann, 2007; Morrison, Clark, Tiddeman, and Penton−Voak, 2010). As for male faces, facial masculinity is positively related to testosterone level (Kasperk et al., 1997; Penton−Voak and Chen, 2004; Pound, Penton−Voak, and Surridge, 2009; Roney, Hanson, Durante, and Maestripieri, 2006), and the level of testosterone is usually associated with status and dominance (e.g., Josephs, Sellers, Newman, and Mehta, 2006). Hence, it came as no surprise that males with masculine facial cues tend to have higher dominance

(Boothroyd, Jones, Burt, and Perrett, 2007; Perrett et al., 1998), which is often a signal to status attainment in groups (Anderson and Kilduff, 2009). Taken together, whereas femininity in females often signals higher reproductive value, masculinity in males is associated with dominance and status.

Given that the majority of recent studies on red effect have been conducted on American or European participants (for an exception, see Elliot et al., 2010, Experiment 4), a replication in a different culture is especially important. In China, red has traditionally been rich in meanings. Red was a symbol for status and stood for official titles in ancient official belief systems, and in folk belief systems red is still related to jubilation, auspiciousness, and fortune, such that traditional festivals like the Spring Festival have always been noted for red decorations (Chen, 2007). Critically, red in China is also closely entwined with images of females and sex. "Red face," a Chinese word, is a byword for young girl, and it could refer to an intimate female friend of a man (yet usually without a sexual relationship) (Shang, 2008). Traditional Chinese weddings are excessive in their use of red, and colloquially "fallen red" refers specifically to the act of "deflowering a virgin" and the blood stain that may be left (Eberhard, 2013). The evolutionary view that the meaning of red may at least partially stem from human biology would be given more credence if we found evidence in China that red can also enhance women's sexual attractiveness, but not general attractiveness, and that this effect is moderated by facial sexual dimorphism cues.

It is currently unknown whether all young targets are perceived as more sexually attractive when presented along with red, regardless of their masculine or feminine cues. It is the intention of this study to investigate this issue. We hypothesize based on evolutionary theory that women with feminine traits are more susceptible to the

red effect. Conversely, masculine traits signal relatively lower reproductive value and therefore may lower or even eliminate the red effect. As for men, in line with Elliot et al. (2010), we hypothesize that red only elevates the sexual attractiveness of masculine men. Conversely, if a sociocultural mechanism is the only viable explanation and the red effect stems entirely from the association between red and sex, fertility, or romance, the effect of red should be uniform despite the varying facial cues of sexual dimorphism.

2 Methods

2.1 Participants

A total of 299 students (149 men and 150 women) from a university in Wuhan participated voluntarily for a gift. We deleted the data from 16 participants because they did not follow the instructions, and their responses on the questionnaire were uninterpretable. This resulted in a final sample of 283 participants (139 males and 144 females). The mean age of participants was 20.95 years ($SD = 1.95$), and the range was from 18 to 26 years. All participants were heterosexual and without histories of mental disorder. The participants had normal or corrected to normal vision, and had normal color vision.

2.2 Design

The present study adopted a $2 \times 3 \times 2 \times 2$ mixed design, in which three between−subject variables were gender of participants (male, female), clothing color (red, blue, white), and cues of sexual dimorphism (masculine, feminine). The within−subject variable was the gender of the target (male, female). The dependent variables included 1) sexual attractiveness, 2) general attractiveness, and 3) warmth and competence. The sexual attractiveness was only rated on a target of opposite sex with participants, so the analysis of sexual attractiveness involves only a $2 \times 3 \times 2$ between−subject design.

2.3 Materials

Unlike previous research on red effects, we generated our stimuli using computer− synthesized (rather than natural) photos in order to manipulate the sexual dimorphism cues (DeBruine, Jones, Crawford, Welling, and Little, 2010; Wen and Zuo, 2012). First, we created an averaged facial archetype of both sexes using computer graphic techniques. The face images were obtained from a large database of a university containing facial photographs of male and female graduates posed with the same background, uniform luminance, and neutral facial expressions. A total of 321 images comprising 144 men and 177 women were available, from which 32 images were selected for each gender. Images with eyeglasses, moustaches, or jewelry were excluded. All 64 images were used to generate two average images with FantaMorph 4.0 software. We marked 179 key points in each face that, as a whole, delineated the shape and contour of the face and its delicate features. Two photos were averaged along the values registered by the 179 key points, and the averaged images were further averaged in an identical manner. The processes were repeated several times until we obtained a single averaged image for each sex (Wen and Zuo, 2012), which is shown in Figure 1.

With the archetype images created, we then chose another photo for each sex from the same database to create

Figure. 1. Averaged facial archetype of both sexes used in stimuli construction

assess perceived general attractiveness: "How attractive do you find this person?", "How good-looking do you think the person is?", and "If you meet this person face to face, how attractive do you think he/she is?" The Cronbach's alpha is .78 based on the opposite sex rating and .81 in the same sex rating. Due to the medium nature of the reliability, we also generated factor scores in each condition instead of simply summing the three items.

We conducted an exploratory factor analysis using the principal axis method to extract common factors for the opposite sex ratings. The scree plot suggested a single factor as most suitable, and it accounted for 69.17% of the total variances. Because there was only one factor, no rotation was needed. Based on the results, we generated a standardized index of perceived general attractiveness using a regression method. The index has a mean of zero.

The case is similar in same sex ratings. Only one factor emerged, which accounted for 73.07% of total variances. Hence, we generated a similar standardized index.

Warmth and competence. In addition to attractiveness ratings, the studies by Elliot et al. (2008) and Schwarz et al. (2013) both measured other traits of the targets such as sympathy, intelligence, agreeableness, kindness, etc. Based on the stereotype content model (Fiske, Cuddy, Glick, and Xu, 2002), the current study adopted two traits: competence and warmth. These traits could loosely encapsulate the dimensions of the aforementioned traits. The item for competence is "How competent do you think the person is?", and the item for warmth is "How warm do you think the person is?"

2.5 Procedure

Upon arriving at the laboratory, participants were randomly assigned to one of the 12 conditions formed by the $2 \times 3 \times 2$ design and were told that the goal of the experiment is to understand the impression formation process. The participants were then handed a folder containing a facial photo and a questionnaire. The participants were instructed to look at the photo for 5 seconds before filling out the questionnaire, which included demographic information and the measures of dependent variables. After completing the questionnaire, the participants were thanked, debriefed, given a gift, and dismissed. The Committee on Research Ethics approved the entire procedure and materials.

3 Results

3.1 Sexual attractiveness

We conducted a 2 (participant gender: male vs. female) $\times$ 3 (color of cloth: red, blue, and white) $\times$ 2 (sexual dimorphism cues: masculine vs. feminine) three-way between-subject ANOVA on the standardized index of sexual attractiveness. The results indicated a significant main effect of participant gender, $F(1,271) =7.80$, $p=0.006$, $\eta_p^2=0.03$. Specifically, male participants ($M=0.14$, $SD=0.81$) found the opposite sex target more sexually attractive than did female participants ($M=-0.13$, $SD=0.81$). The main effect of clothes color was marginally significant, $F(2,271) =2.80$, $p=0.06$, $\eta_p^2=0.02$. Whereas participants found the target wearing red ($M =0.15$, $SD=0.90$) slightly more sexually attractive than the one wearing white ($M =-0.13$, $SD=0.76$) ($p=0.06$), the differences between the targets wearing blue ($M=-0.02$, $SD=0.82$) vs. white ($p=0.75$), and blue vs. red ($p=0.37$) were both not significant. The main effect of sexual dimorphism cues did not reach significance, $F(1,271) =1.42$, $p=0.23$, $\eta_p^2<0.01$.

The two-way interaction between gender and sexual dimorphism cues was significant, $F(1,271)=6.32$, $p=0.01$, $\eta p^2=0.02$, and further analysis revealed that compared to masculine targets, male participants rated feminine targets as more sexually attractive, $F(1,271) =6.74$, $p=0.01$, $\eta_p^2=0.02$, but ratings from female participants did not differ, $F(1,271) =0.89$, $p=0.35$, $\eta_p^2<0.01$. All the remaining two-way interactions were insignificant ($ps >0.17$).

Figure. 3. Sexual attractiveness as a unction of sexual dimorphism, color of clothes, and participants gender
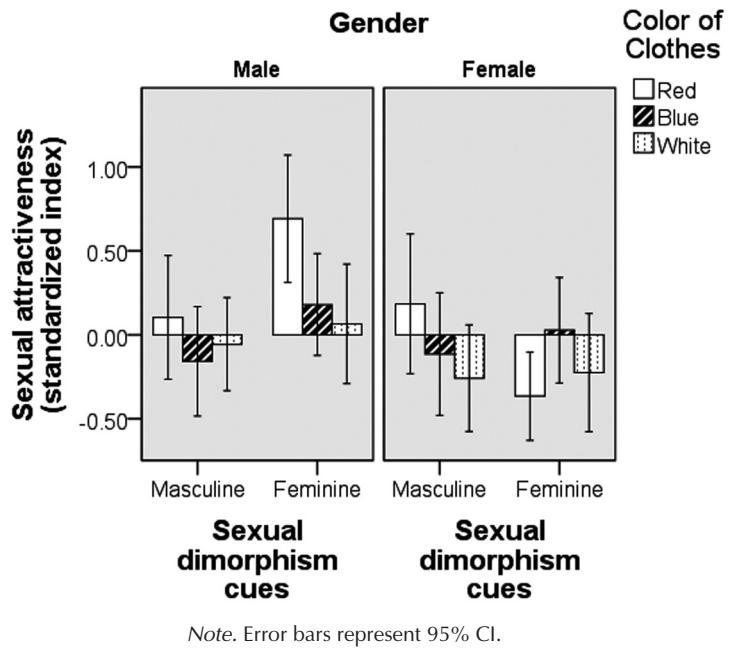
Note. Error bars represent 95% CI.

However, in accordance with our hypothesis we did find a significant three-way interaction among gender, color, and sexual dimorphism cues, $F(2,271) = 3.10$, $p=0.047$, $\eta_p^2=0.02$. To decompose the three-way interaction, we first tested the two-way interactions (or simple interaction) between color and sexual dimorphism separately under male and female conditions. For male participants, the two-way interaction was not significant, $F(2,271)=1.00$, $p=0.37$, $\eta_p^2<0.01$, but it indicated a trend toward significance for female participants, $F(2,271) =2.58$, $p=0.08$, $\eta_p^2=0.02$. Female participants rated masculine targets with red clothing more sexually desirable than feminine red targets, $F(1,271)=5.43$, $p=0.02$, $\eta_p^2=0.02$, and showed no such preference for masculinity in other color conditions (ps>0.5). To directly test the red effect, we planned a contrast between target with red and the other two colors, and found a trend of red to elevate the sexual attractiveness ratings of masculine targets, as rated by women, though it only reach marginal significance, $F(1,271)=3.25$, $p=0.07$, $\eta_p^2=0.01$.

Alternately, decomposing the three-way interaction from the perspective of sexual dimorphism proved to be clearer. The two-way interaction between gender and color of clothes was not significant under the masculine condition, $F(2,271)=0.45$, $p=0.64$, $\eta_p^2<0.01$, yet it was significant under the feminine condition, $F(2,271)=4.50$, $p=0.01$, $\eta_p^2=0.03$.

We examined the second-order simple effect associated with feminine conditions using Sidak corrections. The difference between photos with different colors was significant when rated by men, $F(2,271)= 3.96$, $p=0.02$, $\eta_p^2=0.03$. Consistent with our hypothesis, we planned a contrast between photos with red and the mean of the other two colors, and found that the red target received higher ratings on sexual attractiveness from male participants than targets with the two other colors, $F(1,271) =7.81$, $p=0.006$, $\eta_p^2=0.03$. However, the difference was completely nonsignificant when rated by women, $F(2,271) =1.56$,$p=0.21$, $\eta_p^2=0.01$ (see Figure 3).

Together, the results suggest that the red effect exists only for feminine women rated by male participants, but not for masculine women. Furthermore, we found a discernible trend for red to enhance the sexual attractiveness of masculine men in comparison with feminine men, as

Figure. 4. Perceived general attractiveness as a function of sexual dimorphism, color of clothes, and participants´ gender
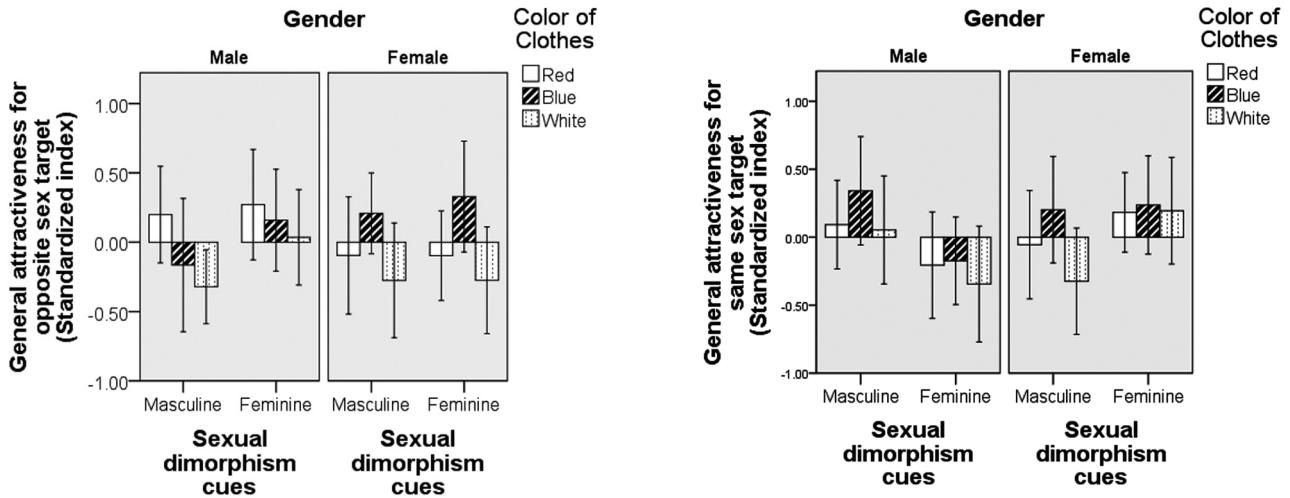
Note. Gender of the target: the left is opposite sex, the right is same sex. Error bars represent 95% CI.

rated by female participants, but the trend did not reach significance.

3.2 General attractiveness

We conducted a 2 (participant gender: male vs. female) × 3 (color of cloth: red, blue, and white) × 2 (sexual dimorphism cues: masculine vs. feminine) × 2 (target gender) four-way mixed design ANOVA on the standardized index of general attractiveness. The target gender is the within-subject variable, and the three other variables are all between-subject. The analysis showed a significant main effect of color of clothes, $F(2,270) = 4.64$, $p=0.01$, $\eta_p^2=0.03$. Post-hoc analysis with Sidak correction found that, overall, participants thought targets wearing blue clothes generally were more attractive than ones wearing white (MBlue−MWhite=0.30, SE =0.10, p= 0.008), and no significant differences were found between red and blue, or white and red (ps>0.13).

Additionally, the three-way interaction among target gender, participant gender, and sexual dimorphism cues was significant, $F(1,270)$ =11.00, p=0.001, η_p^2=0.04. Decomposing the three-way interaction based on the target gender, we found that the two-way interaction between participant gender and sexual dimorphism was significant only for same-sex targets, $F(1,270)$ =10.14, p=0.002, η_p^2

=0.04, but not for opposite sex targets, $F(1,270)$ =0.92, p =0.34, η_p^2<0.01. Further analyzing the second-order simple effects under the same-sex target condition, we found that male participants rated feminine same-sex targets (M=−0.24, SE=0.11) as less generally attractive than masculine same-sex targets (M =0.16, SE =0.10), $F(1,270)$ =7.25, p=0.008, η_p^2=0.03; meanwhile, female participants perceived feminine same-sex targets (M= 0.21, SE=0.10) as more generally attractive than masculine ones (M =−0.06, SE=0.11), though the difference was only marginally significant, $F(1,270)$=3.25, p=0.07, η_p^2=0.01, as shown in Figure 4.

All the other main effects, two-way interactions, three-way interactions and four-way interaction were insignificant (ps > 0.1).

3.3 Warmth

We conducted a similar 2 (participant gender: male vs. female) × 3 (color of cloth: red, blue, and white) × 2 (sexual dimorphism cues: masculine vs. feminine) × 2 (target gender) four-way mixed design ANOVA on warmth ratings. The target gender is a within-subject variable, and the rest are between-subject variables. As shown in Figure 5, the results revealed a significant main effect of target gender, $F(1,270)$=20.73, p<0.001, η_p^2=0.07; namely,

participants rated the opposite sex (M=3.29, SE=0.08) as less warm than same−sex targets (M=3.71, SE =0.07). The main effect of participant gender was also significant, $F(1,270)$=6.23, p =0.01, η_p^2=0.02. Compared with female participants (M =3.35, SE=0.08), male participants tended to give higher warmth ratings to all the targets (M =3.65, SE =0.09). However, the two−way interaction between participant gender and target gender was not significant, $F(1,270)$=2.13, p=0.15, η_p^2=0.01. Furthermore, we found a marginally significant two−way interaction between color of clothes and participant gender, $F(2,270)$=2.70, p=0.07, η_p^2=0.02. Further analysis of simple effects showed that male participants found targets wearing red warmer than did female participants, $F(1,270)$=7.48, p= 0.007, η_p^2=0.03, whereas for targets wearing white, male participants gave only slightly warmer ratings than did female participants, $F(1,270)$ =3.69, p=0.06, η_p^2=0.01; for targets wearing blue, there was no significant gender difference, $F(1,270)$ =0.15, p=0.70, η_p^2<0.01. Alternately, the differences across color of clothes under both male and female conditions were all insignificant, ps >0.11. Planned contrasts between red clothing and the two other colors reached marginal significance in male participants, $F(1,270)$ =2.87, p=0.09, η_p^2=0.01, but not in females

participants (p =0.55), suggesting a slight tendency for men to rate red targets warmer than other targets. All the remaining main effects, two−way interactions, three−way interactions, and four−way interaction were insignificant, ps >0.14.

3.4 Competence

A similar four−way mixed−design ANOVA was conducted on competence ratings. As seen in Figure 6, the results showed a significant two−way interaction between participant gender and target gender, $F(1,270)$ =14.46, p<0.001, η_p^2=0.05. Specifically, rating opposite sex targets, male participants tended to give higher competence ratings (M=4.66, SE=0.09) than did female participants (M =4.25, SE=0.09), $F(1,270)$ =9.81, p=0.002, η_p^2=0.04, yet when rating same−sex targets, male participants tended to give lower competence rating (M=4.41, SE=0.09) than female participants (M=4.65, SE=0.09), though it only reached marginal significance, $F(1,270)$ =3.27, p=0.07, η_p^2=0.01. In addition, the two−way interaction between color of clothes and participant gender was significant, $F(2,270)$=6.09, p=0.003, η_p^2=0.04. Simple effects analysis showed that the difference across color of clothing was significant in male participants, $F(2,270)$=4.29, p= 0.02, η_p^2=0.03, but not in female participants, $F(2,270)$= 2.00,

Figure. 5. Warmth rating as a function of sexual dimorphism, color of clothes and participants gender
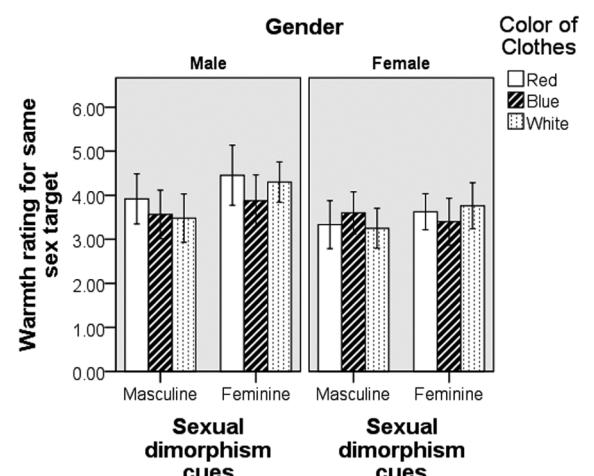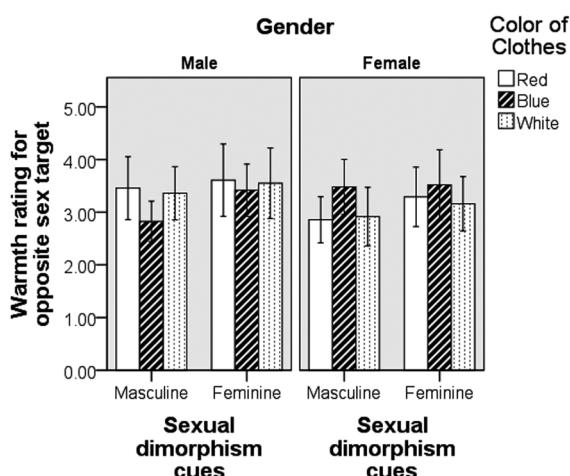
Note. Gender of the target: the left is opposite sex, the right is same sex. Error bars represent 95% CI.

p =0.14, η_p^2 =0.02. Planned contrasts between the color red and the two other colors in the male condition showed that male participants found targets in red more competent than other targets, $F(1,270)$ =6.16, p =0.01, η_p^2=0.02.

All other main effects, two-way interactions, three-way interactions and four-way interaction were not significant, ps >0.18.

relatively lower reproductive value and conflict with the information about fertility presumably imparted by the color red (Law-Smith et al., 2005; Morrison et al., 2010), thereby offsetting or canceling the red effect.

Men with masculine facial cues were perceived by female participants as more sexually attractive when wearing red clothes, but this comparison only reached marginal significance. This tendency is consonant with our
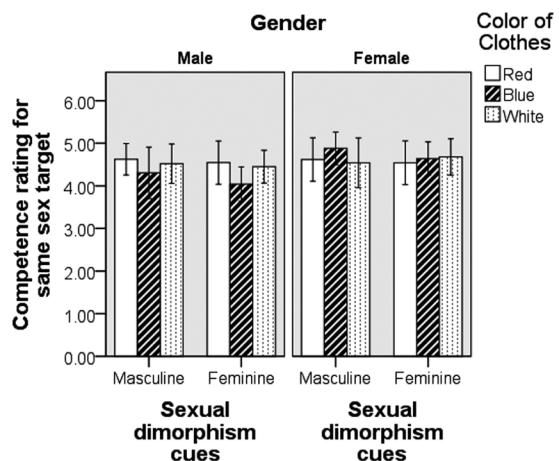
Figure. 6. Competence rating as a function of sexual dimorphism, color of clothes, and participants gender

Note. Gender of the target: the left is opposite sex, the right is same sex. Error bars represent 95% CI.

4 Discussion

The present experiment provides strong support for our hypothesized moderation of sexual dimorphism cues on the red effect and also provides a replication of the red effect in a Chinese context. The color red was shown to have a boosting effect on the sexual attractiveness of women with feminine facial traits when perceived by male participants. However, women with masculine facial traits did not benefit from red clothes. These findings could be explained from an evolutionary biological perspective, as feminine traits in women signal higher reproductive value. The most critical information in mating with the opposite sex would be gleaned from many cues. The perception of heightened reproductive value by presenting both red clothes and feminine traits enhances the sexual attractiveness of the female. In contrast, masculine traits in women may signal

hypothesis. Because masculinity and red both signal status in males, and females are likely to favor males with status and resources, in the current study we indeed showed that women tend to rate men with a red-and-masculinity combination as more sexually desirable. Meanwhile, two factors may have contributed to its failure to reach significance. First, the association between the color red and males is multi-faceted. Although red may enhance men's facial attractiveness, a man associated with too much red could be perceived as aggressive (Stephen, Oldham, Perrett, and Barton, 2012). This is further supported by the link between the facial reddening of a man and the blood suffusion brought about by anger or aggression (Drummond and Quah, 2001). Second, masculine facial cues often reflect levels of testosterone (Penton-Voak and Chen, 2004; Roney et al., 2006). Testosterone is related to various personality traits that may have negative connotations,

such as dominance or aggressiveness (Boothroyd et al., 2007; Hughes, Dispenza, and Gallup Jr., 2004; Perrett et al., 1998), and the perception of these traits may be more evident when targets are wearing neutral facial expressions (Hareli, Shomrat, and Hess, 2009; Tracy and Beall, 2011), as in the current study. Thus, the masculine cues in the current experiment may have triggered contradictory perceptions of the male target, which, as a result, could have offset the red effect of enhancing the sexual attractiveness of men. Given the dearth of studies on how red affects male sexual attractiveness, future research is needed to clarify the red effect across genders.

The present study also measured the perceived general attraction of the targets. The results demonstrated that male participants did not rate the opposite sex target paired with red as higher on the three measures, relative to the other colors. This result is generally consistent with previous studies (Elliot et al., 2010, 2013; Schwarz and Singer, 2013), yet it should be noted that Elliot and Niesta (2008) did document an effect of red on general attractiveness, though it was smaller in effect size. This slight disparity in results warrants further theoretical and empirical consideration.

Averaging over three colors, we found that whereas female participants did not favor either feminine or masculine women, the male participants perceived feminine men as less generally attractive. This could lend support to the gender dichotomization tendency that is characteristic of a male gender role (Bosson, and Michniewicz, 2013).

As for warmth and competence ratings, we found that, averaging over the sexual dimorphism cues, male participants tended to rate female targets wearing red as warmer and more competent than the targets wearing white or blue. This result on warmth and competence ratings slightly deviated from ratings on other related trait words in Elliot et al. (2010). Considering that it was not

moderated by sexual dimorphism cues, other mechanisms may have been responsible, such as halo effect. The close link between red and general positivity in Chinese culture, along with the status of red as being especially tied with females in Chinese culture, may also exert influences (Eberhard, 2013; Shang, 2008).

This study also has some limitations. First, as in Schwarz and Singer (2013), we used solely self-report ratings that are well documented to be susceptible to social desirability biases, especially on delicate issues such as direct judgment of others' attractiveness and of sexual interest toward a stranger. It could be more accurate and precise to employ biological measures to index one's sexual attraction to the targets.

In same sex ratings, we measured only perception of general attraction, warmth, and competence, dropping sexual attraction. It is counterintuitive to measure heterosexuals' sexual attraction to same-sex targets. Nevertheless, in future studies, same or modified items also measuring the perceived sexual attractiveness of same-sex participants could aid in the discriminant validity of the current measures.

Furthermore, a study to examine the effects of red in different mating or task contexts would be fruitful. Research has already suggested that the red effect may have completely different meanings in working or achievement contexts (Elliot and Maier, 2012, 2014; Elliot, Maier, Moller, Friedman, and Meinhardt, 2007). Thus, the moderation of sexual dimorphism on the red effect may be further moderated by context (i.e., perhaps the feminine female paired with red would be perceived as less competent in working conditions).

So far, studies about the red effect have focused on the perceivers, though it is likely that the color of clothing may also exert influences over the wearers. It is plausible that the color of clothing may change the sexual awareness of its wearer, and it could be the case that wearing red clothes

is a reflection of the sexual intention of the wearer from the beginning. The combination of data from the red-wearer and the perceiver would better capture the significance of red and would be an interesting avenue for future studies.

Another limitation of this study is that we did not directly test the alternative explanations. For example, one might argue that the red stimuli simply activated mental representation of traditional wedding ceremonies in China or the "red-light" district in the West due to cultural schema or stereotypes. Female newlyweds or female sex workers are supposed to have more feminine traits. Consistent with this argument, the reason that the red effect is specific to feminine women could be easily sought in social conditioning rather than evolution. Though such reasoning has certain inherent difficulties (e.g., Why did red not invoke an image of the red flag, but activated images of the "red-light" district?), we were not able to test these issues directly in this study. The cultural mechanism cannot be excluded from the red effect on sexual attraction, and it may operate at a different level and exert influence through social learning and acculturation processes, such as through the perceived sexual receptiveness of the individual wearing red clothes (Pazda et al., 2012). Moreover, Roberts et al. (2010) found that in certain settings, black dressing elevated perceived sexual attractiveness as well, which undermines the uniqueness of red and its purported evolutionary underpinnings. In light of these considerations, further research is needed to clarify the contributions of both cultural and biological mechanisms.

The simple perceptual mechanism that reduces the color effect to color contrast or luminance, as in Schwarz and Singer (2013), can be ruled out because the color configuration used in the current study is identical to previous research, and this interpretation could not explain the moderation role of sexual dimorphism cues. However, we did not use a spectrophotometer to check the color parameters of printed materials; this should be corrected in future replications.

Collectively, previous research has shown that red stimuli could enhance an individual's sexual attraction. The current investigation adds to this literature by showing that this effect is moderated by the facial sexual dimorphism cues of the targets. Furthermore, it is the first study to systematically examine the moderating roles of facial cues in explaining the red effect. This finding adds to the growing literature on the red effect and its boundary conditions, and the implications may stimulate future research.

References

Anderson, C., and Kilduff, G. (2009). Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance. *Journal of Personality and Social Psychology, 96*, 491–503.

Boothroyd, L. G., Jones, B. C., Burt, D. M., and Perrett, D. I. (2007). Partner characteristics associated with masculinity, health and maturity in male faces. *Personality and Individual Differences, 43*, 1161–1173.

Bosson, J. K., and Michniewicz, K. S. (2013). Gender dichotomization at the level of ingroup identity: What it is, and why men use it more than women. *Journal of Personality and Social Psychology, 105*, 425–442.

Buss, D. M. (2008). *Evolutionary psychology: The new science of mind* (3rd ed.). Boston: Allyn and Bacon.

Chen, J. (2007). *Worship and taboos about the colour RED: A study of RED in Chinese belief systems.* Unpublished master's thesis, Sichuan University, Chengdu, China.

DeBruine, L. M., Jones, B. C., Crawford, J. R., Welling, L. L., and Little, A. C. (2010). The health of a nation predicts their mate preferences: Cross-cultural

variation in women's preferences for masculinized male faces. *Proceedings of the Royal Society B: Biological Sciences, 277*, 2405–2410.

Deschner, T., Heistermann, M., Hodges, K., and Boesch, C. (2004). Female sexual swelling size, timing of ovulation, and male behavior in wild West African chimpanzees. *Hormones and Behavior, 46*, 204–215.

DiStefano, C., Zhu, M., and M?ndril?, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research and Evaluation, 14*, 1–20.

Drummond, P. D., and Quah, S. H. (2001). The effect of expressing anger on cardiovascular reactivity and facial blood flow in Chinese and Caucasians. *Psychophysiology, 38*, 190–196.

Eberhard, W. (2013). *Dictionary of Chinese symbols: Hidden symbols in Chinese life and thought*. London: Routledge.

Elliot, A. J., and Maier, M. A. (2012). Color-in-context theory. In M. P. Zanna, P. Devine, J. M. Olson, and A. Plant (Eds.), *Advances in experimental social psychology* (Vol. 45, pp. 61–125). Salt Lake City: Academic Press.

Elliot, A. J., and Maier, M. A. (2014). Color psychology: Effects of perceiving color on psychological functioning in humans. *Annual Review of Psychology, 65*, 95–120.

Elliot, A. J., Maier, M. A., Moller, A. C., Friedman, R., and Meinhardt, J. (2007). Color and psychological functioning: The effect of red on performance attainment. *Journal of Experimental Psychology: General, 136*, 154–168.

Elliot, A. J., and Niesta, D. (2008). Romantic red: Red enhances men's attraction to women. *Journal of Personality and Social Psychology, 95*, 1150–1164.

Elliot, A. J., Niesta Kayser, D., Greitemeyer, T., Lichtenfeld, S., Gramzow, R. H., Maier, M. A., and Liu, H. (2010). Red, rank, and romance in women viewing men. *Journal of Experimental Psychology: General, 139*, 399–417.

Elliot, A. J., Tracy, J. L., Pazda, A. D., and Beall, A. T. (2013). Red enhances women's attractiveness to men: First evidence suggesting universality. *Journal of Experimental Social Psychology, 49*, 165–168.

Fiske, S. T., Cuddy, A. J. C., Glick, P., and Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth, respectively, follow from perceived status and competition. *Journal of Personality and Social Psychology, 82*, 878–902

Gesquiere, L. R., Wango, E., Alberts, S. C., and Altmann, J. (2007). Mechanisms of sexual selection: Sexual swellings and estrogen concentrations as fertility indicators and cues for male consort decisions in wild baboons. *Hormones and Behaviors, 51*, 114–125.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6*, 430–450.

Guéguen, N. (2012a). Color and women hitchhikers' attractiveness: Gentlemen drivers prefer red. *Color Research and Application, 37*, 76–78.

Guéguen, N. (2012b). Does red lipstick really attract men? An evaluation in a bar. *International Journal of Psychological Studies, 4*, 202–209.

Guéguen, N., and Jacob, C. (2012). Lipstick and tipping behavior: When red lipstick enhance waitresses tips. *International Journal of Hospital Management, 31*, 1333–1335.

Guéguen, N., and Jacob, C. (2013). Color and cyber-attractiveness: Red enhances men's attraction to women's internet personal ads. *Color Research and Application, 38*, 309–312.

Guéguen, N., and Jacob, C. (2014). Clothing color and tipping: Gentlemen patrons give more tips to waitresses with red cloths. *Journal of Hospitality and Tourism Research, 38*, 275–280.

Hareli, S., Shomrat, N., and Hess, U. (2009). Emotional

versus neutral expressions and perceptions of social dominance and submissiveness. *Emotion, 9*, 378–384.

Hughes, S. M., Dispenza, F., and Gallup, G. G., Jr. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior, 25*, 295–304.

Hutchings, J. (2004). Color in folklore and tradition-The principles. *Color Research and Application, 29*, 57–66.

Josephs, R. A., Sellers, J. G., Newman, M. L., and Mehta, P. H. (2006). The mismatch effect: When testosterone and status are at odds. *Journal of Personality and Social Psychology, 90*, 990–1013.

Kasperk, C., Helmboldt, A., Borcsok, I., Heuthe, S., Cloos, O., Niethard, F., and Ziegler, R. (1997). Skeletal site-dependent expression of the androgen receptor in human osteoblastic cell populations. *Calcified Tissue, 61*, 464–473

Kaya, N., and Epps, H. H. (2004). Relationship between color and emotion: A study of college students. *College Student Journal, 38*, 396–405.

Law-Smith, M. J., Perrett, D. I., Jones, B. C., Cornwell, R. E., Moore, F. R., Feinberg, D. R., . . . Hillier, S. G. (2005). Facial appearance is a cue to oestrogen levels in women. *Proceedings of the Royal Society Series B: Biological Sciences, 273*, 135–140.

Lee, Y. (2006). *Man as the prayer*: The origin and nature of human kind. New York: Trafford.

Lynn, B. M., McCord, J. L., and Halliwell, J. R. (2007). Effects of menstrual cycle and sex on postexercise hemodynamics. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology, 292*, 1260–1270.

Meier, B. P., D'Agostino, P. R., Elliot, A. J., Maier, M. A., and Wilkowski, B. M. (2012). Color in context: Psychological context moderates the influence of red on approach- and avoidance-motivated behavior. PLOS

ONE, 7, e40333.

Morrison, E. R., Clark, A. P., Tiddeman, B. P., and Penton-Voak, I. S. (2010). Manipulating shape cues in dynamic human faces: Sexual dimorphism is preferred in female but not male faces. *Ethology, 116*, 1234–1243.

Neto, F. (2002). Colors associated with styles of love. *Perceptual and Motor Skills, 94*, 1303–1310.

Niesta-Kayser, D., Elliot, A. J., and Feltman, R. (2010). Red and romantic behavior in men viewing women. *European Journal of Social Psychology, 40*, 901–908.

Pazda, A. D., Elliot, A. J., and Greitemeyer, T. (2012). Sexy red: Perceived sexual receptivity mediates the red-attraction relation in men viewing woman. *Journal of Experimental Social Psychology, 48*, 787–790.

Pazda, A. D., Elliot, A. J., and Greitemeyer, T. (2014). Perceived sexual receptivity and fashionableness: Separate paths linking red and black to perceived attractiveness. *Color Research and Application, 39*, 208–212.

Penton-Voak, I. S., and Chen, J. Y. (2004). High salivary testosterone is linked to masculine male facial appearance in humans. *Evolution & Human Behavior, 25*, 229–241.

Perrett, D. I., Lee, K. J., Penton-Voak, I., Rowland, S., Yoshikawa, D. M., Burt, D. M., . . . Akamatsu, S. (1998). *Effects of sexual dimorphism on facial attractiveness. Nature, 394*, 884–887.

Pound, N., Penton-Voak, I. S., and Surridge, A. K. (2009). Testosterone responses to competition in men are related to facial masculinity. *Proceedings of the Royal Society B: Biological Sciences, 276*, 153–159.

Re, D. E., Whitehead, R. D., Xiao, D., and Perrett, D. I. (2011). Oxygenated-blood colour change thresholds for perceived facial redness, health, and attractiveness. PLOS ONE, 6, e17859.

Roberts, S. C., Owen, R. C., and Havlicek, J. (2010).

Distinguishing between perceiver and wearer effects in clothing color–associated attributions. *Evolutionary Psychology, 8*, 350‑364.

Roney, J. R., Hanson, K. N., Durante, K. M., and Maestripieri, D. (2006). Reading men's faces: Women's mate attractiveness judgments track men's testosterone and interest in infants. *Proceedings of the Royal Society of London Series B: Biological Sciences, 273*, 2169‑2175.

Schwarz, S., and Singer, M. (2013). Romantic red revisited: Red enhances men's attraction to young, but not menopausal women. *Journal of Experimental Social Psychology, 48*, 161‑164.

Setchell, J. M., and Dixson, A. F. (2001). Circannual changes in the secondary sexual adornments of semifree–ranging male and female mandrills (Mandrillus sphinx). *American Journal of Primatology, 53*, 109‑121.

Shang, X. (2008). Comparison of cultural connotations of "Red" in Chinese and English. *Journal of Linyi Normal University, 30*, 139‑141.

Stephen, I. D., and McKeegan, A. M. (2010). Lip colour affects perceived sex typicality and attractiveness of human faces. *Perception, 39*, 1104‑1110.

Stephen, I. D., Oldham, F. H., Perrett, D. I., and Barton, R. A. (2012). Redness enhances perceived aggression, dominance and attractiveness in men's faces. *Evolutionary Psychology, 10*, 562‑572.

Tracy, J. L., and Beall, A. T. (2011). Happy guys finish last: The impact of emotion expressions on sexual attraction. *Emotion, 11*, 1379‑1387.

Wen, F., and Zuo, B. (2012). The effects of transformed gender facial features on face preference of college students: Based on the test of computer graphics and eye movement tracks. *Acta Psychologica Sinica, 44*, 14‑29.

Perceptual simulation in gender categorization: Associations between gender, vertical height, and spatial size

Xiaobin Zhang[1], Qiong Li[1], Kendal J. Eskine[2], Bin Zuo[3,*]

([1] School of Psychology, Northwest Normal University, Lanzhou, China)

([2] Department of Psychological Sciences, Loyola University, New Orleans, Louisiana, United States of America)

([3] School of Psychology, Central China Normal University, Wuhan, China)

Abstract The current studies extend perceptual symbol systems theory to the processing of gender categorization by revealing that gender categorization recruits perceptual simulations of spatial height and size dimensions. In study 1, categorization of male faces were faster when the faces were in the "up" (i.e., higher on the vertical axis) rather than the "down" (i.e., lower on the vertical axis) position and vice versa for female face categorization. Study 2 found that responses to male names depicted in larger font were faster than male names depicted in smaller font, whereas opposite response patterns were given for female names. Study 3 confirmed that the effect in Study 2 was not due to metaphoric relationships between gender and social power. Together, these findings suggest that representation of gender (social categorization) also involves processes of perceptual simulation.

Keywords gender categorization; perceptual simulation;vertical height; spatial size

1 Introduction

Emerging evidence has revealed that human thought draws from one's embodiment, which refers both to actual bodily states and to simulations of embodied experiences in the brain's modality-specific systems for perception, action, and introspection. For example, Borghi, Glenberg, and

Kaschak showed that verifying whether a given object (e.g., a car) had a certain part (e.g., a roof) involved perceptual simulation processes such that when participants' reaction movement (upward or downward) corresponded with the position of the part relative to the object (e.g., roof vs. upward), the reaction was quicker. In another study, after watching an action cartoon, participants were asked to

describe the cartoon to a listener when the cartoon was no longer present. The result showed that participants who were prevented from gesturing (keeping stationary) processed the cartoon"s description significantly slower than control participants. Similarly, Tucker and Ellis found representations of grapes and hammers can be activated through simulations of motor processes involved in precision and power grips, respectively.

Such findings can be explained through a perceptual symbol systems account, which undergirds such "embodied cognition"effects. According to perceptual symbol systems theory, conceptual representations are tied to their perceptual basis and conceptual processing involves the partial simulation of those perceptual experiences that initially accompanied category exem-plars. Perceptual symbols are raw materials that make up the variable constructions (i.e., simulations) and they draw from all senses, including proprioception, introspection, and motor programs, and they are derived from multiple sources of direct experience.

Many studies have shown that processing abstract and concrete concepts activates modality-specific simulation of physical space. In one study, two words (e.g., root and branch) were presented above each other, and their order either followed the canonical arrangement (i.e., branch above root) or a contradictory arrangement (i.e., root above branch). Participants were asked to judge whether the two were related or not. The results showed that reactions were quicker when the arrangement of the words followed the canonical arrangement of the objects. Some studies demonstrated that judging valence involves perceptual simulation of the vertical spatial dimension, on which good is up and bad is down. Other studies showed that the representation *TIME* involves the perceptual simulation of both horizontal space dimension and spatial size dimensions. Along these lines, additional research has revealed thatthe representationandprocessing of

SOCIAL POWER also recruits vertical spatial perceptual simulation such that power =up and powerless=down. Additional findings indicate that power was also represented in terms of size cues in which power=big and powerless=small .

It has been argued that gender is the most frequently utilized domain in human categorization, yet there is a sparse amount of research investigating its embodiment. Some research has shown that gender category (male and female) also be grounded in sensorimotor metaphors. For example, Slepian and colleagues found that *MALE* was associated with the proprioceptive experience of "tough" and *FEMALE* was associated with the proprioceptive experience of "tender", and the authors explained the results by the perspective that the largest trait difference between the male and female was the difference in tenderness. In addition to traits like "tender" and "tough", it can be argued that stature is another important and salient physical difference between males and females. On average, men are taller and larger than women. According to the perceptual symbols account, the representation of gender may involve the perceptual simulation of experiences rooted in physical dimensions of vertical height and spatial size. The purpose of the present research was to demonstrate that the processes involved with gender categorization activate perceptual simulations involved with spatial dimensions. Based on the above analysis, we hypothesized that the representation of gender (male and female) will also involve the processing of perceptional simulation such that *MALE* categories will be processed faster along lines of greater size and higher verticality and vice versa for *FEMALES*.

2 Study 1

We conducted study 1 to determine if the representation and processing of gender involves the perceptual simulation

of vertical dimension. We predicted that the male faces would be judged faster if they were presented at the top of the computer screen, but the female faces would be judged faster if they were presented at the bottom of the computer screen.

2.1 Methods

Ethics statement. This study was reviewed and approved by the committee for the protection of subjects at Central China Normal University, School of Psychology Ethics Committee. Written consent was also obtained from each participant before the experiment according to the established guidelines of the committee. This procedure was followed in studies 2 and 3 as well.

Participants and design. Participants were 43 Chinese undergraduates (23 females) each whom were offered a notebook as compensation. The study had a 2 (facial gender: male or female)62 (position: top or bottom) repeated−measures design.

Stimulus, materials,and procedure. Eighty−eight white and black photographs of faces (44 males and 44 females) were selected on the basis of earlier prior testing. All picture files were standardized in size to 100×120, and all the faces displayed neutral expressions. Participants arrived to the laboratory individ− ually and were greeted by a male experimenter. They were seated facing the computer screen and told that the study was investigating aspects of face perception. The procedure was the same as the one used by. In each trial, a fixation cross was firstly presented at the center of the screen for 300 ms. Following this central cue, a subsequent fixation cross was flashed for 300 ms either at the 40% position (from top to down, above the central cue) or at the 60% position (from top to down, below the central cue) of the screen. A third fixation cross was flashed for 300 ms either at the 30% position (from top to down, above the central cue) or at the 70% position (from top to down, below the central cue) of the screen (in the same vertical direction as the second cross). The face pictures then appeared either at the 25% position (from

top to down, above the central cue) or at the 75% (from top to down, below the central cue) position of the screen for 2000 ms (in the same vertical direction as the third cross), or disappeared when the participants made a response. The spatial cues (cross) were intended to direct attention to the spot of the picture"s appearance, and thereby reduce random spatial exploration and additional error variance . All of the fixations and pictures appeared centered horizontally on the screen. Participants were required to report, by means of a key press, whether each face depicted a male or female target as quickly and accurately as possible, and the response keys were counterbalanced across the sample. If the response was inaccurate, the word "incorrect"appeared in a red font for 1500 ms. Accurate trials were separated by a blank screen for 500 ms. On completion of the experiment, participants were debriefed and dismissed.

2.2 Results and discussion

Mean categorization latencies served as the dependent measure of interest. Given the presence of extreme responses in the data set, response times that were slower than 2.5 standard deviations were excluded from the analysis, as were trials on which errors were committed . This resulted in 0.09% (350) of the data being excluded from the statistical analysis. Latencies were then log−transformed to normalize their distribution. For ease of interpretation, however, the untransformed means are reported in Figure 1. The transformed latencies were submitted to a 2 (facial gender: male vs. female) $\times 2$(position: top vs. bottom) ANOVA. The analysis revealed no main effects of position [$F(1, 42) =0.005$, $p=0.94$, $\eta^2 <0.001$] and facial gender [$F(1,42)=2.690$, $p=0.108$, $\eta^2 =0.060$]. As we expected, the facial gender $\times$ position interaction was significant [$F(1, 42) =9.028$, $p=0.004, \eta^2 <0.177$; see fig. 1]. Simple effects analysis demonstrated that responses to male faces in the up position ($M=548.78$, $SD=90.30$) were significantly faster than in the down position ($M=566.85, SD=108.48$),

$[F(1, 42)=4.58, p= 0.038, \eta^2=0.098]$. Alternatively, responses to female faces in the up position (M=552.58, SD=99.71) were significantly slower than in the down position (M=539.79, SD=90.98), $[F(1, 42)=5.30, p=0.026, \eta^2=0.112]$.

The results of study 1 indicated that the representation of gender involved the perceptual simulation of vertical space. If the gender−linked spatial height (e.g., male face was presented on the top of screen) was consistent with the actual physical feature of gender, gender categorization was facilitated. If gender and spatial height were incongruent, gender categorization was degraded.
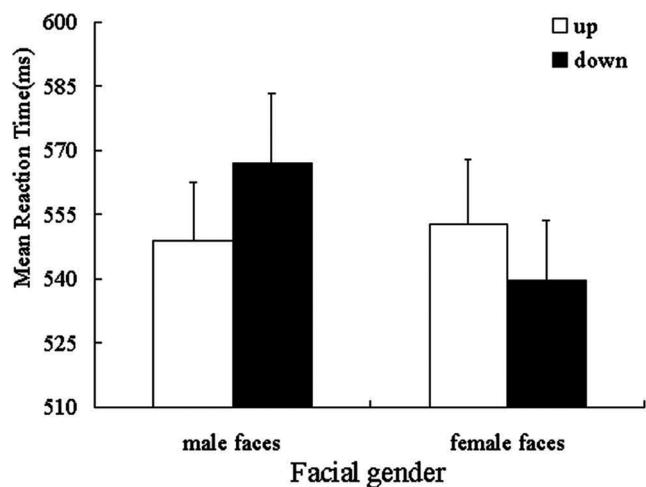
Figure. 1. Mean reaction latency as a function of facial gender and position (study 1). Error bars Indicate standard error of the mean.
doi:10.1371/journal.pone.0089768.g001

3 Study 2

In addition to vertical height, size is also a salient physical difference between males and females. We conducted study 2 to determine if the representation and processing of gender involves the perceptual simulation of spatial size. We predicted that male names would be judged faster if they were presented in bigger font size, whereas female names would be judged faster if they were presented in smaller font size.

3.1 Methods

Participants and design. Participants were 48 Chinese undergraduates (23 males) who were offered a notebook as compensation. The experiment had a 2 (name type: male names or female names) × 2 (name font size: big or small) repeated− measures design.

Stimulus, materials, and procedure. Eighty names (40 male names and 40 female names) were selected on the basis of earlier pilot testing. One hundred and twenty−two common names (66 male names, 66 female names, not including the names of famous people) were chosen from the internet, which 28 participants (14 males) rated on 7−point scales (extremely suited to female, "1" to extremely suited to male, "7"), "the extent the name can be used for male or female". We then chose the highest scored 40 male names (the lowest score is 6.07, $M = 6.44$, $SD = 0.219$), and we chose the lowest scored 40 female names (the highest score is 1.36, M=1.36, SD= 0.73). The scores of the male name were significantly higher than the female name, $t(39) = 219.44$, $p<0.001$. Finally, male names and female names were matched with the same 40 common family name.

Participants arrived at the laboratory individually and were greeted by a male experimenter. They were seated facing the computer screen and told that the study was investigated people"s ability to classify names by gender. In each trial, a fixation cross was firstly presented at the center of the screen for 800 ms, at which point a name appeared at the center of the screen for 2000 ms or disappeared when the participants made a response. Twenty male (female) names were presented at the screen in large font size (70 point), and twenty different male (female) names were presented on the screen in small font size (25 point). The inter−trial interval was 250 ms. Participants were required to report, by means of a key press, whether each name was a male name or female name as quickly and accurately as possible, and the

response keys were counterbalanced across the sample. If the response was inaccurate(slower than2000 ms), theword "incorrect" ("please be quicker") appeared in a red font for 1000 ms. On completion of the experiment, participants were debriefed and dismissed.

3.2 Results and discussion

Mean categorization latencies served as the dependent measure of interest. Given the presence of extreme responses in the data set, response times that were slower than 2.5 standard deviations were excluded from the analysis, as were trials on which errors were committed. This resulted in 0.05% (191) of the data being excluded from the statistical analysis. Latencies were then log–transformed to normalize their distribution. For ease of interpretation, however, the untransformed means are reported in Figure 2. The transformed latencies were submitted to a 2 (name type: male names or female names) × 2 (name font size: big or small) ANOVA. The analysis revealed no main effects of name gender [$F(1, 47)=1.69$, $p=0.20$, $\eta^2=0.035$]. The main effect of name size was significant [$F(1, 47)=4.48$, $p=0.04$, $\eta^2=0.087$], such that participants were faster to categorize big font size names($M=589.20$, $SD=78.90$) than small font size names ($M=598.16$, $SD=86.62$). Most importantly, as predicted, the name type × name size interaction was significant[$F(1,47)= 78.26$, $p<0.001$, $\eta^2=0.625$; see fig. 2]. Simple effects analysis demonstrated that responses to big sized male names ($M=575.70$, $SD=69.33$) were significantly faster than small sized male names ($M=618.12$, $SD=85.34$), [$F(1, 47)=57.30$, $p<0.001$, $\eta^2=0.549$]. Contrarily, responses to the big sized female names ($M=602.70$, $SD=88.47$) were significantly slower than small sized female names ($M=578.20$, $SD=87.91$), [$F(1,47)=22.44$, $p<0.001$, $\eta^2=0.323$].

The results of study 2 indicated that representation of gender also involved the perceptual simulation of spatial size such that male was associated with larger spatial size,

and female was associated with smaller spatial size.

4 Study 3

The results of studies 1 and 2 suggest that representations of gender involved perceptual simulation processes of spatial height and size dimensions. In study 1, participants showed a congruency effect in categorizing male and female faces, high and low, respectively. In study 2, participants showed a congruency effect in categorizing male and female names, in large and small font, respectively. But these effects can also be driven by the metaphors of social power. According to the Chinese tradition culture, the power and social status of men were higher than women, and the men were thought to make greater contributions to society than women. Even now, men occupy more senior official careers than women in China. Until January 1, 2010, there were only three women in the 26 ministers of Chinese government departments, accounting for 11.5% (Institute of Women, the All China Women's Federation). Previous research has demonstrated that vertical positions were important metaphors of power, and the powerful was associated with up and the powerless
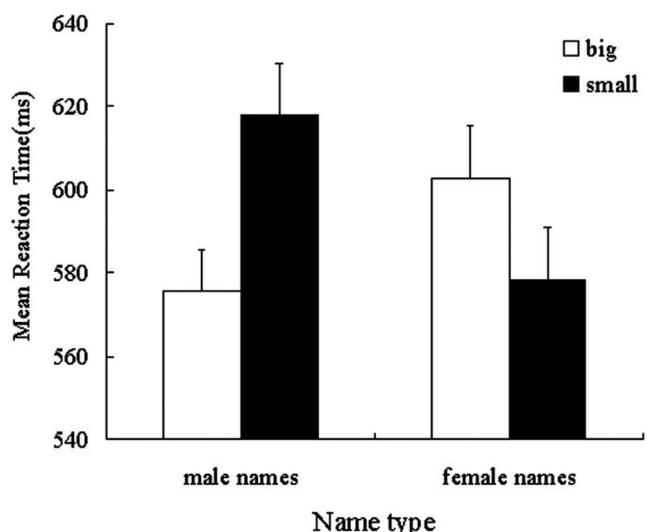
Figure 2. Mean reaction latency as a function of name type and name font size (study 2). Error bars indicate standard error of the mean.
doi:10.1371/journal.pone.0089768.g002

was associated with down, and mental representation of power was also associated with size cues. Participants reacted faster when powerful groups appeared on top and powerless groups appeared at the bottom, and participants reacted faster when the names of powerful groups appeared in big size (compared to small size) and vice versa for powerless groups. Based on the above analysis, the effect found in studies 1 and 2 may also result from the metaphors of power. To examine the possibility that these results are outcomes of metaphor endorsement, rather than perceptual symbols, in study 3 we recruited a group of participants who do not endorse the power−verticality metaphor for gender.

4.1 Methods

Participants and design. Two hundred and two undergraduates (184 females, 18 males; age ranged from 18 to 24) were chosen randomly to complete Attitudes Towards Women Scale (AWS), a 25−item questionnaire. The AWS measures traditional and conservative attitudes of women's place, including separate factors of rights, position relative to men, freedom, family role, and legal rights for college−aged participants. The 25 items are measured on a scale ranging from 1 (strongly disagree) to 7 (strongly agree), where lower scores indicate more traditional, antifeminist views and higher scores indicate more positive and pro−feminist attitudes. The AWS has been used in Korea, Taiwan, and China and was found to have good validity and reliability in these samples. We then chose the 38 highest scoring undergraduates (37 females, 1 male, age ranged from 18 to 22) to complete the experiment. Their scores ranged from 141 to 173 (M=148.86, SD=5.99). The experiment had a 2 (name type: male names or female names) ×2 (name font size: big or small) repeated−measures design.

Stimulus, materials, and procedure. The procedure was the same as study 2.

4.2 Results and discussion

Mean categorization latencies served as the dependent measure of interest. Given the presence of extreme responses in the data set, response times that were slower than 2.5 standard deviations were excluded from the analysis, as were trials on which errors were committed. This resulted in 0.05% (139) of the data being excluded from the statistical analysis. Latencies were then log−transformed to normalize their distribution. For ease of interpretation, however, the untransformed means are reported in Figure 3. The transformed latencies were submitted to a 2 (name type: male names or female names) × 2(name font size: big or small) ANOVA. The analysis revealed no main effect of name font size [$F(1, 37)$ =2.88, p=0.098, η^2=0.072]. The main effect of name type was significant [$F(1, 37)$=4.48, p=0.04, η^2=0.108], such that participants were faster to categorize female names (M=590.69, SD=74.15)than male names (M=603.16, SD=74.23). As in study 2, the name type × name size interaction was significant [$F(1, 37)$=42.19, p<0.001, η^2=0.53; see fig. 3]. Simple effects analyses demonstrated that responses to big sized male names (M=584.33,SD=71.87) were significantly faster than smaller sized male names (M=621.99, SD=76.59), [$F(1, 37)$=28.73, p<0.001, η^2=0.44]. Conversely,
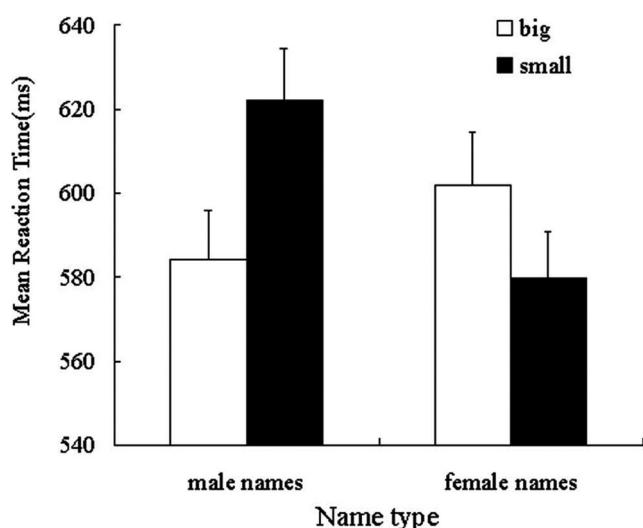
Figure 3. Mean reaction latency as a function of name type and name font size (study 3). Error bars indicate standard error of the mean.
doi:10.1371/journal.pone.0089768.g003

responses to big sized female names (M=601.70, SD=79.42) were significantly slower than small sized female names (M=579.67, SD=68.88), [$F(1, 37)$=12.20, p=0.001, η^2=0.25].Including AWS scores as a covariate, we repeated the above analysis as an ANCOVA. The analysis revealed no main effect of AWS scores $F(1, 36)$=2.29, p=0.67, η^2=0.005]. The interactions between AWS scores and other factors were also not significant, ps>0.28.

The results of study 3 indicated that feminists, who do not endorse the power−verticality metaphor for gender, represent and process gender in similar spatial and size dimensions, thus confirmed that the effects reported in studies 1 and 2 are not a result of "social power" metaphors.

5 General discussion

A wide array of research now supports the view that human cognition is grounded in and shaped by sensorimotor experiences and that our conceptual representations include sensory, motor, and introspective activations that are recruited into partial simulations, which reenact various embodied states. While previous research has revealed that gender−category representa−tions include sensorimotor information related to handling hard and soft objects and proprioceptive experience (toughness) , no research to date has explored how salient gender differences in size affect the processing of gender representations. In the present research, we have provided evidence that the representation and processing of gender activates simulations involved with vertical height and spatial size dimensions, with *MALE* being associated with up and big and *FEMALE* being associated with down and small. Importantly, by recruiting feminist participants in a separate study and replicating the observed effects, we ruled out the alternative explanation

that concepts involving "social power" might be driving these effects. These results are consistent with perceptual symbol systems theory by spotlighting how processing concepts can be affected by their perceptual basis, which we have extended to include the social categorization of gender in terms of vertical height and spatial size.

While the present findings could potentially be explained by either conceptual metaphor or perceptual symbol accounts, we proposed the latter approach appears to have the most explan atory power. First, it has been argued elsewhere that some metaphors (such as vertical position) are more basic than others. One implication would be that basic metaphors (vertical position) are so prevalent because they draw from similar embodied experiences and thus use perceptual simulation in order to ground their meaning. That is perceptual simulation may be the basic of some basic metaphors. Second, the results from the feminist participants suggest that metaphors are not being used, as they would naturally recruit different metaphors regarding size and gender. Third,metaphorserves to only make abstract relationships more concrete. In actuality, males are on average taller than females. Additionally, on average, men are larger than women. Yet men being larger and taller than women are not abstract in this sense, and it can be seen directly and this correlation then, is based in concrete sensory experience. Finally, a perceptual account is more parsimonious and is better aligned with one's general empirical experiences in that men and women typically correspond to specific physical sizes.

Ourdataprovide furtherevidence thatearly perceptual processes contribute to social−categorical thinking. Whereas prior research demonstrated that facial cues and visual acuity can affect the speed of social categorization, the studies presented here demonstrated vertical height and spatial size can also contribute to social−categorical thinking. Just as visual cues are present on every human

face and are therefore likely to have a ubiquitous influence on categorical thinking about other people, spatial position and size are normally present for males and females and are therefore likely to have a ubiquitous influence on gender categorization. In particular, a person"s position (e.g., sitting on a high or low stool) or stature (e.g., taking up more or less physical space) might influence how they are judged and potentially how they judge others in the social world.

Based on present studies, further research should further explore whether and how the present findings influence formation of specific gender stereotypes. The previous studies showed that the high position was associated with good and the low position was associated with bad. We can infer that because processing the concept of male (female) involved perceptual stimulation associating with high (low) position which is relevant with positive(negative) valance respectively, the perceptual simulation involving gender categorization may play an important role in the formation of negative stereotypes towards women. For example, a lot of studies have shown that people had the stereotype that woman can not reach high achievements in mathematics. Future research should explore the formation of this negative female stereotype by the perspective of representation of concept of gender (male and female) based on the present study. Further, it would be interesting to test whether inconsistent perceptual–gender properties (e.g., very tall women and very small men) influence social judgments in stereotype inconsistent ways. Specifically, additional research could explore if the activation of female stereotypes to (such as females are timid and so on) who are in high positions may be inhibited compared to females in low positions, and whether the activation of male stereotypes (such as males are rude and so on) who are in low positions may be inhibited compared to males in high positions.

References

Barsalou, L. W. (2008). Cognitive and neura1 contributions to understanding the conceptual system. *Current Directions in Psychologica1 Science, 17,* 91–95.

Barsalou, L. W. (1999) Perceptual symbol systems. *Behavioral and Brain Sciences, 22,* 577–609.

Borghi, A. M., Glenberg, A. M., Kaschak, M. P. (2004) Putting words in perspective. *Memory & Cognition, 32,* 863–873.

Boroditsky, L. (2001). Does language shape thought?: Mandarin and English speakers' conceptions of time. *Cognitive Psychology, 43,* 1–22.

Boroditsky, L. (2000). Metaphoric structuring: Understanding time through spatial metaphors. *Cognition, 75,* 1–28.

Boroditsky, L., & Ramscar, M. (2002) The roles of body and mind in abstract thought. *Psychological Science, 13,* 185–189.

Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *The Journal of Abnormal and Social Psychology, 42,* 33–44.

Bruner, J. S., & Postman, L. (1948). Symbolic value as an organizing factor in perception. *The Journal of Social Psychology, 27,* 203–208.

Byrne, Z. S., Felker, S., Vacha–Haase, T., & Rickard, K. M. (2011) A comparison of responses on the Attitudes Toward Women scale and Attitudes Toward Feminism scale. *Measurement and Evaluation in Counseling and Development, 44,* 248–264.

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition, 106,* 579–593.

Chia, R., Allred, L. J., & Jerzak, P. A. (1997) Attitudes toward women in Taiwan and China: Current status,

problems, and suggestions for future research. *Psychology of Women Quarterly, 21,* 137–150.

Cloutier, J., Mason, M. F., Macrae, & C. N. (2005). The perceptual determinants of person construal: Reopening the social cognitive toolbox. *Journal of Personality and Social Psychology, 88,* 885–894.

Crawford, L. E., Margolies, S. M., Drake, J. T., Murphy, M. E. (2006). Affect biases memory of location: Evidence for the spatial representation of affect. *Cognition and Emotion, 20,* 1153–1169.

Daugherty, C. G., & Dambrot, F. H. (1986). Reliability of the Attitudes toward Women Scale. *Educational and Psychological Measurement, 46,* 449–453.

Eccles, J. S., & Jacobs, J. E., Hld, R. E. (1990). Gender role stereotypes, expectancy effects, and parents' socialization of gender differences. *Journa1 of Socia1 Issues, 46,* 183–201.

Feingold, A. (1994), Gender differences in personality: A meta–analysis. *Psychological Bulletin, 116,* 429–456.

Fennema, E. H., & Sherman, J. A. (1977). Sex–related differences in mathematics achievement, spatial visualization, and sociocultural factors. *American Educational Research Journal, 14,* 51–71.

Gentner, D., Imai, M., & Boroditsky, L. (2002), As time goes by: Evidence for two systems in processing spaceR time metaphors. *Language and cognitive processes, 17,* 537–565.

Giessner, S. R., & Schubert, T. W. (2007). High in the hierarchy: How vertical location and judgments of leaders' power are interrelated. *Organizational Behavior and Human Decision Processes, 104,* 30–44.

Haslam, N., Rothschild, L., & Ernst, D. (2000). Essentialist beliefs about social categories. *British Journal of Social Psychology, 39,* 113–127.

Institute of Women, The All China Women's Federation. (2011). The main status of the international women's political participation in 2010. *Briefing of Information Research, 3,* 1–10.

Landau, M. J., Meier, B. P., & Keefer, L. A. (2010). A Metaphor–Enriched Social Cognition. *Psychological Bulletin, 136,* 1045–1067.

Meier, B. P., & Robinson, M. D. (2006). Does "feelingdown" mean seeing down? Depressive symptoms and vertical selective attention. *Journal of Research in Personality, 40,* 451–461.

Meier, B. P., & Robinsonm, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science, 15,* 243–247.

Moeller, S. K., Robinson, M. D., & Zabelina, D. L. (2008) Personality dominance and preferential use of the vertical dimension of space. *Psychological Science, 19,* 355–361.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth–Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9,* 184–211.

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114,* 510–532.

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological Science, 7,* 226–231.

Spence, J. T., Helmreich, R., & Stapp, J. A. (1973) A short version of the Attitudes Toward Women Scale (AWS). *Psychonomic Bulletin, 2,* 219–220.

Schnall, S. (2013). *Are there basic metaphors?.* In: Landau M, Robinson MD, Meier BP. Metaphorical thought in social life. Washington, DC: American Psychological Association

Schubert, T. W., Waldzus, S., & Giessner, S. R. (2009). Control over the association of power and size. *Social Cognition, 27,* 1–19.

Chia, R., Allred, L. J., & Jerzak, P. A. (1997). Attitudes toward women in Taiwan and China: Current status,

problems, and suggestions for future research. *Psychology of Women Quarterly, 21*, 137–150.

Slepian, M. L., Weisbuch, M., Rule, N. O., & Ambady, N. (2011). Tough and tender: Embodied categorization of gender. *Psychological Science, 22*, 26–28.

Smith, R. L., & Bradley, D. W. (1980). In defense of the Attitudes toward Women Scale: an affirmation of validity and reliability. *Psychological Reports, 47*, 511–522.

Spence, J. T., Helmreich, R., & Stapp, J. A. (1973). A short version of the Attitudes Toward Women Scale (AWS). *Psychonomic Bulletin, 2,* 219–220.

Spence, J. T., & Helmreich, R. (1972). The Attitudes Toward Women Scale: an objective instrument to measure attitudes toward the rights and roles of women in contemporary society. *Journal Supplement Abstract Service,* 2, 66.

Swim, J. K. (1994). Perceived versus meta−analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology, 66,* 21–36.

Tucker, M., & Ellis, R. (2001). The potentiation of grasp types during visual object categorization. *Visual Cognition, 8*, 769–800.

Wapner, S., Werner, H., & Krus, D. M. (1957). The effect of success and failure on space localization. *Journal of Personality, 25*, 752–756.

Zwaan, R. A., & Yaxley, R. H. (2003). Spatial iconicity affects semantic relatedness judgments. *Psychonomic Bulletin & Review, 10*, 954–958.

Gaze direction and brightness can affect self-reported emotion

Xiaobin Zhang [1], Qiong Li [1], Bin Zuo [2, *]

（ [1] School of Psychology, Northwest Normal University, Lanzhou, China ）

（ [2] School of Psychology, Central China Normal University, Wuhan, China ）

Abstract Previous studies revealed that emotion (pleased or depressed) could bias perception in a metaphorically consistent manner (e.g., happy=white (up), depressed=dark (down)). The present study extended this view by investigating whether these metaphors can also affect the emotion of an observer in a metaphorically consistent manner. In Experiment 1, after gazing at a black screen, participants became more depressed and less pleased temporarily. Conversely, after gazing at a white screen, participants became more pleased and less depressed temporarily. Results from Experiment 2 revealed that after gazing at the top of the screen, participants felt more pleased and less depressed temporarily but felt the reverse when gazing at the bottom of the screen. These results suggest that metaphors can, at least temporarily, affect the emotion of an observer along a pleased–depressed dimension.

1 Introduction

According to metaphor representation theory (Lakoff & Johnson,1999), people conceptualize non–perceptual states in perceptual terms. For example, a growing body of studies has suggested that vertical position and brightness are metaphors of affect: "up" and "brightness" are associated with good, and "down" and "black" have negative connotations (Crawford, Margolies, Drake, & Murphy, 2006; Meier & Robinson, 2004; Meier, Robinson, & Clore, 2004). In addition to affect, emotions of pleased and happy are considered to be light in color or high in vertical position as compared to emotion of depressed

states. Specifically, people typically depict "pleased" and "depressed" with its corresponding metaphors as up (light) and down (darkness), respectively (Meier & Robinson, 2006), so we contend that 'brightness'/'spatial position' may only affect emotion on pleaseded–pressed dimension not on other emotions (such as proud, scared). Our prediction is based upon the metaphoric mapping of 'brightness' / 'spatial position' and emotion on happye– depressed dimension. Common, everyday expressions highlight these metaphors: "I feel up today (when one feels happy)" "I' m at a low point (when onefeelsdepressed)," and "A smile brightens one's face (when one feels happy)." These metaphoric mappings seem to be a

---

* Corresponding author. E–mail address: zhangxiaobin624@163.com (B. Zuo).

universal phenomenon, as the expression 'bright smile' is manifested in many languages, including English, German, Italian, Korean, Chinese, and Russian. In addition, some studies have corroborated the association between emotion (on pleasede depressed) and its metaphors, such as spatial position and bright ness (up/bright−down/darkness) (Meier & Robinson, 2006; Song, Vonasch, Meier, & Bargh, 2012). specifically, Meier and Robinson (2006) argue that neuroticism and depressed symptoms can affect participants' vertical selective attention. Subjects who score higher on measures of neuroticism or depressed symptoms are faster at detecting lower (versus higher) spatial attention targets. Song et al. (2012) confirmed the metaphoric mapping of brightness and smiling whereby people judge smiling faces as perceptually brighter than frowning faces.

In fact, prior to psychologists using concepts such as metaphor and embodied cognition to explain the association between emotion and its metaphors, researchers had outlined the association between emotion and certain perceptual inputs (Teasdale,1993;Thayer, 2003).Some earlier studies have found that emotion can affect perception in a way that is congruent with its metaphors (i.e., pleased=up; depressed=down) (Fisher, 1964; Wapner, Werner, & Krus, 1957). One study revealed that participants who felt happy (i.e., they just received an A on a midterm exam) exhibited an upward bias when horizontally bisecting a lu− minous square. However, participants who felt sadness (i.e., just received an F on a midterm exam) exhibited a downward bias on this task (Wapner et al., 1957).

Several prior studies suggest that metaphors might be bidirectional. For instance, abstract metaphorical expressions are often grounded in physical experiences, and concepts within abstract domains influence related physical experiences (Landau, Meier, & Keefer, 2010; Williams, Huang, & Bargh, 2009). Based on the asso

ciation between emotion and its metaphors, prior studies have explored how emotion on the pleused−depressed dimension affects perceptual judgment (vertical position and brightness judgments), and confirmed that emotion on the pleased−depressed dimension were related with the metaphors of up position/ brightness and down position/darkness respectively (Fisher, 1964; Meier & Robinson, 2006; Song et al., 2012; Wapner et al., 1957), such as depressed symptoms can affect participants' vertical se− lective attention (Meier & Robinson, 2006) and smiling faces were judged as perceptually brighter than frowning faces (Song et al.,2012), but there was no study directly testing whether and how these physical metaphors (vertical position or brightness) affect emotion on pleased−depressed dimension. However, the idea that environmental input might affect emotion is not entirely novel. Several studies have revealed that color can temporarily influence emotional judgments (Adams & Osgood, 1973; Fetterman, Robinson, & Meier, 2012; Hevner, 1935; Keith& James, 1975; Valdez & Mehrabian, 1994). Within the field of clinical psychol ogy, several studies have shown that phototherapy (daily exposures to bright light) is effective in combating certain depressed disor− ders, such as seasonal affective disorder (Terman, Terman, & Ross,1998), antepartum depression (Oren et al., 2002), and general depression (Kripke, 1998).

Although several previous studies have shown that environmental input and bodily sensations can influence emotion, we are not aware of a study that has directly explored whether perception relevant to two important metaphors (vertical position or bright ness) of emotion can temporarily influence the emotional state (on the pleased− depressed dimension) of an observer in a way that is congruent with these metaphors. In the current study, we explored whether vertical position and brightness could affect the emotion of an observer on the pleased− depressed dimension.

2 Experiment 1

We conducted Experiment 1 to determine whether brightness could affect the emotion of aperceiver in a metaphorically consistent manner. We expected that prime condition can temporarily affect the emotion of a perceiver along the pleased−depressed dimensions.

2.1 Participants and design

The participants were 53 Chinese undergraduates (10 men) who were offered 10 RMB as compensation. Five participants (three men, two women) were excluded from the analysis, because four of them did not follow instructions when completing the emotion questionnaire and one subject reported not paying much attention to the screen. All participants (age range: 18e25, mean age: 21.5 years) were randomly assigned to the two prime conditions; each prime condition included 24 participants. We used a 2 (prime condition: white or black) independent groups design.

2.2 Stimulus materials and procedure

Participants arrived at the laboratory, individually, and were greeted by a male experimenter. Each participant was told that s/he must complete two unrelated tasks. During the first task, participants were seated facing a computer screen (Dell Computer, 19− inch display monitor, resolution: 1280 x 1024) and were told that the purpose of the task was to explore the ability in focusing one's attention on a screen. In fact, the purpose of the first task was to prime color metaphors (brightness: black or white). On each trial, a fixation cross was first presented at the center of the screen for 1000 ms. Next, the whole screen became white or black depending on the prime condition; the color was displayed for one minute. During this period, participants were required to look at the screen continuously, but are allowed to blink. After 1 min, participants rested for 5 s. Participants completed this process seven times. At the end of the experiment, participants reported the extent to which they focused on the screen on a 9−point scale (1 ="not at all" to 9= "extremely") by pressing a digit key on the keyboard.

Participants were then asked to complete an emotion (pleasededepressed dimension) questionnaire. We chose the words "happy" and "pleased" to represent the pleased dimension, and the words "depressed" and "repressed" to represent the depressed dimen sion; these words were chose according to the BFS (Abele−Brehm & Brehm, 1986), which was a scale of mood. The questionnaire contained four items: "pleased" and "happy" represented the pleased dimension, and "depressed" and "repressed" represented the depressed dimension. Participants were asked to rate how they felt presently and rated the intensity of each mood (four emotion words) separately on 9−point scale (form 1 to 9, 1 ="not at all" to 9 = "extremely"). On completion of the experiment, participants were debriefed as to their awareness of the study hypotheses, and none of the participants were able to identify the purpose of the study.

2.3 Results and discussion

Mean scores on the pleased and depressed dimensions served as the dependent measure. The data were submitted to a MANOVA. The analysis revealed, within the depressed dimension, scores on the black condition (M=3.71, SE=0.35) were significantly higher than scores on the white dimension (M=2.25, SE=0.35), [$F(1,46)$ η^2=8.64, p<0.01, η^2=0.16; see Fig. 1]. Within the pleased dimension, scores on the white dimension (M=4.98, SE=0.32) were higher than scores on the black dimension, but the difference was not significant (M=4.46, SE=0.32), [$F(1, 46)$ =1.29, p>0.05, η^2=0.03; see Fig. 1]. The correlation between items "pleased" and"happy" was significant (r=0.80, p<0.01) and the correlation between items "depressed" and "repressed" was also significant (r=0.79, p<0.01).

We also analyzed the difference of the scores on two dimensions in different prime conditions. The analysis demonstrated that in the white prime condition, scores on the pleased dimension (M=4.98, SE=0.33) were significantly higher than scores on the depressed dimension (M=2.25, SE=0.34), [$t(23)$ =4.45, p< 0.01, d=1.70]. The prior study also showed that the scores on the positive emotion items of PANAS were higher than scores on the negative emotion items of PANAS (Watson, Clark, &Tellegen, 1988). Conversely, within the black prime condition, there were no significant differ ences between scores on the pleased (M=4.46, SE=0.32) and depressed dimensions (M=3.71, SE=0.36), [$t(23)$ =1.54, p =0.14, d=1.17]. These results indicated that compared to the white prime condition, under the black prime condition
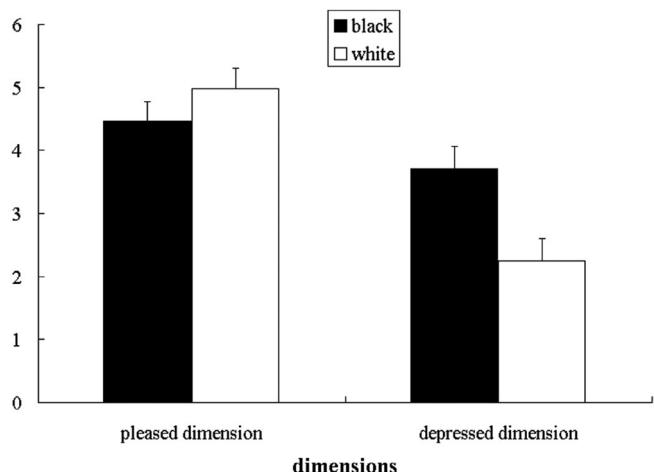
Fig. 1. Mean scores on emotion as a function of brightness and dimensions (pleased–depressed) (study 1).

the difference of scores between depressed dimension and pleased dimension decreased, that was our experimental manipulation was effective.

Results from Experiment 1 revealed that brightness affects emotion felt by a perceiver in a metaphorically consistent manner at least partly.

3 Experiment 2

In addition to brightness, vertical position is also thought to provide a metaphor of emotion along the pleased–depressed dimension. Weconducted Experiment 2 to determine if gaze directed to certain spatial positions could affect emotions along the pleaseded–presseddimension. We expected that participants would feel more pleased and less depressed temporarily if primed by stimuli presented at the top of the computer screen, but would feel the opposite if primed by stimuli presented at the bottom of the computer screen.

3.1 Participants and design

The participants were 73 Chinese undergraduates (44 women)who were offered 10 RMB as compensation. Five participants (3men, 2 women) were excluded from the analysis because they didnot follow instructions when completing the emotion questionnaire. All participants (agerange from18 to 24,mean age 21.1 years)were randomly assigned to the two priming conditions. All par ticipants got (greater than) the correct rate of 80% in the task ofdiscriminating letters from digits. Each priming condition consistedof 34 participants. We implemented a 2 (prime condition: up or down) × 2 (participant gender: male or female) (because the number of male participants was too small and gender was not the important factor we interested in present study, so wedid not makegender as a factor in Experiment 1) between subject design.

3.2 Stimulus materials and procedure

Participants arrived at the laboratory, individually, and were greeted by a male experimenter. Each participant was told that s/hemust complete two unrelated tasks. The purpose of the first task was to prime metaphors (vertical position: top or bottom), but participants were told that the purpose of the first task was to explore the ability to discriminate letters from digits. In the first task, participants were seated facing a computer screen (DellComputer,19−inch display monitor, resolution= 1280 × 1024), and the height of participants' eyes was the same as the height of the center of the screen (we

made it by adjusting the height of chair according to the height of participants). The procedure was as follows: during each trial, a fixation cross was presented at the centerof the screen for 300 ms. Following this central cue, a letter (from Ato Z but excluding "O" due to difficulty in discriminating between"O" and the digit "0") or digit (from 1 to 25) appeared either at the top or bottom of the screen, depending on the priming condition. Each letter and digit appeared 4 times, so the whole experiment contained 200 trials. Participants were required to report, by means of a key press, whether each target was a letter or digit as quickly and accurately as possible. If the response was inaccurate, the word "incorrect" appeared in red for 1000 ms at the top or bottom of screen, depending on the prime condition. A 500 ms blank screen separated accurate trials.

Next, participants were asked to complete one emotion questionnaire, which contained eight items: two pleased dimension words,"pleased" and "happy," and two depressed dimension words, "depressed" and "repressed." In addition to these four items, we added another four words, which were chosen from the PANAS and were unrelated to the pleased−depressed dimension (Watson et al., 1988). The positive items were "enthusiastic" and "proud," and the negative items were "scared" and "hostile." Participants were asked to rate how they felt presently and rated the intensity of each mood (eight emotion words) separately on 9−point scale (form1 to 9, 1="not at all" to 9="extremely"). Scores on these items were our dependent measure. On completion of the experiment, participants were debriefed as to their awareness of the study hypotheses, and none of the participants were able to identify the purpose of the study.

3.3 Results and discussion

Mean scores on the pleased and depressed dimensions served as the dependent measure. The data were submitted to a MANOVA.The analysis demonstrated that for the pleased dimension, scores in the top prime condition (M=5.83, SE=0.30) were significantly higher than scores in the bottom prime condition (M=4.99,SE=0.34), [$F(1, 64)$=3.45, p=0.06, η^2=0.05; see Fig. 2]. The main effect of gender [$F(1, 64)$=1.39, p=0.24, η^2=0.02] and the interaction between gender and prime were not significant [$F(1,64)$=0.45, p=0.51, η^2< 0.01]. Conversely, for the depressed dimension, scores in the bottom prime condition (M=3.86,SE=0.38) were significantly higher than scores in the top primecondition (M=2.86, SE=0.34) [$F(1, 64)$=3.89, p=0.05, η^2=0.06; see Fig. 2]. The main effect of gender [$F(1, 64)$=1.85, p=0.18, η^2=0.03] was not significant and the interaction between gender and prime was significant [$F(1, 64)$=4.12, p=0.05, η^2=0.06]. The correlation between items "pleased" and "happy" was significant(r=0.79, p<0.01) and the correlation between items "depressed"and "repressed" was significant (r=0.80, p< 0.01).

Scores for the other four emotion words unrelated to the pleased−depressed dimension (two positive words: enthusiastic and proud; two negative words: scared and hostile) were also analyzed with a MANOVA. The analysis demonstrated that for the positive dimension, there were no significant differences between scores in the top (M=5.29, SE=0.31) and bottom prime conditions (M=5.24, SE=0.35), [$F(1, 64)$=0.01, p=0.91, η^2< 0.01; see Fig. 2]. The main effect of gender [$F(1, 64)$ =7.34, p< 0.01, η^2= 0.10] was signiifcant, and the interaction between gender and prime conditions was not signiifcant [$F(1, 64)$=0.01, p=0.91, η^2< 0.01]. For the negative dimension, there were also no signiifcant differences between scores in the top (M=1.76, SE=0.23) and bottom prime conditions (M=1.89, SE=0.26), [$F(1, 64)$=0.14, p=0.71, η^2< 0.01; see Fig. 2]. The main effect of gender [$F(1, 64)$=0.67, p=0.42, η^2= 0.01] and the interaction between gender and prime were notsigniifcant [$F(1, 64)$=0.04, p=0.85, η^2< 0.01]. The correlation between items "enthusiastic" and "proud" was signifcant (r=0.56, p<0.01) and the correlation between

items "scared" and "hostile" was signiifcant (*r*=0.48, *p*<0.01).

Results from Experiment 2 revealed that gaze position (upward vs. downward) also affects emotions experienced along the pleasede depressed dimension. When gazing
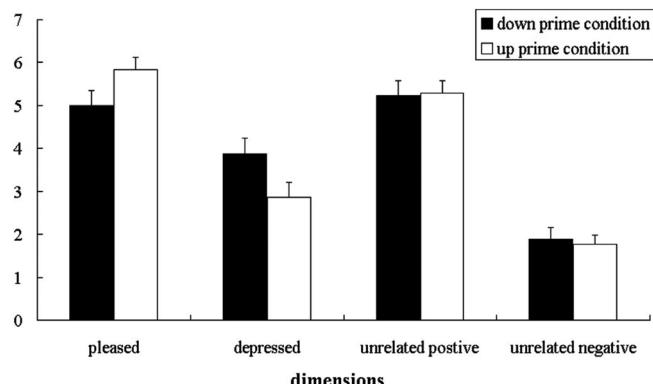
Fig. 2. Mean scores on emotion as a function of gazing vertical positions and dimensions (study 2).

at the top of the screen, participants felt more pleased and less depressed temporarily but felt more depressed and less pleased after gazing at the bottom of the screen temporarily. Most importantly, the gaze position manipulation did not affect ratings on emotional states unrelated to the pleased–depressed dimension, indicating that the prime conditions only affected emotions relevant to their metaphors.

4 General discussion

The present study was the first (to our knowledge) to show that metaphors of emotion (vertical gaze position and brightness) can affect emotional experience along the pleased–depressed dimension in a metaphorically consistent manner. Specifically, gazing at the top of the screen, or looking at a white screen, was associated with participants feeling more pleased and less depressed temporarily, while focusing attention at the bottom of a screen, or viewing a black screen, was associated with feeling more depressed and less pleased temporarily. Prior studies had suggested that emotions do affect

perception (brightness and vertical position judgments) in a metaphorically consistent manner (Meier & Robinson, 2006; Song et al., 2012). Our study has added to this literature by showing that these metaphors (vertical gaze position and brightness) also affect emotional experience along the pleased–depressed dimension. These findings are in line with previous studies suggesting that metaphors of emotion along the pleased–depressed dimension might be bidirectional (Landau et al., 2010; Williams et al.,2009). One prior study found that physical coldness could significantly increase feelings of a specific emotion: loneliness (Bargh & Shalev, 2012). Another recent study found that people use the color red as a perceptual metaphor to understand anger (Fetterman et al., 2012). Our results add to research showing that metaphors (vertical gaze position and brightness) of emotion can infiuence, at least temporarily, emotional experience along the pleased–depressed dimension of an observer.

People typically depict "pleased" and "depressed" with its corresponding metaphors as up (light) and down (darkness), respectively (Meier & Robinson, 2006). Common, everyday expressions highlight these metaphors: "I feel up today," and "A smile brightens one's face." These metaphoric mappings seem to be a universal phenomenon. We chose the pleased–depressed dimension over other positive/negative emotion, because some studies have corroborated the association between emotion (on pleasede– depressed) and its metaphors, such as spatial position and brightness (Meier & Robinson, 2006; Song et al., 2012). There were no study proved the association between the metaphors (such as spatial position and brightness) and general positive and negative emotion. The study 2 also showed that vertical gaze position did not affect ratings on emotional states unrelated to the pleased–depressed dimension, that is positive emotion items (enthusiastic and proud) and negative emotion items (scared and hostile), which were chose from PANAS.

Researchers focusing on metaphor and embodiment have increasingly suggested that the study in this area should attempt to examine mechanisms, moderators, or the implications of embodiment/metaphor in addition to illustrating effects. Our study pushed the envelope by examining the impact of metaphor and embodiment on emotional states. Emotions are important psychological phenomena, and our emotions can affect several cognitive processes and behaviors. Thus, our results have important implications for real life scenarios. Certain environments, such as brightness and the direction of our gaze,can affect our emotions along the pleased−depressed dimension. Thus, we must be mindful of how these factors can inifuence our emotions (i.e., avoid looking at black objects for long periods of time, paint our walls in lighter colors, avoid looking at objects low in vertical position etc.). Furthermore, there was very little cross−cultural work in this area, and our study firstly explored the metaphor of emotion based on the non−western participant samples. Our findings also open up some interesting avenues for future research. Additional studies should explore whether these metaphors of emotion actually inifuence behaviors. For instance, these metaphors could affect our behavior by diverting gaze away from black objects/scenes to avoid aggressive behaviors and focusing on white/light colored objects/scenes to help promote prosocial behaviors. Furthermore, it will be interesting to determine whether the inifuence of these metaphors have any sort of lasting effects on emotional experience and behavior.

References

Abele−Brehm, A., & Brehm, W. (1986). Zur Konzeptualisierung und Messung von Befindlichkeit: Die Entwicklung der "Be?ndlichkeitsskalen" (BFS) [The conceptualization and measurement of mood: The development of the "Mood Survey"]. *Diagnostica, 32*, 209e228.

Adams, F. M., & Osgood, C. E. (1973). A cross−cultural study of the affective meanings of color. *Journal of Cross−Cultural Psychology, 4*(2), 135e156.

Bargh, J. A., & Shalev, I. (2012). The substitutability of physical and social warmth in daily life. *Emotion, 12*(1), 154e162.

Crawford, L. E., Margolies, S. M., Drake, J. T., & Murphy, M. E. (2006). Affect biases memory of location: Evidence for the spatial representation of affect. *Cognition & Emotion, 20*(8), 1153e1169.

Fetterman, A. K., Robinson, M. D., & Meier, B. P. (2012). Anger as "seeing red": Re− action time evidence foran implicit perceptualassociation. *Cognition & Emotion, 1*(1), 1e14.

Fisher, S. (1964). Depressive affect and perception of upedown. *Journal of Psychiatric Research, 2*(1), 25e30.

Hevner, K. (1935). Experimental studies of the affective value of colors and lines. *Journal of Applied Psychology, 19*(4), 385e398.

Keith, W. J., & James, F. S. (1975). Effects of four psychological primary colors on anxiety state. *Perceptual and Motor Skills, 41*(1), 207e210.

Kripke, D. F. (1998). Light treatment for nonseasonal depression: Speed, ef?cacy, and combined treatment1. *Journal of Affective Disorders, 49*(2), 109e117.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the ?esh: The embodied mind and its challenges to western thought*. New York, NY: Basic Books.

Landau, M. J., Meier, B. P., & Keefer, L. A. (2010). A metaphor−enriched social cognition. *Psychological Bulletin, 136*(6), 1045e1067.

Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations be− tween affect and vertical position. *Psychological Science, 15*(4), 243e247.

Meier, B. P., & Robinson, M. D. (2006). Does "feeling down" mean seeing down? Depressive symptoms and

vertical selective attention. *Journal of Research in Personality, 40*(4), 451e461.

Meier, B. P., Robinson, M. D., & Clore, G. L. (2004). Why good guys wear white automatic inferences about stimulus valence based on brightness. *Psychological Science, 15*(2), 82–86.

Oren, D. A., Wisner, K. L., Spinelli, M., Epperson, C. N., Peindl, K. S., Terman, J. S., et al. (2002). An open trial of morning light therapy for treatment of antepartum depression. *American Journal of Psychiatry, 159*(4), 666–669.

Song, H., Vonasch, A. J., Meier, B. P., & Bargh, J. A. (2012). Brighten up: Smiles facilitate perceptual judgment of facial lightness. *Journal of Experimental Social Psychology, 48*(1), 450–452.

Teasdale, J. D. (1993). Emotion and two kinds of meaning: Cognitive therapy and applied cognitive science. *Behaviour Research and Therapy, 31*(4), 339–354.

Terman, M., Terman, J. S., & Ross, D. C. (1998). A controlled trial of timed bright light and negative air ionization fortreatment of winter depression. *Archives of General Psychiatry, 55*(10), 875–882.

Thayer, R. E. (2003). *Ca1m energy: How people regul α te mood with food and exercise.* New York, N Y: Oxford University Press.

Va1dez P.&MehrabianA.(1994). Effects of co10r on emotions. *journal of Experimental Psycholy: General, 123*(4), 394–409.

Wapner, S., Werner, H., & Krus, D. M. (1957). The etfect of success and failure on space 1ocalization. *journal of Personality, 25*(6), 752–756.

Watson, D., Clark, L. A., & Tellegen, A.(1988). Development and validation of brief measures of positive and negative affect: The PANA5 sca1es. *Journal of Personality and Social Psychology, 54,* 1063–1070.

Williams, L. E., Huang, J. Y., & Bargh, J. A. (2009). The scaffo1ded mind: Higher menta1 processes are grounded in ear1y experience of the physica1 wor1d. *European journal of Social Psychology, 39*(7), 1257–1267.

发展心理研究所团队成员

周宗奎，博士，教授，心理学院院长，青少年网络心理与行为教育部重点实验室主任，主要研究领域为人格与社会性发展、青少年网络心理与行为研究、儿童青少年心理健康教育等。

谷传华，博士，教授，主要研究领域为创造性发展与培养，社会性与人格发展。

范翠英，博士，教授，发展心理研究所所长，主要研究领域为儿童青少年心理发展与心理健康教育、青少年网络欺负、网络道德心理研究等。

莫书亮，博士，副教授，主要研究领域为儿童和青少年社会认知发展，异常儿童社会认知发展及干预。

孙晓军，博士，副教授，发展心理研究所副所长，主要研究领域为人格与社会性发展、青少年网络交往行为、留守儿童的心理问题研究等。

田媛，博士，讲师，主要研究领域为网络心理与行为、媒体学习评测、青少年公正价值观等问题。

刘勤学，博士，讲师，主要研究领域为儿童青少年问题行为与健康发展、网络心理与行为研究、亲密关系与家庭治疗。

孔繁昌,博士,讲师,主要研究领域为社会性与人格发展、自我与健康、网络使用与青少年发展。

李董平，博士，讲师，主要研究领域为处境不利儿童青少年的心理发展，青少年问题行为研究。

童年中期同伴关系与孤独感的中介变量检验 *

周宗奎 [1,2]，孙晓军 [1]，赵冬梅 [1]，Hsueh Yeh [3]

([1] 华中师范大学心理学院 , 武汉 430079)

([2] 首都师范大学学习与认知实验室 , 北京 100089)

([3] Co llege of Education, University of Memphis)

摘　要　以 571 名小学三、四、五、六年级的儿童为被试 , 考察了儿童社会喜好、友谊质量、社交自我知觉与孤独感的关系 , 检验了社交自我知觉在同伴关系变量与孤独感间的中介作用。结果表明 , 社会喜好、友谊质量、社交自我知觉和孤独感间相关显著 , 并且存在显著的性别差异 ; 社交自我知觉在同伴关系变量与孤独感间存在中介的作用 ; 独立的中介效应检验中 , 社会喜好、友谊质量均通过社交自我知觉的中介作用与孤独感发生联系 , 同时 , 也存在直接的联系 ; 综合模型中 , 社会喜好只通过社交自我知觉的中介作用与孤独感产生联系 , 不存在直接效应 , 而友谊质量与孤独感既存在中介的联系 , 同时也存在直接联系。

关键词　社会喜好；友谊质量；社交自我知觉；孤独感；中介作用。

分类号　B844

1　问题提出

孤独是一种消极的、弥漫性的心理状态 , 儿童长期处于此状态会导致适应不良。关于孤独感的研究中 , 有两个主要的理论 , 即社会需要理论和认知加工观。社会需要理论认为 , 人生来就有与人保持交往的需要 , 除非人际交往需要得到满足 , 否则就会产生孤独感。而认知加工理论认为 , 孤独感的产生不是因为人类固有的社会交往需要得不到满足 , 而是当一个人对觉知到的人际关系不满意时 , 孤独感才会产生。换句话说 , 当一个人意识到他想要或希望的人际交往关系与实际现状之间有差距时 , 孤独感才会产生。

在众多对儿童孤独感的研究中 , 研究者普遍发现 , 社会技能、问题行为、人格特征、家庭环境等因素都与儿童的孤独感有关 , 而同伴关系则一直是研究者在研究孤独感时重点考察的一个因素。同伴关系作为同龄人之间或心理发展水平相当的个体间在交往过程中建立和发展起来的一种人际关系 , 可以分为四个水平：个体特征水平、人际交互水平、双向关系水平和群体水平。本研究选取的社交自我知觉、友谊质量和同伴接纳就分别处于同伴交往经验的个体水平、双向关系水平和群体水平。

Bush 和 Ladd 的研究表明 , 被拒绝儿童经历了

同伴较多的消极对待,更可能表现出较高的孤独感;Asher 等人的研究则发现,3~6 年级不受欢迎的儿童(被忽视型和被拒绝型儿童)报告了显著高于受欢迎儿童的孤独感;俞国良等人的研究也发现,儿童的同伴接受性与孤独感有显著的负相关。大量的研究都证实了同伴接纳性与孤独感间的密切联系。儿童的社交地位越不利,同伴接纳性越低,其体验到的孤独感就越强。

同时,作为同伴关系另一水平的指标,友谊关系与孤独感间也存在着较强的联系。Hodges 等人的研究发现,友谊质量对儿童的孤独感具有显著的预测作用;受欺负儿童所受到的伤害以及体验到的孤独感会因为拥有一个支持性的朋友而得以减轻;Demir 的研究也发现,对于与异性或同性的友谊不满的青少年报告的孤独感高于感到满意的青少年,并且亲密朋友的数量越多孤独感水平就越低。

而国内研究中,同时从同伴接纳性和友谊关系这两个维度来探讨同伴关系与孤独感的研究较少。实际上这两个指标的具体内涵是有区别的,同伴接纳性考察的是群体水平上的同伴关系,而友谊关系重点强调的是同伴间双向水平的同伴关系。有研究也指出了同伴接纳与友谊在儿童青少年的发展中具有不同的功能。因此,有必要综合这两个维度探讨它们对孤独感的影响,这是本研究所关注的一个方面。

另一方面,Asher 指出,孤独是个体对自己社交状况的一种主观体验。人本主义心理学者也认为当一个人的社会关系网络的数量和质量低于他的期望时,孤独感就产生了。而同伴关系作为一种客观的社交地位,在预测主观孤独感时可能存在偏差。例如,有研究发现,一些受欢迎儿童报告了极高水平的孤独感,而一些被拒绝型儿童却报告了极低水平的孤独感。这种现象一方面印证了孤独是一种主观体验,另一方面说明,只从客观的同伴关系角度来考察孤独感,不能充分解释同一社交地位群体内部的个体差异。

基于上述原因,一些研究者加入了认知因素来研究儿童的孤独感。社交自我知觉作为个体对自身社交地位的主观评价,引起了研究者的广泛兴趣,将客观的社交地位与主观的社交自我知觉结合起来,考察二者对孤独感的影响成为了孤独感研究新的取向。如周宗奎等人在考察二者对孤独感的研究中发现,比起客观社交地位,主观的社交自我知觉对于孤独感具有更强的预测力。在本研究中,我们就试图综合客观的同伴关系和主观的社交自我知觉来考察它们对于孤独感的影响。同时,Hymel 等人曾从社会认知的角度解释孤独感与社会地位的关系,他们认为,儿童的孤独感与儿童在同伴中的实际社交地位之间是以社会认知过程为中介的,个人的人际关系知觉水平是重要的中介变量之一。因此,本研究中,我们假设客观的同伴关系除了与孤独感具有直接联系外,还可以通过主观社交自我知觉的中介作用与孤独感产生间接联系,即主观的社交自我知觉在同伴关系和孤独感间存在部分中介作用。

综上所述,本研究中,我们将综合同伴接纳性和友谊质量这两个指标考察同伴关系对孤独感的影响,在此基础上,检验儿童社交自我知觉在此过程中的部分中介作用。

2 研究方法

2.1 被试的选定

武汉市一所小学的三、四、六年级各两个班,五年级三个班,共 9 个班 580 人。回收有效问卷 571 份,回收率为 98.45%。其中,男生 310 人,女生 261 人;三年级 123 人,四年级 122 人,五年级 200 人,六年级 126 人,各年级男女生分布具体情况如表 1 所示。三、四、五、六年级学生的平均年龄分别为 9.16、10.14、11.11 和 12.12。

2.2 研究工具

2.2.1 同伴提名 给儿童提供一份班级名单表,

表 1 各年级男女生人数分布表

被试	三年级	四年级	五年级	六年级
男生	69	70	100	71
女生	54	52	100	55
总计	123	122	200	126

注:表内各数据均表示被试的人数

要求他们选出自己在班内最喜欢的 3 个同学和最不喜欢的 3 个同学。然后，将每个学生所获得的最喜欢和最不喜欢的提名数除以班级的总人数，分别得到积极提名和消极提名的比例，二者之差表示社会喜好 (sp)，即受欢迎程度，作为被试同伴接纳性的指标，这种记分方法得到了研究者的认可，具有较高的效度。

2.2.2 友谊质量问卷 采用《友谊质量问卷》(Parker & Asher, 1993) 的简表，共 18 个项目，原量表有 40 个项目。包括肯定与关心、帮助与指导、陪伴与娱乐、亲密袒露与交流、冲突解决策略、冲突与背叛这六个友谊维度。Cronbach's α 为 0.76。

2.2.3 儿童自我知觉 (PCSC) 量表 PCSC 量表 (The Perceived Competence Scale for Children) 是 Harter(1982) 编制的儿童自我知觉问卷，原问卷包含的四个维度分别是：社交自我知觉、认知自我知觉、运动技能自我知觉和一般自我知觉，本研究只选用社交自我知觉这一维度，Cronbach's α 为 0.78。该量表给被试同时呈现两个描述性的句子 (如：一些孩子觉得交朋友很困难 vs. 另外一些孩子觉得交朋友很容易)，首先要求被试确定他 / 她更符合哪一句的描述，然后再确定他 / 她是有点符合该描述还是完全符合，分别记为 1~4 分。计算该维度所有项目总分的平均分，得到儿童的社交自我知觉分。

2.2.4 儿童孤独量表 采用 Asher 等人 1984 年编制的专用于 3~6 年级学生的儿童孤独量表 (Chilren's Loneliness Scale)，该量表包括 16 个孤独项目 (10 条指向孤独，6 条指向非孤独) 和 8 个关于个人爱好的插入项目 (为使被试在回答时放松一些)，因子分析表明插入项目与负荷于单一因子上的 16 个孤独条目无关，16 个孤独项目的 Cronbach's alpha 为 0.92。计算 16 个项目的平均分 (反向记分的题目先要进行转换)，得到儿童的孤独感得分，得分越高，表示孤独感越强，这一记分方法在国内外都被广泛采用。

2.3 数据收集与分析

由经过培训后的心理学专业研究生主持，采用团体施测的方式进行。施测时以班级为单位，由主试讲明要求，解释指导语，必要时给予个人指导以确保被试正确理解问卷。所有数据于 2003 年 6 月收集完毕，全部数据由 Filemaker 4.0 录入，利用 SPSS11.5 和 LISREL 8.30 软件进行统计处理，主要采用多元方差分析和结构方程模型等统计方法。

3 研究结果与分析

3.1 描述性统计分析结果

本研究中，所考察的各变量间的相关如表 2 所示，由表 2 可知，除了友谊质量的冲突、背叛维度与社会喜好和社交自我知觉相关不显著以外，各变量间均存在极其显著的相关，这就满足了我们进行中介效应检验的前提条件；同时，各问卷的 Cronbach's α 均大于推荐值 (0.70)，表明信度较高。

3.2 性别、班级的差异检验

本研究中，由于各因变量 (社会喜好、友谊质量

表 2 变量相关矩阵 (N=571)

	M	SD	1 社会喜好	2 肯定与关心	3 帮助与指导	4 陪伴与娱乐	5 亲密袒露	6 冲突解决	7 冲突背叛	8 社交知觉	9 孤独
1	0.003	0.098									
2	2.520	0.952	0.124**								
3	2.551	1.052	0.137**	0.459***	(0.76)						
4	3.148	0.902	0.219***	0.436***	0.450***						
5	2.615	1.080	0.129**	0.461***	0.545***	0.417***					
6	2.991	0.991	0.134**	0.371***	0.429***	0.383***	0.391***				
7	4.164	0.991	0.036	0.129**	0.180**	0.150**	0.146**	0.216**			
8	2.760	0.428	0.133**	0.185***	0.107*	0.179***	0.113**	0.146**	−0.036	(0.78)	
9	1.876	0.769	−0.278**	−0.360***	−0.288***	−0.296***	−0.284***	−0.300**	−0.134**	−0.510***	(0.92)

注: *p<0.05 **p<0.01; ***p<0.001; ( ) 内数据为各量表的内部一致性系数其中，(0.176) 是友谊质量整个量表的内部一致性系数。

各维度、社交自我知觉、孤独感)之间存在显著相关,所以不应通过多次方差分析来对班级的差异进行检验,这一过程应通过 MANOVA 进行多元方差分析来完成。多元方差分析的检验统计量通常用 Wilks 的 Λ,得到的是精确的 F 值。班级差异的多元方差分析结果表明,班级的主效应不显著($\Lambda=0.965$, $F(8, 550) = 2.243$, $p= 0.058$, ns)。

T 检验的结果表明,除了友谊质量的冲突与解决维度和社交自我知觉以外,在所有其它因变量上均存在显著的性别差异。在社会喜好($M_{男}=-0.011$,$M_{女}=0.02$, $t(569) =-3.387$, $p<0.001$)、友谊质量的肯定与关心($M_{男}=2.417$, $M_{女}=2.641$, $t(569)=-2.818$, $p<0.01$)、帮助与指导($M_{男}=2.370$, $M_{女}=2.766$, $t(569) =-4.56$, $p<0.001$)、陪伴与娱乐($M_{男}=3.053$, $M_{女}=3.261$, $t(569) =-2.759$, $p<0.01$)、亲密与袒露($M_{男}=2.361$, $M_{女} = 2.917$, $t(569)=-6.332$, $p<0.001$)、冲突与背叛($M_{男}=4.051$, $M_{女}=4.298$, $t(569) =-3.256$, $p< 0.001$)维度上,男生的得分均显著低于女生,而在孤独感的得分上,则是男生显著高于女生($M_{男} = 1.959$, $M_{女} = 1.775$, $t(562)= 2.836$, $p<0.01$),表现出更高的孤独感体验。

3.3 中介效应的检验

如果自变量 X 通过影响变量 M 来影响 Y,则我们就称 M 为中介变量。本研究中,我们假设社交自我知觉是同伴关系与孤独感间的中介变量(如图 1 所示),图中 a、b、c、c' 均表示相应的标准回归系数。

研究者对中介效应进行检验时较多地运用了 3 种方法:Sobel 检验($Z =ab / \sqrt{b^2S^2a + a^2_b}$)、Goodman I 检验($Z=ab/ \sqrt{b^2S_b^2+a^2_{sb}+S^2_aS_b^2}$)和 Goodman II 检验($Z=ab/ \sqrt{b^2S_b^2+a_s^2{}_b^2-S^2aS_b^2}$)。本研究中,样本容量较大($N$=571),因此,3 种方法的检验功效差别不大。

通过结构方程模型分析,求得对于社会喜好而言,a、b 值为 0.36 和 -0.76, 对应的标准误分别为 0.06 和 0.09, c' 为 -0.10(χ^2=105.50, df=33; RM-SEA = 0.062; SRMR = 0.043; GFI= 0.96; IFI = 0.94;CFI 0.94; NNFI =0.92); 相应的, 对于友谊质量, a、b 值分别为 0.47 和 -0.51, 标准误分别为 0.17 和 0.22, c' 为 -0.13(χ^2= 220.24, df=75; RMSEA =0.051; SRMR = 0.054; GFI=0.96; IFI = 0.94; CFI =0.96; NNFI=0.93)。将这些结果分别代入上述三个公式,各方法的检验结果见表 3。

由表 3 可知,各种检验的的结果都表明了中介变量的显著作用。即社会喜好或友谊质量都可以通过社交自我知觉的中介作用与孤独感发生间接联系,其中,对于社会喜好来说,中介效应为 -0.273(a × b, 下
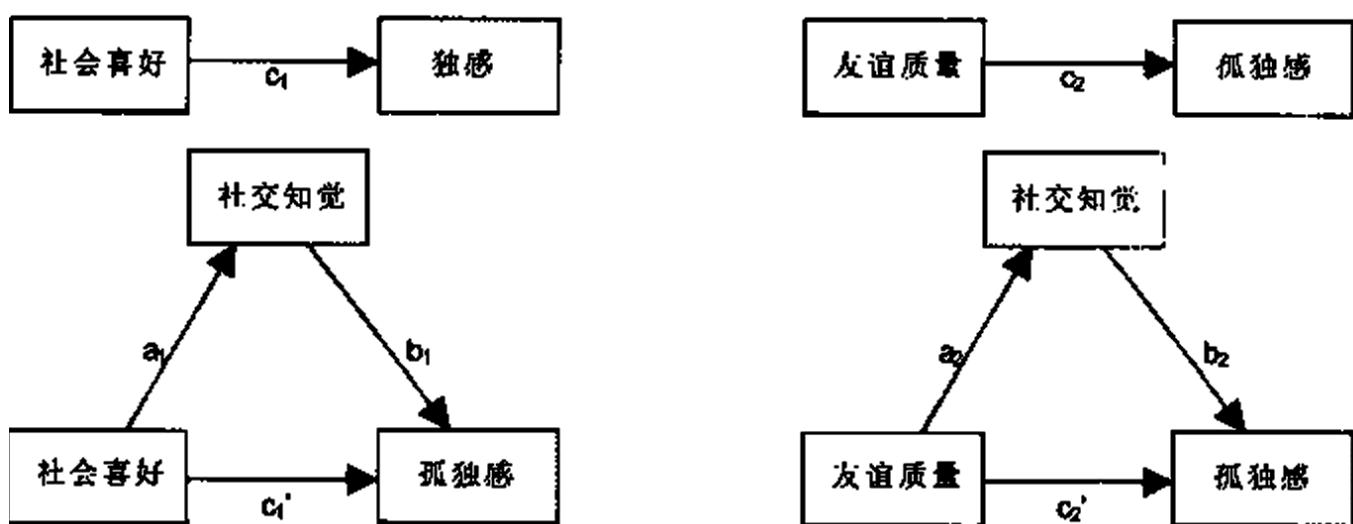
图 1 社交自我知觉在同伴关系与孤独感间的中介作用模式图

表 3 社交自我知觉在同伴关系和孤独感之间的中介作用检验

中介作用的路径	社会喜好 – 社交自我知觉 – 孤独感	友谊质量 – 社交自我知觉 – 孤独感
a(S_a)	0.36（0.06）	0.47（0.07）
b(S_b)	−0.76（0.09）	−0.51（0.12）
Sobe 脸验（Z）	4.891***	3.591***
Goodman I 检验(Z)	4.868***	3.563***
Goodman II 检验（Z）	4.914***	3.620'**

注：*p< 0.05; **p< 0.01; *** p<0.001。

同），总效应为 − 0.373(a ×b +c'，下同），中介效应与总效应的比值为 0.732；而对于友谊质量，其中介效应为 −0.240，总效应为 − 0.37，中介效应与总效应的比值为 0.649，由于中介效应的相对作用均较大（分别为 0.732 和 0.649)，因此，这一中介作用的发现更具有重要意义。

上述 3 种检验证实了社交自我知觉在同伴关系与孤独感间的中介作用的存在，同时，结构方程模型的结果表明，无论对于社会喜好或友谊质量，c' 均达到显著性水平，因此，我们可以进一步判断，社交自我知觉在同伴关系与孤独感间存在部分中介的作用。

综合上述分析的结果，我们构建了如图 2 所示的模型，结构方程模型分析表明，该模型能较好地拟合数据（χ2= 230.34, df=99; RMSEA=0.048; SRMR= 0.045; GFI=0.95; IFI=0.94; CFI=0.94; NNFI =0.93)。

结果表明，友谊质量和孤独感间存在直接联系，直接效应值为 −0.17，同时，友谊质量也可以通过社交自我知觉的中介作用与孤独感产生联系，其中介效应值为 − 0.238(即从友谊质量到孤独感的间接作用路径上的路径系数乘积），总效应值为 −0.408，中介效应与总效应的比值为 0.583；而社会喜好和孤独感间的联系主要是通过社交自我知觉的中介作用来实现的，其中介效应的值为 −0.182，直接效应不显著 (− 0.07)，因此，是一种完全中介的作用，这一结果与前面证明的部分中介作用是不一致的。

4 分析与讨论

4.1 社会喜好、友谊质量、社交自我知觉和孤独感的性别、班级差异

本研究表明，社会喜好、友谊质量、社交自我知觉和孤独感不存在班级差异，但社会喜好、友谊质量（除冲突与解决维度外）和孤独感的性别差异显著。无论是社会喜好或友谊质量，男生的得分均显著低于女生，这一结果与以往的研究结果是一致的。男生往往表现出较多的问题行为，这种问题行为反应在自身的社交行为和策略上就很容易导致其较差的同伴关系，而女生由于社会环境、文化等因素的影响，往往表现得更为收敛，更愿意建立亲密的同伴关系网络，因此，她们的同伴关系自然就比男生要好。而对于孤独感的性别差异，本研究发现，男生的孤独感要显著高于女生，这一结果与我们前期的一项研究的结果是一致的，也与以往的部分研究结果一致。中国的传统文化要求男生更为独立、自强，不鼓励他们表露自己的情绪；而社会化的过程使女生富于乐群性，她们具有更强的亲和动机，更重视建立亲密的同伴关系，也得到更多的同伴支持，因此，女生体验到的孤独感就比较低。但也有研究表明，儿童的孤独感并不存在性别差
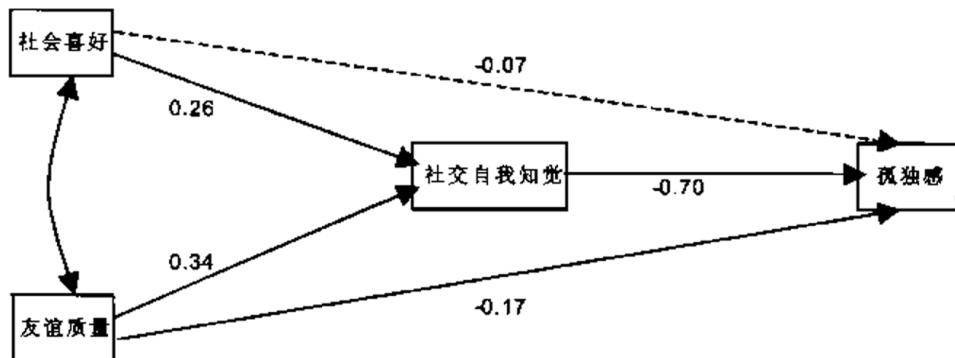
图 2 同伴关系、社交自我知觉与孤独感的关系

异。这可能是由于不同研究中分析的变量不同造成的，我们不能简单地通过这种比较来推断结论，儿童的孤独感是否存在性别差异？为什么存在这些差异？这些都是我们今后的研究中有待解决的问题。

4.2 中介效应的检验

从相关分析的结果来看，本研究中考察的各变量间（同伴接纳、友谊质量、社交自我知觉、孤独感）具有极其显著的相关，其中，同伴接纳、积极友谊质量（友谊质量的前五个维度）和社交自我知觉间存在正相关，表明儿童同伴接纳性越高，其积极友谊质量和社交自我知觉相应也较高，反之，儿童同伴接纳性越低，则其积极友谊质量和社交自我知觉也较低，儿童主观的社交自我知觉是以客观的同伴关系（同伴接纳、友谊质量）为基础的；同时，同伴接纳、积极友谊质量、社交自我知觉三变量与孤独感均存在显著的负相关，表明它们呈互为消长的关系，即同伴接纳性越高、积极友谊质量越好、社交自我知觉越强，则其体验到的孤独感就越低，反之，其体验到的孤独感就越强，这一结论与大多数研究的结果是一致的。

独立的中介效应检验结果表明，社交自我知觉在同伴关系（同伴接纳、友谊质量）与孤独感间起着中介作用，并且是部分中介的作用。由于中介效应的相对作用均较大（分别为 0.732 和 0.649），因此，这一中介作用的发现就更具有实际意义。无论是同伴接纳性或友谊质量，通过社交自我知觉的中介作用对孤独感产生影响的模式是一致的，即较好的同伴关系（同伴接纳性较强或友谊质量较好）背景下，儿童往往具有较高的社交自我知觉水平（路径系数为正数，二者分别为 0.36 和 0.47），并且儿童的孤独感体验也较低（路径系数为负数，二者分别为 – 0.76 和 – 0.51），这一结果与大多数研究的结论也是一致的。社交自我知觉作为个体对自身社交状况的评价或认知，是直接以其客观的同伴关系状况为基础的，本研究也证实了两者间显著的正相关关系，某一客观的同伴关系状况往往会导致个体产生相应的社交自我知觉水平；而另一方面，研究者发现，主观的社交自我知觉对于孤独感具有较强的预测力，儿童的社交自我知觉水平越高，其

孤独感就越低，二者间存在显著的负相关。因此，我们可以推断，儿童的同伴关系能正向预测其社交自我知觉水平，而社交自我知觉的水平又能负向预测其体验到的孤独感强度，本研究的结果证实了这一推断。

图 2 所构建的模型表明，社会喜好和友谊质量共同对儿童的孤独感产生影响。值得注意的是，从模型中各条路径系数的值来看，我们可以发现，当综合社会喜好和友谊质量来考察同伴关系对孤独感的影响时，同伴接纳性较低的儿童体验到的孤独感会因为较高的友谊质量而降低；相似地，友谊质量较低的儿童体验到的孤独感也会由于其较高的同伴接纳性而降低。Shaffer 指出，拥有一个或多个亲密的朋友可以为儿童提供一个情感上的安全网络，这种安全感可以帮助儿童更积极地迎接新的挑战，而且可以帮助儿童承受所面临的压力（如父母离异、同伴拒绝等）；一些研究则发现，受欺负儿童所受到的伤害以及体验到的孤独感会因为拥有一个支持性的朋友而得以减轻。因此，如果儿童同伴接纳性较低并且友谊质量也较差，那么从整个模型来看，他体验到的孤独感就会更加强烈，对于这类儿童，心理学工作者也应给予更多的关注。

值得注意的是，在综合社会喜好和友谊质量来共同考察社交自我知觉的中介作用时，我们发现了一些不一致的结果。独立的中介效应检验中，社交自我知觉在社会喜好和孤独感间起部分中介的作用，而在综合模型中，社会喜好对孤独感的直接效应并不显著，社交自我知觉起到的是完全中介的作用。造成这一结果的主要原因可能是友谊质量与社会喜好间存在较强的共线性，分别考察时并没有考虑到它们的共线性，而在综合模型中，由于共线性的存在，可能相互间产生某种抑制作用，从而造成社会喜好对孤独感直接效应的不显著；另一方面，更重要的是，心理学研究中一直都强调整体不等于部分之和，这个结果也提醒了研究者在研究中应特别注意研究结论的适用范围。

4.3 中介作用的意义

本研究中，我们证实了作为认知因素的社交自我知觉，在同伴关系与主观的情绪体验（孤独感）间确实存在中介作用。实际上，研究者一直很关注认知成

分在同伴关系和情绪体验间的中介作用，也做过一定的研究，结果基本都证实了这一中介作用的存在。例如，Boivin 等人的研究表明，消极的同伴关系对儿童抑郁的影响依赖于其体验到的孤独感及其对社交环境的认知；有研究者也提出了相似的结论，认为同伴拒绝与情绪体验（孤独、抑郁）间的联系是复杂的，有赖于儿童对他的社交情境的知觉；Valas 和 Sletta 则认为，儿童对待自己的社交情境的知觉有助于解释社会行为和同伴拒绝对其内在情绪体验的作用机制等等。这些研究中，共同之处就是认为认知成分在儿童的同伴关系和情绪体验间存在中介作用，本研究的结果与这些研究的结论也是一致的。

　　认知成分在同伴关系和情绪体验间的中介作用的发现具有一定的实际意义。一方面，它提示我们，同伴关系对情绪体验的影响是复杂的，并不是简单的一一对应的关系，会受到其它因素（例如认知因素）的影响；另一方面，更为重要的是，这种中介作用的发现将有助于对儿童消极情绪体验的干预，可以从认知层面入手，通过改善儿童的社交认知从而有效地预防其消极的情绪体验。总之，这一中介作用的发现，无论是对于该领域理论的丰富，或是对于实践工作的开展，都具有重要的意义。

　　需要强调的一点是，本研究中所考察的四个变量（社会喜好、友谊质量、社交自我知觉、孤独感）间可能存在更为复杂的关系，它们之间可能存在与我们的假设完全相反的关系，具体来说，有可能孤独感导致儿童社交自我知觉较低，进一步造成其同伴关系较差等等。本研究中，我们的模型是依据已有理论的支持来构建的，前面我们已经论述过。作为相关研究来说，本研究并不能从严格意义上确定变量间的因果关系，只是提供了这四个变量间可能存在的某种关系，为研究者提供一种选择，严格的因果关系必须通过实验研究才能获取。

5　结论

　　本研究的结果表明，社交自我知觉在同伴关系和孤独感间存在中介的作用，当单独考察社会喜好或友谊质量与孤独感间的关系时，社交自我自觉起部分中介的作用；而综合考察同伴关系和孤独感间的联系时，友谊质量主要通过社交自我知觉的中介作用与孤独感产生联系，同时也与孤独感间存在直接的联系；而社会喜好与孤独感间的联系则是通过社交自我知觉的中介作用来实现的，不存在直接效应。

参考文献

Asher S R, Hyme lS, Renshaw P D. (1984). Loneliness in children. *Child Development, 55*, 1456–1464

Asher S R, Wheeler V A.(1985). Children's loneliness: A comparison of rejected and neglected peer status . *Journal of Consulting and Clinical Psychology, 53*, 500–505

BoivinM, Hymel S, BukowskiW M.(1995). The roles of social withdrawal, peer rejection, and victimization by peers in predicting loneliness and depressed mood in childhood. *Development and Psychopathology, 7, 765–785*

Bush E S, Ladd G W.(2001). Peer rejection as an antecedent of Yong Children's school Adjustment: An Examination of Mediating Processes .*Developmental Psychology, 37*, 550–560

Gauze C, Bukowski W M, Aquan –Assee J, et al .(1996). Interactions between family environment and friendship and associations with well–being during early adolescence. *Child Development, 67*, 2201– 2216

Hodge EV E, BoivinM, Vitaro F, et al .(1999). The power of friendship: Protection against an esca lating cycle of peer victimization. *Developmental Psychology, 35*, 94–104

Hymel S, VaillancourtT, McDougall P,et al.(2004). Peer acceptance and rejection in childhood. In:Smith P K, HartCH. (Ed.) *Blackwell Handbook of Childhood Social Development* ,65–284

Lau S, DennisW K, Patrick C, et al.(1999). Facets of Lone liness and Depression Among Chinese Children and Adolescents . *The Journal of Social Psychology, 139,* 713–725

Margalit M.(1994). Loneliness among children with special need. *New York: Springer–Verlag*, 15 ~ 17

Parker JG, A sherS R.(1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology, 29*(4), 611–621

Peplau L A, Perl manD.(1984). Loneliness research: A survey of empirical findings. In:L A Peplau, S E Goldston (Eds) .Preventing the harmful consequences of severe and persistent loneliness. Rockville, MD: National Institute of Mental Health, 13–47

Rotenberg K J, Hymel S.(1999). *Loneliness in Childhood and Adolescence*. UK: Cambridge University Press,

Rotenberg K J, Hymel S.(1999). *Loneliness in Childhood and Adolescence*. Cambridge:Cambridge University Press,

RubinK H, BukowskiW, Parker JG. (1998). Peer Interactions , Relationships, and Groups. In W Damon, E Nancy. Handbook of Child Psychology. Vol. 3:Social , Emotional, and Personality Development . Sth ed. New York: W iley, 619–700

Schwatz D, Dodge K A, PettiG S, et al .(2000). Friendship as a moderating factor in the pathway between early harsh home environment and later victimization in the peer group. *Developmental Psychology, 36*, 646–662

ShafferD R. (2002). Developmental Psychology: childhood and adolescence(6th ed.). *Thomson learning*, 575–617

Uruk A C, Demir A.(2003). The ro le of peers and families in predicting the loneliness level of adolescents . *The Journal of Psychology, 137* (2), 179–194

ValasH, Sletta O. (1996). Social behavior, peer relations , loneliness and self–perceptions in middle school

children: A mediational model .Paper presented at the Ⅺ Ⅴ thbiennial meeting of the International Society for the Study of Behavioral Development, Quebec City, Canada, WarmanD M, Cohen R.(2000) Stability of Aggressive Behaviors and Children's Peer Relationships.*Social Development, 26*, 277–290

Wang Z Y, Wang J S, Chen H C. (2000). The influence of training on rejected and neglected children (in Chinese) . Psychological Development and Education, 16 (1), 6–11.

[王争艳 , 王京生 , 陈会昌 .(2000). 促进被拒绝和被忽视幼儿的同伴交往的三种训练方法 .*心理发展与教育*, 16 (1), 6–11.]

Wen Z L, Chang L, HauK T, et al .(2004). Testing and application of the mediating effects (in Chinese) . *Acta Psychologica Sinica, 36* (5), 614–620.

[温忠麟 , 张雷 , 侯杰泰等 . (2004). 中介效应检验程序及其应用 .*心理学报 , 36* (5), 614–620.]

Xin Z Q, Chi L P. (2003). The relationship between family functioning and children' s loneliness: the role of mediator (in Chinese) . *Acta Psychologica Sinica, 35* (2), 216–221.

[辛自强 , 池丽萍 . (2003). 家庭功能与儿童孤独感的关系 : 中介的作用 .*心理学报 , 35* (2), 216–221.]

Yu G L, Xin Z Q, Luo X L. (2003). The characteristics of loneliness, peer acceptance and their relation to family functioning (in Chinese) *Acta Psycho logica Sinica, 32* (1), 59–64.

[俞国良 , 辛自强 , 罗晓路 . (2000). 学习不良儿童孤独感、同伴接受性的特点及其与家庭功能的关系 .*心理学报 , 32*(1), 59–64.]

Zhou Z K. *Children' s Social Skill (in Chinese)*. (2002). Wuhan: Central China Normal University Press.

[周宗奎 . 儿童的社会技能 . 武汉 : 华中师范大学出版社 ,2002.]

Zhou Z K, Fan C Y.(2001). Primary school children' s social anxiety and loneliness (in Chinese) . *Psycho*

logical Science, 24 (4), 442–442.

[周宗奎 , 范翠英 . (2001). 小学儿童社交焦虑与孤独感研究 . *心理科学 , 24* (4), 442–442.]

Zhou Z K, Zhao D M, Chen J, et al .(2003). Lone liness as a fuction of sociometric status and self – perceived social competence in middle childhood (in Chinese) . *Psychological Development and Education, 19* (4), 70– 74.

[周宗奎 , 赵冬梅 , 陈晶等 .(2003). 童年中期儿童社交地位、 社交自我知觉与孤独感的关系研究 . *心理发展与教育 , 19*(4), 70–74.]

Zou H.(1993). Children＇s loneliness and peer relationships . *Psychological Development and Education, 9* (2), 16–18

[邹泓 .(1993). 儿童的孤独感与同伴关系 . *心理发展与教育 , 9* (2), 16–18[

Zou H. (1998). Developmental function and influencing factors of peer relationship (in Chinese) . *Psychological Development and Education, 14 (2)*, 40–46 .

[邹泓 . 同伴关系的发展功能及其影响因素 . *心理发展与教育 , 14* (2), 40–46]

The test of the mediator variable between peerrelationship and loneliness in middle childhood

Zhou Zongkui[1, 2], Sun Xiaojun[1] , Zhao Dongmei [1] , Hsueh Yeh [3]

([1] School of Psychology, Central China Normal University, Wuhan 430079)

([2] Learning & Cognition Lab, Capital Normal University, Beijing 100089)

([3] College of Education, University of Memphis)

Abstract 571 elementary school children from the third grade to the fifth are investigated in June 2003. The relationships among social preference, friendship quality, self-perceived social competence and loneliness are examined, the mediator effect of self-perceived social competence between peer relationship and loneliness is also tested. The results indicated that, the relationships among social preference, friendship quality, self-perceived social competence and loneliness are significant, and the gender differences is significant;self-perceived social competence has mediator effec tbetween peer relationship and loneliness;in separate analyses, there are not only indirect relation between loneliness and socia lpreference or friendship quality through the mediator effect of self-perceived social competence, but also direct relation;in integrate analysis, there are direct and indirect relations between loneliness and friendship, there are only indirect relation between loneliness and social preference.

Keywords　social preference; friendship quality; self-perceived social competence; loneliness; mediator effect

儿童的同伴交往与孤独感：一项 2 年纵向研究 *

周宗奎，赵冬梅，孙晓军，定险峰

(华中师范大学心理学院，武汉 430079)

摘　要　用提名法和问卷法对小学 3、4 年级 274 名儿童进行两年追踪调查，采用交叉滞后设计，考察了同伴交往 4 个特征水平上的变量与孤独感的相互预测关系。结果发现：(1) 自我报告的同伴交往变量——友谊质量和社交自我知觉与孤独感之间的交叉滞后效应显著；而同伴评定的同伴交往变量——积极提名、消极提名、受同伴欺负、消极退缩与孤独感之间的交叉滞后效应不显著；另外，前测的互选朋友数能显著负向预测后测的孤独感，前测的孤独感不能显著预测后测的互选朋友数。(2) 在排除了早期其它变量的效应之后，积极提名分、消极提名分、受同伴欺负得分、消极退缩得分、互选朋友数在两年之间仍然呈现出高度的稳定性；友谊质量和社交自我知觉以及孤独感则表现出中等程度的稳定性。

关键词　同伴交往；孤独感；交叉滞后回归。

分类号 B844

1　问题提出

儿童同伴关系领域的研究最早可追溯到 20 世纪 30 年代，例如，皮亚杰在他的早期着作中曾经论述了同伴关系在社会能力发展中的作用。60、70 年代来自灵长类动物的实验研究和人类的相关研究支持了这样一个假设：早期的同伴关系不良将导致儿童青少年短期或长期的社会适应困难。在此背景下，新一轮同伴关系的研究迅速崛起。在考察早期同伴交往困难对后期适应问题的影响时，研究者把儿童青少年的适应问题区分为，外化和内化问题。外化的问题是表现于外的、指向他人的问题，如攻击行为、反社会行为、活动过度以及其他一些行为障碍综合症的表现；而内化的问题，如孤独、焦虑、抑郁等。其中孤独感就是本研究要考察的适应指标。但是，在上世纪 60、70 年代，心理学领域的一些研究者认为儿童不会体验到孤独感，直到 80 年代才开始对儿童孤独感的研究。

其实孤独感是很常见的现象，不仅青少年和成人会体验到，而且儿童也能体验到，它可能产生于对归属感的普遍需要，即个体与其所关心的人建立一种稳定的社会关系的需要。因此，孤独感是对这种社会关系受到威胁时的认知和情绪反应。有研究者发现在对孤独感的描述中，儿童非常关注与同龄伙伴的人际关系。而儿童与同伴的交往可以分为 4 个水平：个体特征水平、人际互动水平、双向关系水平和群体接纳

水平。本研究涉及的社交自我知觉处于个体水平；受欺负和社交退缩处于人际互动水平；互选朋友数和友谊质量处于双向关系水平；积极提名和消极提名处于群体水平。而关系水平和群体水平就是我们平常所提到的友谊和同伴接纳这两种同伴关系。研究发现，孤独感与同伴交往各个水平的变量都有联系。

大量研究表明，儿童的孤独感与其同伴接纳水平相关非常显著，儿童的社交地位不同孤独感也有差异，儿童的社交地位越不利，其孤独感就越强。也有研究发现同伴接受性在问题行为与孤独感之间起着中介作用。Bush 的研究结果表明，被拒绝儿童经历了同伴较多的消极对待，更可能表现出较少的班级参与行为和更多的孤独感。

友谊关系充满感情色彩，它为双方提供亲密交流与袒露的机会。对于友谊可以从以下 3 个层次来考察：两个个体之间是否存在友谊；友谊的范围（即儿童拥有的相互认可的朋友的数量）；友谊的质量（即朋友之间提供的支持、陪伴或冲突水平）。例如，Parker 和 Asher 的研究表明，除了较低的同伴接纳性外，只有少数几个朋友或者甚至没有朋友的儿童，他们的某种非常重要的社交需要得不到满足，从而导致了孤独感的产生。朋友的陪伴和支持可减轻和消除儿童的孤独感。其他研究也表明，青少年的孤独感随着对同性和异性友谊满意感的增加而下降。可见儿童友谊的数量和质量都会影响到他们体验到的孤独感。

在小学里，受同伴欺负是一种经常发生的、对儿童的心理社会适应具有消极影响的现象。受同伴欺负，会使儿童产生对同伴的不信任感和对学校环境的不安全感，进而会体验到孤独。研究一致认为，受同伴欺负与孤独感相关，受欺负儿童比不受欺负的儿童报告了更高的孤独感。

与攻击行为不同，社交退缩反映了一种过分控制的行为模式。消极退缩 (passive withdraw) 的儿童往往因为社交焦虑而把自己从同伴群体中孤立出来。因此，与同龄人相比，消极退缩的儿童与同伴交往的次数较少，独处的时间更多，从而增加了他们体验孤立和孤独的机会。

鉴于孤独感是个体对自己社交状况的一种主观体验，个体对自己社交状况的评价会影响到他们体验到的孤独感。有研究考察了儿童的孤独感与他们对自己和他人的表征之间的联系，结果发现，孤独感与较低的自我觉知的能力、自尊以及自责的归因方式之间，存在着横向和纵向的联系。最近的研究也发现，与社会喜好相比，社交自我知觉对孤独感有更强的预测力，并且纵向研究也发现，社交自我知觉的上升和下降显著地影响到孤独感的降低和增强。

从以上研究可以看出，以往研究主要是从横向的角度考察了同伴交往对孤独感的影响，而较少有研究反过来考察消极的情绪——孤独感对同伴交往的影响。有研究者认为，情绪对儿童同伴交往的成功与否起着非常重要的作用。情绪控制是儿童能够与同伴友好相处的一个重要方面。有较多积极情绪的儿童也更受同伴欢迎，反之，情绪反复无常以及有较多消极情绪的儿童更可能被同伴拒绝。Vorbach 等人的研究证实了情绪控制能力对拥有高质量友谊的重要性。另外，Nolen 等人的追踪研究发现，体验到较高抑郁水平的儿童相对于那些抑郁水平较低的儿童保持了比较悲观的认知风格。由此，本研究也将考察孤独感是否也会导致儿童对自己社交能力的消极评价，即社交自我知觉。本研究假设，同伴交往变量能显著预测孤独感，而孤独感也对同伴交往变量有显著的预测效应。

为验证上述假设，本研究于 2002 年 6 月对被试进行了相关数据的收集，2 年以后，对同样的样本收集了同样的数据。这一追踪研究设计将允许本研究从发展的角度来检验儿童的同伴交往与孤独感之间的相互影响。

2 研究方法

2.1 研究对象

本研究抽取武汉市某小学 3、4 年级的学生为研究对象。2002 年 6 月 (T1) 测查了 3、4 年级的儿童，

平均年龄分别为 9.1 和 10.1, 2004 年 6 月 (T2) 对这些儿童再次施测,这时他们已经分别升入 5、6 年级,有效被试共有 274 人。被试具体情况见表 1。

2.2 研究工具

同伴提名 (Peer Nomination)　限定提名法,集体施测。给儿童提供一份班级名单表,要求他们在问卷上填上自己在班内最喜欢的三个同学和最不喜欢的三个同学。然后计算每个学生被他人提名的积极分数 (L M) 和消极分数 (LL),并在班级内标准化。

表 1　被试的性别、年级分布

被试	三年级	四年级	总计（n）
男生（n）	55	88	143
女生（n）	46	85	131
总计（n）	101	173	274

友谊质量问卷 (Friendship Quality Question naire) 该量表评价与最好朋友的友谊质量。共 18 个项目,是 40 项的《友谊质量问卷》(Parker & Asher, 1993) 的简表。它选用了原量表六个友谊维度(肯定与关心、帮助与指导、陪伴与娱乐、亲密袒露与交流、冲突解决策略、冲突与背叛) 的项目中负荷最高的三个。我们采用回译程序得到中文项目。将冲突与背叛的项目反向记分后,再将 18 个项目相加得到友谊质量总分。实测 Cronbach α 系数为 0.83。翻译后的量表适于测量中国儿童的友谊质量。

朋友提名 (Friend Nomination)　非限定提名法,集体施测。给儿童提供一份班级名单,要求他们根据自己的实际情况,将朋友的编号填写在问卷上。有几个写几个。然后计算每个儿童的互选朋友数,在班级内标准化。

班级戏剧问卷 (Class Play)　Masten 等 (1985) 编制的"班级戏剧问卷"被认为是测量儿童社会行为方面信、效度较高的工具。该问卷包括 6 个因素:社交/领导性、受欺侮、被排斥、消极 – 孤立、关系攻击、外部攻击。采用回译程序得到中文项目。施测时发给每个被试一份全班同学的名单,要求被试根据每个角色的要求在名单中选出一个或几个适合扮演该角色的同学。统计时,首先计算出每个项目的得分,每个儿童被其他儿童提名的次数之和为该儿童在项目上的得分;其次,计算各维度的平均分,一个维度上各项目得分的平均值为该维度的平均分;最后,计算各维度的标准分,将各维度得分在班级内标准化,计算出每个被试在 6 个因子上的标准分。根据研究需要,本研究选取其中的两个维度受欺侮和消极 – 孤立的 Cronbach α 系数分别为 0.95 和 0.78。

PCSC 量表 (The Perceived Competence Scale for Children) 是 Harter（1982）编制的儿童自我知觉问卷,

表 2　同伴交往各变量与孤独感的相关矩阵

变量	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1.LO1															
2.LM1	−0.31**														
3.LL1	0.08	−0.24**													
4.FQ1	−0.44**	0.27**	−0.07												
5.MF1	−0.37**	0.63**	−0.27**	0.36**											
6.HS1	−0.56**	0.25**	−0.17**	0.38**	0.29**										
7.VI1	0.14*	0.22**	0.602**	−0.08	−0.29**	−0.19**									
8.PW1	0.12*	−0.109	0.27**	−0.10	−0.17**	−0.18**	0.49**								
9.LO2	0.40**	−0.28**	0.14*	−0.34**	−0.36**	−0.37**	0.23**	0.19**							
10.LM2	−0.30**	0.63**	−0.21**	−0.25**	0.42**	0.22**	−0.23**	−0.15**	−0.30**						
11.LL2	0.11	−0.26**	0.74**	0.14*	−0.28**	−0.23**	0.58**	0.26**	0.12*	−0.25**					
12.FQ2	−0.27**	0.26**	−0.15*	0.30**	0.30**	0.15*	−0.20**	−0.13*	−0.44**	0.26**	−0.14*				
13.MF2	−0.31**	0.47**	−0.20**	0.31**	0.51**	0.20**	0.14*	−0.07	−0.34**	0.45**	−0.24**	0.29**			
14.HS2	−0.45**	0.30**	−0.17**	0.31**	0.31**	0.42**	−0.15*	−0.15*	−0.63**	0.26**	−0.14*	0.34**	0.34**		
15.VI2	0.116	−0.15*	0.40*	0.182	−0.35**	−0.19**	0.58**	0.34**	0.15*	−0.17**	0.68**	−0.19**	−0.12*	−0.12*	
16.PW2	0.105	−0.10	0.04	−0.10	−0.06	−0.12	0.37**	0.50**	0.13*	−0.00	0.25**	0.06	−0.01	−0.10	0.625**

注：* $p<0.05$;** $p<0.01$;LO = 孤独感;LM= 积极提名;LL= 消极提名;FQ = 友谊质量;MF= 互选朋友数;HS= 社交自我知觉;VI= 受欺负;PW= 消极退缩。变量名后的 1=T1, 变量名后的 2=T2。

经回译后得到中文版本，预试中的 Cronbach α 系数为 0.68。原量表有四个维度分别是：社交自我知觉、认知自我知觉、运动技能自我知觉和一般自我知觉，本研究只选用社交自我知觉这一维度，计算该维度所有项目总分的平均分。

儿童孤独量表 (Children's Loneliness Scale) 采用 Asher 等人 1984 年编制的专用于 3~6 年级学生的儿童孤独量表，经回译后得到中文版本。该量表包括 16 个孤独项目和 8 个关于个人爱好的插入项目，因子分析表明插入项目与负荷于单一因子上的 16 个孤独项目无关。实测的 Cronbach α 系数为 0.92，信度很高。计算 16 个项目的平均分（反向记分的题目先行转换），所得总分越高，表示孤独感越强。这一记分方法在国内外都被广泛采用。

2.3 统计分析

全部测量结果由 Filemaker 4.0 录入整理，由 SPSS 10.0 统计数据。

3 研究结果

3.1 相关分析

表 2 为各变量之间的相关矩阵。从横向的相关分析可以看出，T1、T2 的同伴交往变量除消极提名、
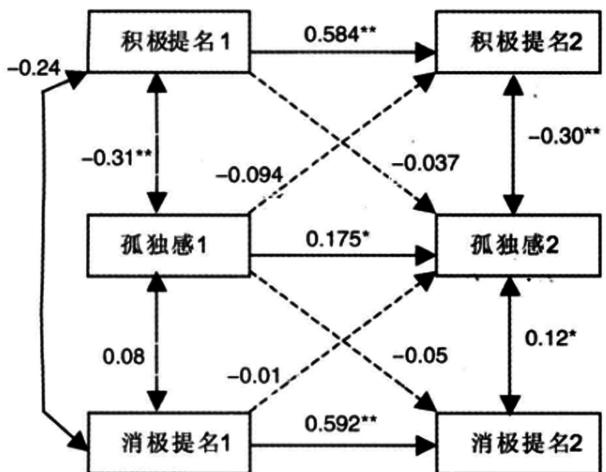
图 1 积极提名、消极提名与孤独感的交叉滞后回归分析
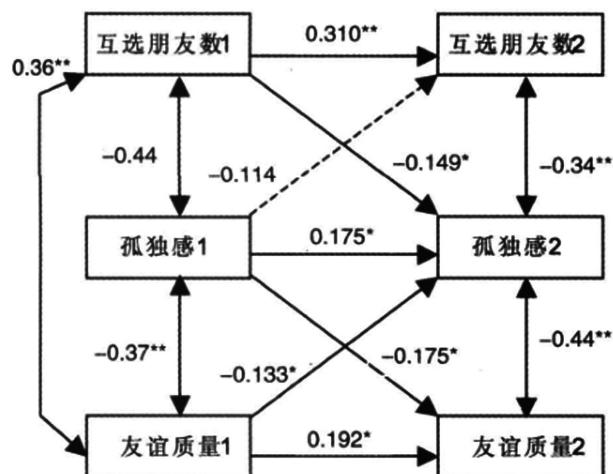
图 2 互选朋友数、友谊质量与孤独感的交叉滞后回归分析
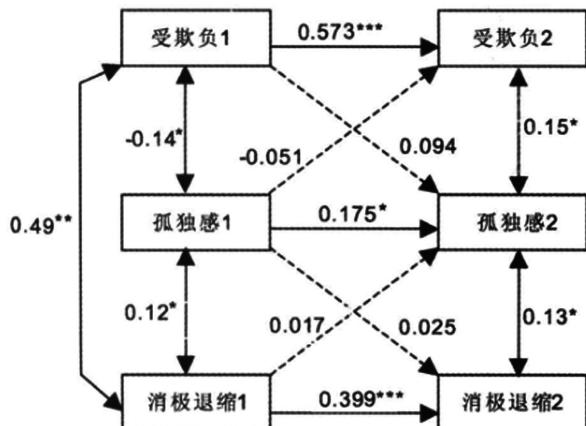
图 3 受欺负、消极退缩与孤独感的交叉滞后回归分析
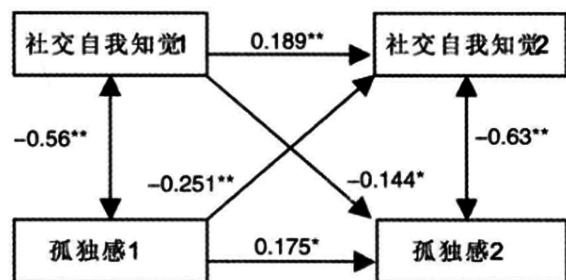
图 4 社交自我知觉与孤独感的交叉滞后回归分析

同伴欺负和消极退缩外，与 T1、T2 的孤独感的相关都达到了显著水平，具体来说，儿童获得积极提名越多、友谊质量越高、互选朋友数越多、社交自我知觉越积极，孤独感越低。受欺负、消极孤立与孤独感的相关在 T1 相关显著，而在 T2 相关不显著。在 T1 和 T2，消极提名与孤独感的相关都未达到显著水平。

3.2 交叉滞后回归

在相关分析的基础上，应用交叉滞后回归分析 (cross- lagged regression analyses)，得出了以下几个模式图，双向箭头实线表示相关显著，虚线表示相关不显著，线旁的数据为相关系数；单向箭头表示采用全部纳入法 (Enter) 进行分层回归获得的分析结果，线旁的数据为标准化偏回归系数 β 值，实线表示回归系数显著，虚线表示不显著。

3.2.1 同伴交往与孤独感之间的相互影响

图 1 的数据显示，在分别控制了前测的其它变量后，前测的积极提名、消极提名与孤独感的交叉滞后效应不显著，即前测的积极提名、消极提名对后测的孤独感没有显著的预测效应，而前测的孤独感也不能显著预测后测的积极提名和消极提名。

图 2 的数据显示，在控制了前测的其它变量后，前测的互选朋友数和友谊质量分别能够显著负向预测后测的孤独感 (β =−0.149, p <0 . 05; β =−0.133, p< 0.05); 前测的孤独感对后测的友谊质量也有显著的负效应 (β =− 0.157, p<0.05)，而对后测的互选朋友数则没有显著的预测意义。

图 3 的数据显示，在控制了前测的其它变量后，前测的受同伴欺负分数、消极退缩得分对后测的孤独感没有显著的预测效应，而前测的孤独感也不能显著预测后测的受同伴欺负分数。

图 4 的数据显示，在控制了前测的其它变量后，前测的社交自我知觉能显著负向预测后测的孤独感，β =−0.144, p<0.05, 而前测的孤独感也能显著负向预测后测的社交自我知觉，β =−0.251, p<0.001。

3.2.2 同伴交往与孤独感的稳定性

在控制了前测的积极提名、消极提名、友谊质量、互选朋友数、同伴欺负、消极退缩和社交自我知觉后，孤独感的稳定性为 β =0.175, p<0.05; 而在控制了前测的其它变量后，积极提名、消极提名、互选朋友数、友谊质量、受同伴欺负、消极退缩和社交自我知觉的稳定性分别为 β =0.584, β =0.592, β = 0.310, ps<0.001; β =0.192, p<0.01; β =0.488, β =0.399, ps<0.001; β =0.189, p<0.01。

4　讨论

4.1 儿童同伴交往与孤独感之间的相互影响

早期同伴关系和后期适应的因果理论模型认为，同伴交往在儿童的发展中扮演多种重要角色，一个儿童的同伴关系质量会影响后期适应结果。本研究结果表明，即使在控制了前测的孤独感和其它同伴交往变量的效应后，前测的友谊质量、互选朋友数、社交自我知觉分别能够显著负向预测后测的孤独感，这与以往研究结果一致；而在控制了前测的同伴交往各变量的作用后，前测的孤独感能显著负向预测后测的友谊质量和社交自我知觉。

首先讨论处于同伴群体关系水平的两个指标 -- 积极提名和消极提名与孤独感之间的关系。群体关系水平反映的是群体对个体的态度：接纳或排斥。儿童获得的积极提名越多，就越能体验到被关心以及归属于群体的感觉，这对减轻和抑止儿童的孤独感有积极的影响。但是本研究结果发现，在分别控制了前测的其它变量后，前测的积极提名、消极提名对后测的孤独感没有显著的预测效应。这可能是因为积极提名和消极提名与其它同伴交往变量之间，有较强的共线性，可能相互间产生某种抑止作用，所以在控制了其它同伴交往变量的作用后，积极提名和消极提名对两年后的孤独感的预测作用也就不显著了。研究也发现，前测的孤独感不能显著预测后测的积极提名和消极提名，即两年前的孤独感的高低并不能显著影响两年后同伴群体对他 / 她的喜欢程度。

儿童的互选朋友数以及自评的友谊质量，反映了儿童双向关系水平的友谊状况，都能预测儿童的社会能力和适应性的变化。本研究表明，在控制了早期的孤独感和其它同伴交往变量的作用后，儿童早期的互选朋友数和友谊质量对两年后的孤独感都有独特的影响。本研究只是整体上考察了互选朋友数对孤独感的效应，没有具体区分有无朋友、朋友多或少与孤独感的关系。Schaffer 总结了没有朋友对儿童发展可能造成的消极后果：如有情绪问题、较低的社会能力、

学校适应性差等。Parker 和 Seal 的研究发现，与拥有较少朋友的儿童相比，有较多朋友的儿童其孤独感较低，但是若把没有朋友的儿童排除在分析之外，友谊数量与孤独感的相关不显著。可见，友谊数量对孤独感的预测力，主要是有没有朋友所导致的效应，朋友数量的多少影响不大。在友谊中被一个人所喜欢与在同伴群体中被许多人所喜欢的体验有着质的不同。当青少年被问及他们希望从朋友那里得到什么或怎样告诉别人谁是他们的最好朋友时，他们常常说好朋友理解他们，值得信任，倾听他们的思想或情感，会为他们分忧解难。可见友谊质量越高，儿童从这种一对一的友谊关系中获得的积极情感支持就越多，从而为他们提供了抵制孤独感的缓冲器。如果儿童的互选朋友较少并且友谊质量也较差，那么他体验到的孤独感就会更加强烈。本研究还发现，在控制了前测的同伴交往各变量后，前测的孤独感仍对后测的友谊质量有独特的影响。以往有实验研究考察了孤独者与非孤独者在谈话中的交往风格，结果发现，在孤独维度上得分高的被试对他们的交谈伙伴表现出相对较小的兴趣，很少与同伴发生互动。并且，孤独者更倾向于扮演"被动的人际交往角色"，即他们不会做太多的努力去加入别人的谈话或活动。显然，孤独者消极被动的交往风格使那些很可能成为朋友的人远离了他们，同时也会降低他们与好朋友的友谊质量。

本研究结果发现，在控制了前测的孤独感和其它同伴交往变量后，消极退缩和受同伴欺负对后测的孤独感没有显著预测效应。这可能是因为，表现出退缩或 / 和攻击行为倾向的儿童，往往更容易成为受欺负的对象，退缩和攻击是与受欺负状况联系最密切的两种行为模式这也可以从表 2 的各相关矩阵中看出，退缩与受欺负在时间 1 和时间 2 的相关分别是 $r=0.49, r=0.625$。因此，当排除了前测的受同伴欺负的作用后，消极退缩对后测的孤独感就没有显著影响了。同样地，排除了前测的消极退缩的作用后，受同伴欺负也不能显著预测后测的孤独感。本研究还表明，在控制了前测的同伴交往各变量的作用后，孤独感不能预测后期的受同伴欺负和消极退缩。这说明即使儿

童体验到较高的孤独感也不能直接导致儿童受欺负和社交孤立，这与以往研究结果不一致，例如，Hawker 和 Boulton 比较了自我报告的压抑、孤独、焦虑和自我价值感对儿童受欺负的预测作用。当控制了这些因素之间的交互作用之后，结果发现它们都与自我报告的受欺负有关，但压抑与孤独是同伴报告的受欺负的最强预测源。这有待进一步验证。

社交自我知觉变量处于同伴交往的个体特征水平，是对自己社交状况的主观评价。本研究发现，儿童早期的社交自我知觉能显著负向预测后期的孤独感。孤独感的认知理论认为孤独感是人们对社交需要未满足的情绪反应，与个体生活的客观特征相比，人们对人际关系的标准和主观知觉与孤独感的关系更直接。理性情绪治疗理论也认为，个人的情绪和行为反应大部分取决于个体本身对事件如何理解、解释、并赋予其意义因此，儿童对自己的社交状况评价越积极，他们体验到的孤独感就越低，反之，体验到的孤独感就越高。本研究也表明，即使在控制了早期的社交自我知觉的影响之后，儿童早期的孤独感依然能显著负向预测两年后他们对自己社交状况或能力的评价。对此我们可以用伤疤效应来解释。Nolen 等人的研究验证了抑郁症的后果——伤疤效应，即与抑郁水平较低的儿童相比，抑郁水平较高的儿童保持了比较悲观的认知风格。类似地，儿童早期越孤独，他们对自己后期社交能力的评价就越消极、悲观。从早期孤独感对后期同伴交往各变量的影响来看，它对社交自我知觉的影响最大，这也可能是因为儿童期的认知风格正处于形成期，更容易受到其他因素的影响。

4.2 两年期间儿童同伴交往与孤独感的稳定性

本研究发现，儿童获得的积极提名、消极提名、受同伴欺负、消极退缩、互选朋友数在两年之间仍然呈现出较高的稳定性；友谊质量和社交自我知觉也具有一定的稳定性。这说明，同伴交往的特征一旦出现，将会一直稳定地持续下去，即同伴交往的稳定性主要是由它自身的特征维持的。此外，孤独感的稳定性仅在 0.05 水平上显著，所以，与同伴交往的某些变量相比，孤独感更容易受到外部环境的影响。Asher 等人

认为, 被同伴接纳的程度、是否被过度欺负、是否有好朋友以及与好朋友的友谊质量影响儿童体验到的孤独感。并且除同伴关系之外, 儿童与师生关系、亲子关系以及手足之间的更广泛的社交网络也与孤独感有关。

5　结论

（1）由自我报告所得的同伴交往变量 (友谊质量和社交自我知觉) 与孤独感之间的相互预测关系显著；而由同伴评定得到的同伴交往变量 (积极提名、消极提名、受同伴欺负、消极退缩) 与孤独感之间的相互预测关系则不显著；另外, 互选朋友数能显著预测两年后的孤独感, 而孤独感不能显著预测两年后的互选朋友数。

（2）由同伴评定所得的积极提名分、消极提名分、受同伴欺负、消极退缩、互选朋友数在两年之间所呈现的稳定性, 高于由自我报告法得到的友谊质量、社交自我知觉以及孤独感, 并且同伴交往变量的稳定性主要是由它自身的特征维持的。

参考文献

Asher S R, Paquette J A. (2003). Loneliness and peer relations in childhood . *Current Directions in Psychological Science, 12*(3), 75–79.

Buhs E S, Ladd G W. (2001). Peer rejection as an Antecedent of Yong mChildren's School Adjustment :An Examination of Mediating Processes . *Developmental Psychology, 37*, 550–560.

Burger JM. 著 . 陈会昌等 . *人格心理学* . 北京 : 中国轻工业出版社 , 2000. 252–253.

Cillessen A H N. (1997). Sociometric status, social self–perceptions, and the development of school adjustment in middle childhood. Paper presented at the biennial meeting of the society for research in Child Development.

Corsini R J , Wedding D, 朱玲意 , 林美熏 , 李立维等译 . (2000) *当代心理治疗的理论与实务* . 台北 : 心理出版社 , 261–266.

HawkerD S J . BoultonM J. (2000). Twenty year's reseach on peer victimiazation and psychosocialmal adjustment: A meta–analytic review of cross– sectional studies. *Child Psychology and Psychiatry, 41*(4), 441–455.

Hymel S, Rubin, K H, Rowden L. (1990). Children' s peer relationship: Longitudinal prediction of internalizing and externalizing problems from middle to late childhood. *Childhood Development, 61*, 2004–2021.

Lei L, Wang Y, Guo B L. (2004). The influence of classroom norms on the relations between social behaviors and peer victimization (in Chinese). *Acta Psychologica Sinica, 36*(5), 563–567.

[雷雳 , 王燕 , 郭伯良 . (2004). 班级行为范式对个体行为与受欺负关系影响的多层分析 . *心理学报 , 36*(5), 563–567.]

Marie V A. (2002). The relationship between emotional competence and social competence among early adolescents, 63(3–B), 1578–1588.

Nolen–Hoeksema S, Girgus J S, SeligmanM E. (1992). Predictors and consequences of childhood depressive symptoms : A 5–year longitudinal study. *Journal of Abnormal Psychology, 101*, 405–433.

ParkerJ G, A sher S R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and fee lings of loneliness and social dissatisfaction. *Developmental Psychology, 29,* 611–621.

Parker J G, Seal J. (1996). Forming, losing, renewing, and replacing friendships: Applying temporal parameters to the assessment of children' s friendship experiences . *Child Developmental, 67*, 2248–2268.

Rotenberg K J, Hymel S. (1993). *Loneliness in Childhood and Adolescence* . Cambridge:Cambridge University

Press, 3–6.

Rubin K H, Bukowski W, Parker J G. (1998). Peer Interactions, Relationships, and Groups . In: Damon W, Nancy Eed . Handbook of Child Psychology, l3, Social, Emotional and Personality Development (5h ed .). New York: Wiley, 619–700.

Santrock J W. (2001). Child Development (9h ed). New York: McGraw– Hill, 453–553.

Schaffer H R. (1997). Social Development Blackwell Publishers Ltd.

Schwatz D, Dodge K A, Petti G S, Bates J E. (2000). The Conduct Problems Prevention Research Group Friendship as a moderating factor in the pathway between early harsh home environment and later victimization in the peer group. *Developmental Psychology, 36*, 646–662

UrukA C, Demir A. (2003). The role of peers and families in predicting the loneliness level of adolescents . *The Journal of Psychology, 137*(2), 179–194.

Xin Z Q, Chi L P. (2003). The relationship between family functioning and children's loneliness: the role of mediator(in Chinese). *Acta Psychologica Sinica, 35*(2), 216–221.]

[辛自强 , 池丽萍 . (2003). 家庭功能与儿童孤独感的关系 : 中介的作用 . *心理学报 , 35*(2), 216–221.]

Yu G L, Xin Z Q, Luo X L. (2000). The characteristics of loneliness, peer acceptance and their relation to family functioning (in Chinese) *Acta Psychologica Sinica, 32*(1), 59–64.

[俞国良 , 辛自强 , 罗晓路 . (2000). 学习不良儿童孤独感、同伴接受性的特点及其与家庭功能的关系 . *心理学报 , 32*(1), 59–64.]

Zhang W X. (1999). *Children's Social Development(in Chinese)*. Beijing Normal University Press.

[张文新 . (1999). *儿童社会性发展* . 北京 : 北京师范大学出版社 , 165–166.]

Zhao D M. (2004). *Factors of Loneliness during Middle Childhood: Peer Acceptance, Friendship Quality and Self–Perceived Social Competence (in Chinese).* Master 's thesis, Central China Normal University.

[赵冬梅 . *童年中期儿童孤独感的影响因素 : 同伴接纳、友谊质量和社交自我知觉* . 硕士毕业论文 , 华中师范大学 , 2004.]

Zhou Z K, Fan C Y. (2001). Primary school children's social anxiety and Loneliness (in Chinese). *Psychological Science, 24*(4), 442–442.

[周宗奎 , 范翠英 . (2001). 小学儿童社交焦虑与孤独感研究 . *心理科学 , 24*(4), 442–442.]

Zhou Z K, Zhao D M, Chen J, et al.(2003). Loneliness as a fuction of sociometric status and self–perceived social competence in middle childhood (in Chinese). *Psychological Development and Education, 19*(4), 70–74.

[周宗奎 , 赵冬梅 , 陈晶 , 蒋京川 , Rachel Hundley. (2003). 童年中期儿童社交地位、社交自我知觉与孤独感的关系研究 . *心理发展与教育 , 19*(4), 70–74.]

Children's experiences with peers and loneliness: A two-year longitudinal study

Zhou Zongkui, Zhao Dongmei, Sun Xiaojun, Ding Xianfeng

(School of Psychology, Central China Normal University, Wuhan 430079, China)

Abstract   Evidences from researches on both primates and human supported the hypothesis that problems in early peer relationship could lead to subsequent psychosocialmaladjust ments, including externalizing and internalizing problems (such as feelings of loneliness and depression) . While children's experiences with peers can be best understood by referring to several levels of social complexity: within individuals, within interactions, within relationships, and within groups . Previous researches had revealed that loneliness was associated with a variety of peer variables at each of the four levels . Most existing researches had focused on the influences of peer variables on loneliness . Less examined was the impact of loneliness on peer variables . The present study was designed to explore the interaction between loneliness and peer variables using a 2 −year longitudinal design.

This study examined 274 children from Grade 3 to Grade 4 in 2002 and 2004. This study used the measures auch as the Perceived Competence Scale for Children, the Revised Class Play, Friendship Nomination and Friendship quality Questionnaire, Sociometric Nominations, and Children's Loneliness Scale to assess self −perceived social competence, social withdraw and peer victimization scores, mutual friends number and best friendship quality, positive nomination and negative nomination, and loneliness score, respectively.

Cross − lagged regression analysis was conducted to examine the mutual predicative relations between loneliness and peer variables within each of the four levels . The results indicated that(1) there is amutually predictive relations between peer variables and loneliness during the two − year interval such that (a) at the group level, after controlling other variab les at Time 1, no significant cross − lagged effects were found between "liked most"nomination, "like least"nomination, and loneliness; (b) at the relationship level, after controlling other variables at Time 1, there was a significant cross− lagged effect between friendship quality and loneliness, suggesting that early friendship quality significantly predicted loneliness two years later, and early loneliness also significantly predicted friendship quality two years later;additionally, number of early mutual friends significantly predicted loneliness two years later, whereas early loneliness didn't predict number of mutual friends two years later;(c) at the interaction level, after controlling other variables at Time 1, no significant cross − lagged effect was found between passivewithdraw, peer victimization, and loneliness; (d) at the individual level, after controlling other variables at Time 1, there was a significant cross − lagged effect of self − perceived social competence on loneliness and vice versa, indicating that early self − perceived social competence significantly predicted loneliness two years later, and early loneliness also significantly predicted self − perceived social competence two years later . (2)With respect to the

stability of peer variables and loneliness during the 2 –yealr interval, the " like most"no mination, " like least"nomination, peer victimization, passive withdraw and number of mutual friends were highly stable, while best friendship quality, self–perceived social competence, and loneliness showed moderate stabilitie.

Conclusions:Forst, mutual predictive relation between self– reported peer variables and loneliness were more stronger than that between peer– rated peer variables and loneliness . Second, the peer– rated variables (ie., " likemost" nom ination, "like least"nomination, peer victimization, passive withdraw and mutual friends number)were more stable than those of self – reported variables (. i e . , friendship quality, self – perceived social competence and loneliness) . Third, once the characters of peer variables merged, they would show continued stability .

Keywords   Children' s experiences with peers; loneliness; cross– lagged regression design

The effect of general creative personality and freedom of task choice on adolescents' social creativity

Chuanhua Gu[1], Biying Hu[2], Zhi Jing[1] ,Flemmings F. Ngwira[2] , Zongkui Zhou[2]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior, Ministry of Education; School of Psychology, Central China Normal University)

([2] Faculty of Education, University of Macau, Macao, China)

Abstract   This study investigated the effect of general creative personality and freedom of task choice on the social creativity of adolescents. The results indicated, first, that senior high school students scored higher than junior high school students. Second, girls scored higher than boys on originality, fluency, flexibility, appropriateness, and utility with regard to creative social problem-solving. Third, freedom of task choice and its interaction with creative personality had significant effects on the originality, appropriateness,utility, flexibility, and fluency of social creativity. Adolescents who completed the task voluntarily scored higher on these dimensions than adolescents who completed it reluctantly and, among the voluntary adolescents, those with high and medium creative personality scored higher than those with low creative personality, whereas no such difference was found among the reluctant adolescents. Adolescents were more likely to show social creativity, and their general creative personality was more likely to be brought into effect under the freedom of task choice condition.

Keywords   social creativity; general creative personality; freedom of task choice; adolescent.

1 Introduction

Since the mid-twentieth century, the most popular way of classifying research on creativity has been the 4Ps: person, process, pressure, and product (Runco, 1997, 2004). In other words, creativity research has generally focused on how individual and environmental factors such as personality and working conditions affect the development of creativity (Feldman, 1999).

1.1 The relationship between personality and creativity

According to Runco (2004), exploring motivation is an important aspect of research on creative personality. Amabile and colleagues (Amabile, 1983, 1996; Amabile, Hill, Hennessey & Tighe, 1994) explored the effects

of social factors (e.g., evaluation, reward and work compulsion, social promotion) on artistic and linguistic creativity, and proposed the intrinsic motivation hypothesis. Applying the intrinsic motivation hypothesis to linguistics and the arts, they confirmed that intrinsic motivation improves individual creativity on a specific task, whereas extrinsic motivation decreases individual creativity on the same task. Environmental factors such as pressure were also found to influence creative performance by affecting the features of the creative person.

The earliest studies on the relationship between personality and creativity examined the typical personality traits of creative people, dating back to Terman's (1906) psychological research on the personality traits of creative children (Simonton, 1999). Such studies found that highly creative people, both adults and children, usually possess some typical personality or character traits. These include not only general creativity traits such as openness, thirst for knowledge, emotionality, independence, self-confidence, and diligence, but also obvious domain specificity (Gu, 2011; Lin, 2009; Torrance, 1962). Feist (1999) found that creative people with artistic talents were more impulsive, anxious, sensitive, irresponsible, insubordinate, independent, inhospitable, and unsociable than creative people with talents in the field of science. Certain personality traits, such as independence and openness, can promote individual creativity while traits such as dependence can impede it (Gu, Zhou, & Wang, 2008; Sternberg & Lubart, 1992). Studies on creative personality traits using the social personality paradigm reinforced the motivating effect of personality on creativity (Feist, 1999). The majority of studies have reached the following conclusions: creative people (including artists and scientists) tend to possess similar creative personality traits, despite the differences between different disciplines, and creative personality traits are closely related to creative performance. Some researchers regard creative personality as a constituent part or subsystem of a creative system,

influenced by the systems theory perspective (Gu, 2005; Hennessey & Amabile, 2010).

1.2 The interaction between personality and the context or environment on creativity

As an important topic in creativity research, and with an urgent need to motivate people's creativity in different working conditions, researchers have been called upon to evaluate the creative environment (pressure). For instance, Amabile and Gryskiewicz (1989) designed the *Work Environment Inventory*, which measures the working conditions for promoting employees' creativity. According to the intrinsic motivation hypothesis, an environment can be considered creative and will help people to exert their creativity if it motivates their intrinsic rather than their extrinsic motivation. In contrast, an environment that encourages people to work for external rewards or praise provides only extrinsic motivation that stifles rather than fosters creativity (Amabile, 1996; Collins & Amabile, 1999). Furthermore, the nature of motivation can be altered by different types of rewards. In a metaanalysis of 60 experimental and non-experimental studies examining the relationship between rewards and creative performance, Byron and Khazanchi (2012) found that creativity-contingent rewards tended to increase creative performance, and were more positively related to creative performance when individuals were allowed greater choice (and were less controlled) and were given more positive, contingent, and task-focused performance feedback. In contrast, performance-contingent or completion-contingent rewards tended to have a slightly negative effect on creative performance.

The investment theory of creativity proposed by Sternberg and Lubart (1992, 1999) stated that intellect, knowledge, thinking style, motivation, personality, and the environment are readily accessible resources for creative people that can be divided into individual and environmental factors, and their interactions produce

creativity.

According to the cognitive−affective system theory of personality (Mischel & Ayduk, 2002; Mischel & Shoda, 1995), the features of situations (i.e., social and interpersonal situations, intrapsychic situations) activate a set of internal reactions—both cognitive and affective—based on the individual's prior experience with those features. The characteristic processing dynamics of the cognitive−affective personality system in relation to the relevant features of situations generate diverse forms of behavior. As the interactive or integrated effect of these individual and environmental factors has been verified in the arts and science domains, the specific individual and environmental factors are believed to have an interactive influence and integrated effects on creativity in social problem−solving.

1.3 Previous research on social creativity

The domain specificity of creativity, such as social creativity, has attracted increasing attention recently (Baer, 2008). Social creativity is the creativity that arises in social interactions or social activity. Specifically, it refers to the individual characteristics that a person uses to solve social problems in original, unique, and appropriate ways (Gu, 2011). Mouchiroud and Lubart (2002) considered social creativity as a continuum with distinct poles: the interpersonal social creativity shown by people in the process of dealing with personal relations (such as individual leadership in everyday life), and the social creativity affecting important aspects of social life (such as the social creativity shown by eminent religious or political leaders). Social creativity is the most common type of creativity, and is a combination of cognition and action and corresponding personality traits, formulated on the basis of basic adaptive skills. It is an important ability for proposing and solving all kinds of social problems, accommodating to and organizing social life, making good social decisions, and maintaining good mental health. It is not, however,

identical to social skills, social problem−solving abilities, or social leadership skills. Both social skills and social problem−solving abilities highlight a person's general social adaptation and the appropriateness and utility of the method to solve problems while social leadership emphasizes the individual ability to manage and organize interpersonal relationships. Social creativity, in contrast, is a higher−level and more comprehensive characteristic, comprising the personality tendency for dealing with all sorts of interpersonal relationships and social problems, and cognitive ability. Hence, it is necessary to study the process of social creativity development, especially the factors that influence it and their interacting mechanisms, on the basis of existing research.

In the past 10 years, researchers have conducted some initial studies on the development of children's social creativity and its relationship with individual, family, school, and peer factors (Mouchiroud & Lubart, 2002; Mouchiroud & Bernoussi, 2008; Gu, 2007, 2009; Gu, Huang, & Zhou, 2008; Gu, Zhang, & Zhou, 2009). Other studies have explored the development of social leaders' creative personalities over the lifespan (Gu, 2011). Research on social creativity can be divided into two categories: studies of historical figures in different disciplines such as history, politics and sociology; and studies of the general population, especially normal children and adolescents. James and Asmus (2001) explored the relationship between creativity in different domains (including social activities). Mouchiroud and Lubart (2002) investigated the social creativity development of early school−age children while Mouchiroud and Bernoussi (2008) studied the relationship between social creativity and factors such as social skills, popularity, and extracurricular activities.

Although the abovementioned studies explored the development of social creativity and its relative ability to a certain extent, they lack a deep exploration

of the interactive influences between various factors. Furthermore, little research has been conducted on adolescents, and even less on the effects of personality and task. Does a general creative personality, a positive psychological tendency toward creative activities, affect adolescents' social creativity on a specific task? Does their attitude toward completing a task (whether compulsory or not) affect their social creativity performance? Do creative tendency and attitude toward the task have an interactive effect on adolescents' social creativity on specific tasks? Studies such as those by Amabile (1983, 1996) emphasized creativity in language and art (including writing, painting, etc.), but seldom referred to creativity in the domains of science or social problem-solving. It is worthwhile exploring, therefore, whether the effects of personality and task features on general creative thinking in a specific task also exist in the social creative domain.

Adolescence is a special period of development because it is full of uncertainty, and is a time when social interaction becomes particularly important. As Frauenknecht and Black (2004) pointed out, social problem-solving may be the single most important skill that a young person can acquire. Social creativity is a significant asset for adolescents seeking the best solutions to social problems in their daily lives. Hence, this study took adolescents as the participants.

1.4 The current study

The aim of the study was to explore the mechanism of the effects of general creative personality (individual factor) and freedom of task choice (environmental factor), and especially their interaction, on adolescents' social creativity. The effect of different combinations of creative personality and task presentation were examined to assess whether they inspired adolescents' social creativity in special problem-solving situations, to provide a reference for social creative education. Specifically, taking the adolescents as the subject and starting from the special

domain of social creativity, this study explored the effect of individual and environmental factors and their interactive influence during the process of social creation. The freedom of task choice was manipulated by controlling the way the task was presented to the participants; that is, some of the participants volunteered to complete the task while others were reluctant to do so. The motivation of the voluntary subjects to complete the task was intrinsic while that of the reluctant subjects was extrinsic.

Prabhu, Sutton and Sauser (2008), in a study of college students, found that some personality traits (e.g., openness to experience, self-efficacy) influenced creativity by affecting intrinsic motivation while a complex relationship existed between the personality traits (e.g., self-efficacy, persistence) of people with different levels of extrinsic motivation and creativity. That is, intrinsic motivation was a mediating variable between personality and creativity while extrinsic motivation was a moderating variable.

Based on the research findings above, this study tested the following hypotheses. First, general creative personality and freedom of task choice will have significant main effects on social creativity; the social creativity of adolescents with high creative personality will be significantly higher than that of adolescents with low creative personality; and the social creativity of adolescents who volunteer to complete the task will be significantly higher than that of adolescents who do so reluctantly. Second, general creative personality has a significant interaction with freedom of task choice. The social creativity of adolescents with different levels of creative personality will differ significantly when completing the task voluntarily (intrinsic motivation), and adolescents with high (low) creative personality will show higher (lower) social creativity. There will be no significant difference between the social creativity of adolescents with different levels of creative personality when they complete the task reluctantly (extrinsic motivation).

The majority of studies have found a close relationship between the development of creativity and various demographic variables, particularly age, gender, birth order and number of siblings (Feldman, 1999). Hence, the current study also investigated whether the social creativity of adolescents differed according to demographic differences (gender, school grade, with/without siblings). Accordingly, the third hypothesis was that there would be significant differences in social creativity between adolescents with different demographic characteristics.

2 Method

2.1 Participants

A sample of 210 healthy adolescents was selected from a high school in Wuhan city in central China; 106 were from 2 junior middle school classes with an average age of 13.61 years (SD=0.60), and 104 were from 2 senior middle school classes with an average age of 15.95 (SD=0.73). The sample consisted of 135 girls and 75 boys. All of the participants could read and write and could generate written responses to hypothetical situations or tasks without any difficulty.

2.2 Measurement

Open-ended questions about social problem situations Adolescents' social creativity was assessed using open-ended questions about social problem situations. Firstly, the high school students were interviewed about their daily life asking them open-ended questions about the kinds of social problems they commonly encountered. The typical social problems they faced were identified accordingly. Finally, the open-ended questions about social problem situations were compiled using the relevant research (Gu, Liu, & Zhou, 2010) as a reference, including six typical situations concerning time management, emotion regulation, learning, establishing interpersonal relationships, heterosexual relationships and peer relationships. Every situation was presented as a story, and the adolescents were asked to provide as many solutions as possible to help the character to solve the problem. The participants could answer the questions in any order, regardless of the sequence of the six situations, and their responses were given anonymously.

Taking relevant research (Gu, 2009, 2011) as a reference, five dimensions—originality, fluency, flexibility, appropriateness, and utility—were measured. Originality is the extent to which a solution differs from others; appropriateness is the extent to which a solution is acceptable to society; utility is the extent to which the solution solves the problem; flexibility refers to the number of categories of solutions; and fluency refers to the fluency of the participant's thinking. The originality, appropriateness, and utility of each answer to each situation or task were judged by a group of five adults with professional experience using a 3-point Likert scale. Fluency was scored as the total number of solutions provided, and flexibility as the number of categories into which all of a participant's solutions could be classified. Each judge was trained and then assessed independently using the same criteria or definition.

For each participant, the total score on each dimension was computed by combining the scores from the five raters and six situations. According to relevant theories (constant probability of success model; Simonton, 1990) and empirical studies (Mouchiroud & Lubart, 2002; Simonton, 1990), there is a positive and significant association between originality (quality of solutions) and fluency (quantity of solutions) in different age groups. The intra-class correlation coefficients for appropriateness, utility, originality, and flexibility, calculated using the inter-rater reliability method suggested by Shrout and Fleiss (1979), ranged from 0.86 to 0.89.

To determine whether each participant's attitude toward the task was voluntary or reluctant, following the

open-ended questions, he or she was asked to answer the question, " Were you interested in answering these questions?" followed by two choices: "Yes" or"No, but I answered them for the sake of completing the task."

The creativity assessment packet The adolescents' general creative personality was measured using the creativity assessment packet (divergent affection parts) revised by Williams, Lin and Wang (1994). This is a 50-item self-report checklist to assess individual curiosity, complexity, risk-taking, and imagination. It has good reliability and validity and the Cronbach's Alpha coefficient obtained in the current study was 0.84.

The participants were divided into three groups (high, medium, low) according to their total creative personality scores. The high group comprised those with a total score one standard deviation above the grade average, the low group comprised those with a total score one standard deviation below the grade average, and the medium group comprised the remainder. Risk-taking refers to the courage to confront failure and criticism, to speculate about and complete tasks even in chaotic contexts; that is, to stand up for one's own opinions. Curiosity refers to the characteristics of exploration and inquisitiveness, a fondness for blurred situations, enjoying working out how things work, and being able to grasp the special status and observe the result. Imagination refers to the ability to visualize, establish images in the mind, conceive things that have not happened, and conjecture what is beyond sensation and real by intuition. Complexity refers to seeking out all possibilities, seeing the distance between possibility and reality, deducing a law from the chaos, and being willing to explore complex problems.

2.3 Procedures

Before commencing the data collection, the participants provided their consent to par ticipate in the study. Students who did not wish to fill out the survey were allowed to work

on other schoolwork while the other students completed the task. The creativity assessment packet was distributed to students in their own classrooms, followed by the questionnaire on social problem situations. A trained undergraduate took charge of each class and gave the students the same instructions: "In everyday life, students are often confronted with difficult situations that demand solutions. Some of the situations faced by students are listed below; please say what the students should do under such conditions." The participants were asked to respond independently and indicate their attitude to the task (voluntary or reluctant) after answering the question "Are you interested in answering these questions?" by selecting either "Yes" or "No, but I will answer them for the sake of completing the task" ("Yes" representing voluntary, "No" representing reluctant). Hence, the study used a 2 (willingness to complete the task: reluctant, voluntary) × 3 (general creative personality: high, medium, low) experimental design. The freedom of task choice referred to whether the participants were willing to accept and complete the task in the current study. The voluntary participants completed the task without feeling forced, due to its intrinsic interest. In contrast, those who completed it reluctantly did so for external reasons rather than intrinsic interest. The data were analyzed using descriptive analyses, t-tests and ANOVA in SPSS version 17.0 (SPSS Inc., Chicago, Illinois).

3 Results

3.1 Variation in social creativity due to demographic variables

As shown in Table 1, there were significant differences in social creativity and its dimensions according to gender and grade. No significant difference was found between children with and without siblings on either of the dimensions.

Social creativity total scores and scores on each dimension (originality, fluency, flexibility, appropriateness, utility) were significantly higher for girls than for boys. Social creativity scores, including scores on each dimension, were higher among senior high school than junior high school students. Children with siblings scored higher than those without siblings on overall social creativity and on each dimension, but the differences were not statistically significant.

3.2 The effect of creative personality and freedom of task choice on adolescents' social creativity

Table 2 displays the means and standard deviations of the six experimental groups on the five dimensions of social creativity.

MANOVA was performed, with the dimensions of social creativity as the dependent variables and creative personality and freedom of task choice as the independent variables. Only freedom of task choice had a marginal main effect on social creativity total scores. Creative personality did not have a main or an interaction effect on social creativity total scores—freedom of task choice: $Wilks'\lambda = 0.95$, $F(5,186) = 2.08$, $p=0.07$, $\eta^2=0.053$; creative personality: $Wilks'\lambda =0.95$, $F(10,372)= 0.98$, $p=0.46$, $\eta^2=0.03$; interaction: $Wilks'\lambda =0.93$, $F(10,372) =1.36$, $p=0.20$, $\eta^2=0.04$. However, there was a significant main and interaction effect of freedom of choice on the dimensions of social creativity (see Table 3).

Further tests showed that voluntary adolescents scored significantly higher than reluctant adolescents on each dimension of social creativity (fluency: $T=4.19$, $df=194$, $p<0.001$; originality: $t=4.42$, $df=194$, $p<0.001$; flexibility: $t=4.33$, $df=194$, $p<0.001$; appropriateness: $t=4.60$, $df=194$, $p<0.001$; utility: $t=4.44$, $df=194$, $p<0.001$). A simple effect test showed that creative personality had a main effect on each dimension of social creativity among voluntary adolescents—fluency: $F(2,125)=3.60$, $p=0.03$, $\eta^2= 0.05$; originality: $F(2,125) =3.86$, $p=0.02$, $\eta^2= 0.06$; flexibility: $F(2,125)=3.86$, $p=0.02$, $\eta^2=0.06$; appropriateness: $F(2,125) =4.12$, $p=0.02$, $\eta^2=0.06$; utility: $F(2,125) =4.16$, $p=0.02$, $\eta^2 =0.06$—and adolescents in the high and

Table 1. Variation of adolescents' creativity on gender, grade and number of children (single or not; M, SD, T)

Variables	Male M(SD)	Female M(SD))	t (208)	Eighth grade M(SD)	Eleventh grade M(SD)	t(208)	Single children M(SD)	Nonsingle–children M(SD)	t(202)
Fluency	11.560(5.749)	13.355(5.128)	−2.327*	10.679(4.835)	14.788(5.200)	−5.932***	12.737(5.443)	13.136(5.205)	−0.043
Originality	12.912(7.554)	16.210(7.199)	−3.126**	11.230(5.448)	18.907(7.294)	−8.652***	15.066(7.578)	15.413(6.730)	−0.28
Flexibility	8.885(3.877)	10.816(3.315)	−3.803***	8.471(3.061)	11.813(3.406)	−7.840***	10.131(3.616)	10.463(3.501)	−0.54
Appropriateness	25.746(13.659)	31.991(11.519)	−3.518***	25.101(11.354)	34.509(12.181)	−5.790***	29.715(12.675)	31.090(11.824)	−0.065
Utility	23.226(12.210)	28.213(10.040)	−3.188**	22.460(10.231)	30.480(10.507)	−5.604***	26.430(11.213)	27.518(9.995)	−0.58
Total	82.330(41.695)	100.586(36.089)	−3.320**	77.943(33.999)	110.500(37.160)	−6.620***	94.080(39.327)	97.623(35.922)	−0.54

*P<0.05, ** P<0.01, *** P<0.001.

Table 2. Means and standard deviations of the six experimental groups on each dimension (*M and SD*)

	Voluntary (*n=128*)					Reluctant (*n=68*)				
	Fluency	Originality	Flexibility	Appropriateness	Utility	Fluency	Originality	Flexibility	Appropriateness	Utility
High creative personality (*n=37*)	15.444 5.522	19.393 8.841	12.089 3.649	35.800 12.704	31.526 10.937	10.500 4.503	11.940 4.838	8.540 2.827	21.8000 11.908	19.140 10.320
Medium creative personality (*n=126*)	14.125 4.884	17.093 7.015	11.146 3.119	33.761 10.726	29.889 9.499	10.290 5.628	11.526 6.173	8.368 3.682	23.942 12.706	21.468 11.463
Low creative personality (*n=33*)	10.923 4.645	12.539 5.486	9.031 3.393	25.323 9.934	22.339 8.748	12.000 5.058	14.190 7.659	10.150 3.410	29.150 12.668	26.270 10.509

Table 3. The Effect of Freedom of Task Choice and Creative Personality on Dimensions of Social Creativity (MANOVA)

Independent variables	Dependent variable s	SS	df	MS	F	η^2
	Fluency	196.72	1	196.72	7.52**	0.04
	Originality	428.424	1	428.424	8.65*	0.044
Willingness	Flexibility	89.874	1	89.874	8.04*	0.041
	Appropriateness	1325.070	1	1325.070	9.80*	0.049
	Utility	944.015	1	944.015	9.06*	0.046
Creative personality	Fluency	34.598	2	17.299	0.66*	0.007
	Originality	81.314	2	40.657	0.82*	0.009
	Flexibility	9.218	2	4.609	0.41*	0.004
	Appropriateness	65.833	2	32.917	0.24*	0.003
	Utility	45.920	2	22.960	0.22*	0.002
Willingness × creative personality	Fluency	176.642	2	88.321	3.38*	0.034
	Originality	391.922	2	195.961	3.96*	0.040
	Flexibility	108.894	2	54.447	4.81*	0.049
	Appropriateness	1452.565	2	726.283	5.37*	0.054
	Utility	1204.717	2	602.358	5.78*	0.057

$**P<0.01, ** P<0.05$

medium group scored significantly higher than those in the low group on each dimension—fluency: mean difference $(MD)_{\text{high-low}}=4.52, p=0.01$; $MD_{\text{medium-low}}=3.202$, $p=0.03$; originality: $MD_{\text{high-low}}=6.85$, $p=0.01$; $MD_{\text{medium-low}}=4.56$, $p=0.04$; flexibility: $MD_{\text{high-low}}=3.06$, $p=0.01$; $MD_{\text{medium-low}}=2.12$, $p=0.03$; appropriateness: $MD_{\text{high-low}}=10.48$, $p=0.01$; $MD_{\text{medium-low}}=8.45$, $p=0.01$; utility: $MD_{\text{high-low}}=9.19$, $p=0.01$; $MD_{\text{medium-low}}=7.55$, $p=0.01$. In contrast, for the reluctant adolescents, creative personality had no significant effect on any dimension.[1]

4 Discussion

4.1 Group difference in adolescents' social creativity

This study extended the relevant literature on creativity by exploring adolescents' creativity from the perspective of Chinese culture. Compared with artistic and scientific creativity, the social creativity investigated in this study was more likely to be domain specific.

The study found that social creativity, including the dimensions of originality, fluency, flexibility, appropriateness, and utility, was higher among senior high school students than junior high school students. This result confirms the findings of previous research; that is, the creativity development trend reported by Torrance (1960), and the social creativity development trend found in studies of primary school children (Gu, Liu, & Zhou, 2010; Mouchiroud & Lubart, 2002). Torrance (1960, 1964) discovered that children's creativity increased from grade 1 to 3, decreased in grade 4, increased again in grade 5, decreased from grade 6 to 7, and then continued to increase until adulthood. As the participants (grade 8 and grade 11) in this study had passed the "decreasing period,"

---

[1] This study also examined the moderating effects of creative personality as a continuous variable on the relationship between freedom of task choice and social creativity, following the procedure for interaction terms involving continuous predictors described by Aiken and West (1991) and Jaccard, Turrisi and Wan (1990). The results showed that the relationship between freedom of task choice and social creativity was moderated by creative personality. The interactions between freedom of task choice and fluency, originality, flexibility, appropriateness, and utility of social problem-solving reached statistical significance (fluency: $b=-0.21$, $t=-2.95$, $p=0.004$; originality: $b=-0.31$, $t=-3.25$, $p=0.001$; flexibility: $b=-0.16$, $t=-3.41$, $p=0.001$; appropriateness: $b=-0.58$, $t=-3.64$, p<.001; utility: $b=-0.55$, $t=-3.96$, p<.001). Overall, these findings support the results reported here.

their cognition and personality were in a period of steady development. As children grow older, their increased social interaction experience and development of formal and postformal thinking promote the development of social creativity. In contrast, the results of this study differ from those on general creative thinking. For example, Yi, Hu, Plucker and McWilliams (2013) found that elementary school children's creative thinking scores were significantly higher than those of middle school children. The differences in creativity development with age suggest that social creativity is domain specific to some extent.

No previous study has reported a similar conclusion concerning the development trend of creativity. Runco and Charles (1997) reviewed the relative research and pointed out that there were peaks and troughs in the development of the potential for creativity (or latent creativity), and creativity performance showed an alternately descending and ascending pattern. These peaks and troughs appear at different ages. Researchers have found different development trajectories of creativity, some multimodal, some U−shaped, and some an inverted J−shape. Different aspects of creative thinking ascend and descend at different speeds. There is significant individual variation in the development of creativity during both childhood and adulthood. A study of Chinese adolescents (Lin, 2009) found that the development trajectory of creative thinking ability followed an inverted V pattern, ascending in junior high school and descending in senior high school. As a new research field, social creativity clearly needs to be investigated further.

This study found a significant gender difference, with girls performing better than boys in originality, flexibility, fluency, appropriateness, and utility during the process of social problem−solving. Other studies have found higher social creativity among girls than boys (Gu, Zhou, & Hu, 2009; Gu, Zhou, & Zhong, 2009). There are two plausible explanations for this phenomenon. First, Chinese culture

has undergone extensive changes since the ongoing reform and opening up. In particular, the implementation of the family planning policy in 1979 has altered parents' perception of gender: the idea that "men are superior to women" has been replaced by "men are equal to women." As most families can only give birth to one child in China, the majority of families have only a single child, thus parents' attitudes toward children and their parenting behavior have nothing to do with a child's gender. In other words, girls receive the same education as boys instead of being discriminated against or maltreated because of their gender. Second, girls mature faster than boys, both physiologically and psychologically, and find it easier to communicate socially due to their advantage in linguistic creativity; hence, they tend to build close interpersonal relationship and obtain more experience of social problem−solving (Gu, Zhou, & Hu, 2009; Lin, 2009). After they reach puberty, girls are concerned with the success of social intercourse in addition to school achievement (Feldman, 2005). Moreover, Chinese culture expects adolescent girls to be quiet and amiable, which might make their solutions to social problems more appropriate and useful. All of the above might explain their advantage in social creativity development. This study provides some evidence for cultural differences in the conception of creativity and creative education, as proposed by Niu and Kaufman (2013).

Although children with siblings scored higher on social creativity and its dimensions than single children, the difference was not significant, which failed to support the hypothesis. The results are, however, in accordance with the conclusion of a study on Chinese only children (Fang & Fang, 2005). Gu and Zhou (2008) found that children's social creativity was associated with parenting style; the expression of warmth and understanding was favorable to children's social creativity while severity and punishment were unfavorable. At present, most children living in cities

in China are "single children," whose parents have more time, energy and financial resources to nurture them and to provide warmth and understanding, which are favorable for the development of their social creativity.

Many media, such as TV programs, newspapers and the Internet, have debated the problems that might arise in the mental development of single children, such as selfishness, willfulness, and anti-socialness. As a result, parents have taken heed of such calls and have made efforts to avoid the projected harmful effects of a lack of siblings by consciously highlighting the development of interpersonal relationships (parent-child, peer, friendship). Hence, single children's social creativity and ability to solve social problems have improved, and the differences between them and children with siblings have diminished accordingly.

4.2 Effect of creative person ality and freedom of task choice on adolescent's social creativtty

Amabile (1983, 1996) and Amabile et al. (1994) found that college students tended to perform better on creative tasks when intrinsically motivated than when extrinsically motivated. When the participants felt free to choose or decide different aspects of the tasks, their intrinsic motivation increased and so did their social creativity. They concluded that choice imposed by others is bad for creativity while choosing for oneself is good for creativity. In other words, freedom of task choice will bring about higher creativity than imposed choice.

This study verified this finding from the perspective of social creativity. It found significant main and interaction effects of freedom of task choice on originality, appropriateness, utility, flexibility, and fluency, with voluntary adolescents scoring higher than reluctant ones on each dimension of social creativity. Furthermore, among the voluntary adolescents, those with high or medium creative personalities scored higher than those with low creative personalities on each dimension while no main effect of personality was found among reluctant

adolescents. This shows that under the free task choice condition, adolescents were more likely to exhibit social creativity with the positive effect of a creative personality. Adolescents with a high creative personality, intensive curiosity, complexity, risk-taking, and imagination were more likely to solve problems positively, raising more flexible, unique, appropriate, and effective solutions. Obviously, the adolescents who completed the task voluntarily were motivated intrinsically while those who completed it reluctantly were motivated extrinsically. The combination of high creative personality and intrinsic motivation inspired high levels of social creativity. However, the personality of reluctant adolescents, whether high or low in creativity, did not affect their social creativity.

Shalley (1995) examined the effects of coaction, expected evaluation, and goal setting on individual creativity and productivity, and found that higher levels of creativity occurred when individuals worked alone. Productivity was high when the participants worked alone under no expectation of evaluation, but was low when they worked alone or were assigned a creativity goal. The results of this study further support Shalley's findings on social creativity.

4.3 Cultivate adolescents'social creativity

Amabile (1983, 1996) pointed out that children's creativity can be increased by allowing them to choose a task freely, by making the task interesting, and by increasing their intrinsic motivation; a compulsory task, the aim to win external praise, obvious control, and long-lasting pressure from peers must be avoided to maintain high creativity. The current findings provide support for this point of view and also take it a step further. To increase their social creativity, it is necessary to allow adolescents to choose a task freely, or at least to ensure they work of their own free will rather than following an external demand. The enjoyment of the task itself, its

relationship with adolescents' real life, and its significance will make adolescents complete the task out of interest or just for the sake of completing it. This should be emphasized when teachers wish to cultivate adolescents' social creativity or train them to solve social problems in a creative way.

To increase adolescents' social creativity, their creative personality, including curiosity, complexity, risk-taking, and imagination when solving various social problems, must be deliberately strengthened. As Sternberg (2003) commented, personality traits such as a willingness to overcome difficulties and take risks, tolerance of ambiguity, and high selfefficacy, motivate people to break away from conventions and think in a unique way. People with such personality traits, together with intrinsic motivation, have higher levels of social creativity. The cultivation of a creative personality, especially in children, is closely associated with a supportive environment and encouragement. Sternberg (2003) pointed out the importance of encouraging children to develop their own ideas, find their own interests and favorite activities, enhance their self-efficacy, and take advisable risks, which are also useful for cultivating social creativity. When facing social problems such as moods, time management, interpersonal relationships, and environmental protection, children should be encouraged to raise their own ideas after thinking and exploring independently, to solve problems bravely and rationally instead of escaping, and to find the best solution by analyzing all of the potential solutions rationally. Through this process, their creative self-efficacy will increase. Creative self-efficacy is the self-judgment of one's imaginative ability and perceived competence in generating novel and adaptive ideas, solutions, and behavior. It has been linked with positive beliefs and outcomes,including students' motivational beliefs and academic aspirations, their willingness to take intellectual risks in the classroom, creativity ratings from

supervisors, and teachers' ratings of creative expression (Beghetto, 2010). The formation of a creative personality will create a long-lasting impetus to improve social creativity.

4.4 Limitations and future research

The freedom of task choice in this research refers to the participants' willingness to accept and fulfill the task. Under the condition of free choice, the participants are considered free to choose and their behavior is intrinsically motivated. Otherwise, if the participants accept the task reluctantly due to outside pressure, they are not free to choose or have a psychological sense of being forced, then their behavior is externally motivated.

As Amabile (1983, 1996) pointed out, if the participants are allowed to choose materials for work (e.g., drawing) freely and actively instead of them being imposed by others, they can demonstrate higher creativity. In future research, the feeling of freedom in choosing the task for the participants in the experimental group could be improved by offering several equally effective tasks to evaluate social creativity. For the comparative group, the participants could be given no other choice except the task given by the experimenter.

In this study, although some students may have been limited in their creative expression by the written format of the assessment, the assessment was still appropriate for Chinese adolescents considering their greater level of inhibition. Chen et al. (1998) found that Chinese children were significantly more inhibited than their Western counterparts, which to some extent provides evidence for the appropriateness of the written format of the assessment because it may have helped the Chinese adolescents to generate creative expression more freely than a face-to-face interview. However, in future research it would be useful to compare the effect of the written format of the assessment with that of interviews on adolescents' creative expression.

In this study, the responses to each question were analyzed at an aggregate level (combined across situations) to measure the adolescents' overall social creativity. That is, the questions captured the common domains of creative action and were homogeneous. However, the domain of social creativity is broad, that is, there are probably multiple sub-domains within social creativity, such as leadership and interpersonal conflict. Future research, therefore, could focus on multiple sub-domains of social creativity and examine the effects of freedom of choice and general creative personality.

The study investigated the relationship between adolescents' social creativity and their gender, age, and whether they had siblings, and explored the difference in adolescents' social creativity between the two groups. However, it did not investigate the effect of freedom of task choice and creative personality and their interaction in each group separately. Future studies could analyze the moderation of demographic variables, especially whether the participants have siblings, on the relationship between freedom of task choice, creative personality, and social creativity.

The participants in this study were mainly from middle schools in China's capital city. In reality, there are still great gaps between cities and rural areas in China, especially in the level of economic development, so there might be considerable differences in the psychology of adolescents from different areas (Gu, Zhou, & Hu, 2009). In future studies, the participants could be sampled from both cities and the countryside to investigate differences between adolescents from areas with different levels of economic development. The sample could be further extended, for example by recruiting participants from China and other countries (e.g., North America) to investigate differences in the variables mentioned above under different cultural backgrounds.

Finally, as Baer (2008) noted, it is still uncertain whether creativity has domain specificity. In future research, the effect of the abovementioned variables on social creativity, scientific creativity, and artistic creativity could be explored together. As Feist (1999) argued, personality has different effects on scientific and artistic creativity. Most scientists are self-controlled, introverted, and independent, whereas artists are mostly impulsive and lack responsibility. Similarly, there is an inimitable tendency toward creative personalities in the social creativity field. Gu, Zhou, and Wang (2008) found that children with higher social creativity tended to be more outgoing, sociable, and easy-going, with lower psychoticism. Therefore, in different creative fields, the effect of the match between creative personality and task choice may be different, which is worth investigating further.

References

Aiken, L.S., & West, S.G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage Publications Inc.

Amabile, T.M. (1983). *The social psychology of creativity*. New York: Springer Verlag.

Amabile, T.M. (1996). *Creativity in context: Update to the social psychology of creativity. Boulder*, CO: Westview.

Amabile, T.M., & Gryskiewicz, N. (1989). The creative environment scales: The work environment inventory. *Crea-tivity Research Journal*, *2*, 231–254.

Amabile, T.M., Hill, K.G., Hennessey, B.A., & Tighe, E. (1994). The work preference inventory: Assessing intrinsic and extrinsic motivational orientations. *Journal of Personality and Social Psychology, 66*, 950–967.

Baer, J. (2008). Commentary: Divergent thinking tests have problems, but this is not the solution. *Psychology of*

Aesthetics, Creativity, and Arts, 2, 89–92.

Beghetto, R.A. (2010). Creativity in the classroom. In J.C. Kaufman & R.J. Sternberg (Eds.), *Handbook of Creativity*(pp. 447–463). Cambridge, UK: Cambridge University Press.

Byron, K., & Khazanchi, S. (2012). Rewards and creative performance: A meta–analytic test of theoretically derived hypotheses. *Psychological Bulletin, 138*, 809–830.

Chen, X., Hastings, P.D., Rubin, K.H., Chen, H., Cen, G., & Stewart, S.L. (1998). Child–rearing attitudes and behavioral inhibition in Chinese and Canadian toddlers: A cross–cultural study. *Developmental Psychology, 34*, 677–686.

Collins, M.A., & Amabile, T.N. (1999). Motivation and creativity. In R.J. Sternberg (Ed.), *Handbook of creativity*(pp. 297–312). New York: Cambridge University Press.

Fang, F., & Fang, G. (2005). *Children development psychology*. Peking: People's Education Press.

Feist, G.J. (1999). The influence of personality on artistic and scientific creativity. In R.J. Sternberg (Ed.), *Handbook of creativity* (pp. 273–296). New York: Cambridge University Press.

Feldman, D.H. (1999). The development of creativity. In R.J. Sternberg (Ed.), *Handbook of creativity* (pp. 169–186).New York: Cambridge University Press.

Feldman, R.S. (2005). Social and personality development. In R.S. Feldman (Ed.), *Development across the life span*(pp. 417–446). Asia: Pearson Education Asia Limited and Peking University Press.

Frauenknecht, M., & Black, D.R. (2004). Problem–solving training for children and adolescents. In E.C. Chang, T.J.D'Zurilla & L.J. Sanna (Eds.), *Social problem solving: Theory, research, training* (pp. 153–170). Washington, DC:American Psychological Association.

Gu, C. (2005). The systematic perspective of creativity and its enlightenment on education. *Educational Researchand Experiment, 3*, 51–55.

Gu, C. (2007). The experimental study on cultivation of primary school children's social creativity tendency. *Educational Research and Experiment, 5*, 65–69.

Gu, C. (2009). The development of social creative tendency of primary school children. *Studies of Psychology and Behavior, 7*, 198–202.

Gu, C. (2011). *Social creativity psychology*. Peking: China Social Science Press.

Gu, C., Huang, C., & Zhou, Z. (2008). The relationship between primary school children's social creativity and social skills and its enlightenment on education. *Educational Research and Experiment, 3*, 56–59.

Gu, C., Liu, Y., & Zhou, Z. (2010). The internal consistency and domain–specificity among primary school children.*Psychological Science, 33*, 616–620.

Gu, C., Zhang, H., & Zhou, Z. (2009). Relation between social creative potential of primary school children and teacher's leadership styles. *Chinese Journal of Clinical Psychology, 17,* 284–286.

Gu, C., & Zhou, Z. (2008). On the relationship between social creative tendency and parenting style among primary school children. *Psychological Development and Education, 2*, 34–38.

Gu, C., Zhou, Z., & Hu, J. (2009). The difference in social creative tendency between primary school children fromurban areas and their counterparts from rural areas and its implications to education. *Chinese Journal of Special Education, 8*, 85–88.

Gu, C., Zhou, Z., & Wang, F. (2008). The relationship of social creative tendency and personality attribute in primary school children. *Chinese Journal of Special Education, 3*, 91–96.

Gu, C., Zhou, Z., & Zhong, M. (2009). The relationship among social creativity, peer relationship and academic achievement of primary school children. *Psychological*

Development and Education, 25, 20–25.

Hennessey, B.A., & Amabile, T.M. (2010). Creativity. *Annual Review of Psychology, 61*, 569–598.

Jaccard, J., Turrisi, R., & Wan, C.K. (1990). Interaction effects in multiple regression (*Sage Series: Quantitative applications in the social sciences*). London: Sage.15 Journal of Creative Behavior

James, K., & Asmus, C. (2001). Personality, cognitive skills, and creativity in different life domains. *CreativityResearch Journal, 13*, 149–159.

Lin, C. (2009). *Research on creative persons and educational innovation*. Beijing: Economic Science Press.

Mischel, W., & Ayduk, O. (2002). Self–regulation in a cognitive–affective personality system: Attentional control in the service of the self. *Self and Identity, 1*, 113–120.

Mischel, W., & Shoda, Y. (1995). A cognitive–affective system theory of personality: Reconceptualizing situations,dispositions, dynamics, and invariance in personality structure. *Psychological Review, 102*, 246–268.

Mouchiroud, C., & Bernoussi, A. (2008). An empirical study of the construct validity of social creativity. *Learning and Individual Difference, 18,* 372–380.

Mouchiroud, C., & Lubart, T. (2002). Social creativity: A cross–sectional study of 6–to 11–year–old children. *International Journal of Behavioral Development, 26*, 60–69.

Niu, W., & Kaufman, J.C. (2013). Creativity of Chinese and American: A synthetic analysis. *Journal of CreativeBehavior, 47*, 77–87.

Prabhu, V., Sutton, C., & Sauser, W. (2008). Creativity and certain personality traits: Understanding the mediatingeffect of intrinsic motivation. *Creativity Research Journal, 20*, 53–66.

Runco, M.A. (1997). Introduction. In M.A. Runco & R.S. Albert (Eds.), *The creativity research handbook* (Vol. 1,pp. ix－xiv). Cresskill, NJ: Hampton Press.

Runco, M.A. (2004). Creativity. *Annual Review of Psychology, 55*, 657–687.

Runco, M.A., & Charles, R. (1997). Developmental trends in creativity. In M.A. Runco (Ed.), *Creativity researchhandbook* (Vol. 1, pp. 115–154). Cresskill, NJ: Hampton Press.

Shalley, C.E. (1995). Effects of coaction, expected evaluation, and goal setting on creativity and productivity. *Academy of Management Journal, 38,* 483–503.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,*420–428.

Simonton, D.K. (1990). Creativity and wisdom in aging. In J.E. Birren & K.W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd edn, pp. 320–329). San Diego, CA: Academic Press.

Simonton, D.K. (1999). Creativity from a historiometric perspective. In R.J. Sternberg (Ed.), *Handbook of creativity*(pp. 117–133). New York: Cambridge University Press.

Sternberg, R.J. (2003). *Wisdom, intelligence, and creativity synthesized*. New York: Cambridge University Press.

Sternberg, R.J., & Lubart, T.I. (1992). Buy low and sell high: An investment approach to creativity. *Human Current Directions in Psychological Science, 1*, 1–5.

Sternberg, R.J., & Lubart, T.I. (1999). The concept of creativity: Prospects and paradigms. In R.J. Sternberg (Ed.),*Handbook of creativity* (pp. 3–15). New York: Cambridge University Press.

Terman, L.M. (1906). Genius and stupidity: A study of some of the intellectual processes of seven 'bright' andseven 'stupid' boys. *Pedagogical Seminary, 13,* 307–373.

Torrance, E.P. (1960). An experimental attempt to increase quantity and quality of ideas. In E.P. Torrance, J. E. Bowers,H. J. Rading, N. Palamatu, & P. R. Krishnarah (Eds.), *The Minnesota studies of creative thinking in the early schoolyears* (pp. 30–31). Research Memorandum BERR–60–1. Minneapolis, MN: University of Minnesota.

Torrance, E.P. (1962). *Guiding creative talent*. Englewood, NJ: Prentice Hall Inc.

Torrance, E.P. (1964). Education and creativity. In C.W. Taylor (Ed.), *Creativity: Progress and potential*. New York: McGraw–Hill.

Williams, F.E., Lin, X., & Wang, M. (1994). *Williams creativity assessment packet*. Taipei: Psychology Press.

Yi, X., Hu, W., Plucker, J.A., & McWilliams, J. (2013). Is there a developmental slump in creativity in China? Therelationship between organizational climate and creativity development in Chinese adolescents. *Journal of Crea–tive Behavior, 47*, 22–40.

Appendix

OPEN QUESTIONS OF SOCIAL PROBLEM SITUATIONS

Hello! In everyday life, students are often confronted with difficult situations that demand solutions. Some of the situations faced by students are listed below; please say what the students should do under such conditions. The questionnaire is anonymous, so don't worry about your responses. Please read the instructions carefully before answering the questions. Just reply according to your first feeling to each item, and do not miss any question. There is no fixed sequence for all the items; you can start from any question you want, but you have to finish all of them.

Are you interested in answering these questions? a. Yes; b. No, but I will answer them for the sake of completing the task

◇ Zezhi is confused because he is unable to arrange time well recently. On one side, he wants to take part in varieties of activities he is more interested in; on the other side, he realizes he does not have enough time for his learning. What should he do? (Please give as many solutions as possible to help him; if your solutions are more than 4, just write

them down successively.)

Solution 1	
Solution 2	
Solution 3	
Solution 4	

◇ Nade has trouble in controlling his recent emotions; he feels unhappy, obsessed by indescribable moods that are out of control. What should he do? (Please give as many solutions as possible to help him; if your solutions are more than 4, just write them down successively.)

Solution 1	
Solution 2	
Solution 3	
Solution 4	

◇ Xizhe is worried that he can't understand the lectures in class. Though he can sometimes understand, he cannot work out the questions given by the teacher. What should he do? (Please give as many solutions as possible to help him; if your solutions are more than 4, just write them down successively.)

Solution 1	
Solution 2	
Solution 3	
Solution 4	

◇ Hengxu thinks that he has few good friends, and he wants to make more friends. What should he do? (Please give as many solutions as possible to help him; if your solutions are more than 4, just write them down successively.)

Solution 1	
Solution 2	
Solution 3	

Solution 4

◇ Zhehao fell in love with a classmate. However, the classmate's parents object to their relationship. What should he do? (Please give as many solutions as possible to help him; if your solutions are more than 4, just write them down successively.)

Solution 1

Solution 2

Solution 3

Solution 4

◇ Zifan now lives with several roommates in the same dormitory, but they have different habits, and Zifan is upset by some roommates' bad habits (for example, some roommates like to be on a phone call until late at night thereby disturbing Zifan's sleep).What should he do? (Please give as many solutions as possible to help him; if your solutions are more than 4, just write them down successively.)

Solution 1

Solution 2

Solution 3

Solution 4

Inhibition ability of food cues between successful and unsuccessful restrained eaters: A two-choice oddball task

Fanchang Kong[1,2], Yan Zhang[3], Hong Chen[2,*]

([1]Key Laboratory of Adolescent CyberPsychology and Behavior (Ministry of Education) and School of Psychology, Central China Normal University, Wuhan, China)

([2]Key Laboratory of Cognition and Personality (Ministry of Education) and Faculty of Psychology, Southwest University, Chongqing, China)

([3] Academy of Educational Science, Huazhong University of Science & Technology, Wuhan. China)

Abstract　Background：Previous studies have presented mixed findings on the inhibition ability in restrained eaters (REs) due to the limited amount of neural evidence and limitations of behavioral measures. The current study explores the neural correlations of the specific inhibition ability among successful restrained eaters (S–REs), unsuccessful restrained eaters (US–REs), and unrestrained eaters (UREs). Methodology and Principal Findings: Three groups of females (with 13 participants in each group) completed a two–choice Oddball task, while the event–related potentials (ERPs) are recorded synchronously. Results indicate that S–REs showed inhibition deficit in processing high–energy food cues whereas US–REs show inhibition deficit in processing both low– and high–energy food cues. Conclusion: Results indicate that S–REs and US–REs differ in terms of specific inhibition ability and that enhanced inhibition is essential to a successful diet.

Keywords　restrained eaters; food cues; ERP; specific inhibition

1 Introduction

Self–control is defined as the exertion of control over the self by the self , which reflects the ability to change one's responses, and accommodates an individual's reaction to the behavioral standard to support the pursuit of a long–term goal. Numerous studies have shown that self–control and self–inhibition have significant roles in successful weight control and food intake. Longitudinal studies confirmed that initial inhibition ability predicted weight gain after one year and overweight after three years Food exposure successfully evoked self–regulation in normal–weight REs but not for the overweight REs because the latter ignored the diet goal. Therefore, systematically

examining the inhibition ability of REs would be beneficial in uncovering the mechanism of a successful diet.

Behavioral studies showed that compared with UREs, REs reported a higher level of compulsive or lower level of inhibition ability. Participants in the impulsive group had a significantly higher caloric intake during a subsequent taste test, whereas those under the inhibition group did not differ from the control group. Therefore, impulsivity is a direct cause of overeating . Women who scored higher in uncontrolled eating (TFEQ-D) were significantly more impulsive on the Barrett Impulsivity Scale-11 (BIS-11), but not on the delay discounting task (DDT) and the go/no-go task. Meanwhile, the women who exhibited higher restraint in eating (TFEQ-R) scores were significantly less impulsive on the go/no-go, task but did not differ in terms of the BIS-11 and the DDT. Therefore, some aspects of overeating may be caused by a poor ability to reflect on decisions. High REs did not inhibit their eating behaviors when faced with highly palatable food and appeared to overeat, a reaction that is expected to result in obesity.

Newer studies showed that the interaction between dietary restraint and trait impulsiveness predicted dieting success. Specifically, a lower level of impulsiveness is associated with greater dieting success among REs, suggesting that less impulsive REs are more likely to become S-REs. However, some studies have reported contrasting findings. For example, REs commit fewer errors compared with UREs regardless of food or nonfood cues. Meule, Vüele, and Kübler (2012) showed that REs did not differ from UREs in terms of resistance to distractor interference, although REs reacted faster to the high-calorie food cues than the neutral pictures. This result indicates that REs demonstrate an attention bias toward high-calorie food cues, which is related to low dieting success. Furthermore, the latest studies showed that impulsive reactions to high-calorie food cues are pronounced when both trait impulsivity and food craving are high. However, low levels of impulsivity can compensate for high levels of trait food craving. Thus, REs probably showed inhibition deficit in the specific stimuli (food cues), but not in the general stimuli.

Some studies used cognitive neuroscience technology, such as event-related potentials (ERPs), to examine the neural correlates of the REs' inhibition ability. Hachl, Hempel, and Pietrowsky (2003) found that REs showed no significant difference in neural correlates, but not in behavioral response. Specifically, larger P2 amplitude of REs was comparable to UREs in the hungry condition, but larger P2 amplitude was comparable to UREs in full condition. In addition, larger late positive potential (LPP) for REs was comparable to UREs in both food and nonfood conditions, suggesting that REs require more cognitive resources to inhibit the interference of unrelated information in processing a target stimuli. In another ERP study, the REs and contrast groups completed two blocks . In the first block, emotional pictures and pictures of high-energy food were presented, and the participants were informed that they were allowed to eat half of the food presented in the first block and that the other half was not available by manipulating the availability. Results showed no significant differences between the REs and the contrast group in the first block. However, a significant difference was observed in the second block. For REs, the non-available food cues elicited more positive LPP than the available food cues, but not for the control group. Therefore, the REs successfully conducted cognitive control on the motivation tendency based on the motivational background to reduce the sensibility to the food cues and attain affirmative results in dieting.

Furthermore, Babiloni, Del Percio, Triggiani, Marzano, Valenzano, Petito et al. (2011) used the standard oddball task to explore the temporal processing between the normal-weight successful dieters and non-dieters. The

results showed that larger frontal–parietal P3 were evoked by food pictures among normal–weight successful dieters than in non–dieters. A new study, which employed a go/no–go task to investigate the neurocognitive correlates of processing food–related stimuli, showed that food images elicited significantly enhanced N2, P3, and slow–wave ERP components compared with the nonfood images, suggesting that processing food–related stimuli elicited distinct patterns of cortical activity. Therefore, a probable difference existing between REs and UREs in terms of the neural correlates of processing food stimuli, such as P2, N2, and P3 components.

In summary, the existing literature showed contrasting results in inhibition ability of REs to food cues importantly because limited neural studies were found, specifically on the dynamic temporal process in REs. Previous studies showed that REs comprise successful restrained eaters (S–REs) and unsuccessful restrained eaters (US–REs). The difference between the neural mechanism of S–REs and US–REs showed significant risks for obesity and other eating–related problems. Therefore, the present study investigated the neural correlates of inhibition ability to food cues in a two–choice oddball task using high time–resolution ERPs. Considering the abovementioned

findings, we hypothesized that S–REs and US–REs differed in N2, P2, and P3 components compared with UREs. Furthermore, S–REs and US–REs showed stronger N2, P2, and P3 components in processing food cues compared with neural cues.

2 Method

2.1 Ethics statement

This study has been approved by the Ethics Committee of the Faculty of Psychology of Southwest University. All participants have signed an informed consent prior to their inclusion in our experiments. The study was designed in accordance with tenets of the Declaration of Helsinki.

2.2 Participants

The final sample included 13 S–REs, 13 US–REs, and 13 UREs from Southwest University in Chongqing. Based on the operationalization described in previous studies , the S–REs who scored more than 2.7 on the restrained eating subscale of the Dutch Eating Behavior Questionnaire (DEBQ), less than 3.0 on the emotional and external eating subscales of the DEBQ, less than 1.8 on the uncontrolled eating subscale of the Three–Factor Eating Questionnaire (TFEQ), and had normal BMI for

Table 1. Descriptive statistics and ANOVA analysis on the grouping variables

Variable	Group	M SD	F & p value
BMI	S–RE	20.22 ± 1.88	
	US–RE	19.62 ± 2.05	$F_{(2,39)}$=0.335, p =0.718
	URE	19.89 ± 1.92	
DEBQ_R	S–RE	26.29 ± 3.05	
	US–RE	26.57 ± 10.48	$F_{(2,39)}$=21.485, p =0.001
	URE	10.50 ± 6.80	
DEBQ_EM	S–RE	20.64 ± ± 8.62	
	US–RE	33.57 ± 16.89	$F_{(2,39)}$= 3.559, p =0.038
	URE	22.07 ± 15.28	
TFEQ_U	S–RE	16.93 ± 3.03	
	US–RE	20.86 ± 2.38	$F_{(2,39)}$= 8.937, p =0.007
	URE	15.71 ± 4.38	

Note: BMI= body mass index, DEBQ_R= Restrained Eating subscale of Dutch Eating Behavior Questionnaire, DEBQ_EM=Emotional and External Eating subscales of DEBQ, TFEQ_U=Uncontrolled eating subscale of Three Factor Eating Questionnaire.

more than six months.US−REs scored more than 2.7 on the Restrained Eating subscale of DEBQ, more than 3.0 on the Emotional Eating subscale and External Eating subscale of DEBQ, more than 1.8 in the Uncontrolled eating subscale of TFEQ, with had a normal BMI for more than 6 months. UREs never or infrequently dieted, scored less than 1.6 on the Restrained Eating subscale of DEBQ, with normal BMI (Table 1).

All participants were healthy right−handed young women ranging in age from 18 to 23 (M=19.3, SD=0.3) and they were Han nationality. None of them reported having physical or psychiatric conditions, an eating disorder, or taking medication, smoking recreational drug or alcohol consumption during the past two years. Furthermore, all participants had a body mass index (BMI) within the average range (M=20.13, SD=1.05), and no significant difference between S−REs, US−RE and UREs (F(2,39)=0.335, p=0.718). Participants were paid 50 RMB for their participation.

2.3 Materials and procedure

A two−choice Oddball task was employed in the second experiment. Standard stimulus was a picture of steel clock, deviant stimuli included 30 high−energy food pictures, 30 low−energy food pictures and 30 neutral pictures. High−energy food pictures differed from low−energy food pictures in the dimension of food content ($t_{(57)}$=23.472, p=0.001), but not in the dimensions of arousal ($F_{(2,87)}$=0.239, p=0.788), happiness ($F_{(2,87)}$=2.162, p=0.121) and familiarity ($F_{(2,87)}$=2.767, p=0.068). Each picture was identical in size (400 by 400 pixels), resolution (72 dots per inch), brightness, and background. Additional 15 pictures (five for each category) were selected in the practice session and not used in the normal session.

There were 6 blocks, with 120 trials in each block. In each block, stand stimuli was 75 percent and deviant stimuli was 25 percent. Trials in each block were presented randomly. In each trial, a black fixation appeared 300 ms

in the gray screen, followed 500−1000 ms grey screen, subsequent target picture presented 1200 ms, and finally a grey screen appeared 1000ms before the next trial. In each group, half of the participants were required to press the "F" key when the standard stimuli presented, and press the "J" key when the deviant stimuli presented. Half of participants were asked to press the opposite key when the standard or deviant stimuli presented . The distance from subjects to screen center was 0.9 m, the visual angle of the participant was less than 6° Levels of hungry/fullness and emotional state were measured by self−report in the Likert scale from "0" (not at all) to "100" (very much) pro− and post−experiment. Moreover, the participant was not required to eat any food but water (250ml) within 3 hours prior to the study.

2.4 ERP recording and analysis

EEG was recorded from 64 scalp sites using tin electrodes mounted in an elastic Ag/AgCl cap, with the linked reference on the left and right mastoids, and a ground electrode was placed on the medial frontal aspect (Brain products, German). Eye movements were monitored with supra− and infra−orbital electrodes and with electrodes on canthi. A bandpass of 0.01−100Hz was used for EEG and electro−oculogram (EOG). Electrode impedance was maintained below 5kΩ. After rejecting trials with eye movements, blink, motion or other artifacts at each channel, the averaging of ERPs was computed off− line with computer algorithms. Trials with EOG artifacts with peak−to peak deflection exceeding ±80μV and those contaminated with artifacts were excluded from averaging.

In the experiment, only responses to the deviant stimuli were analyzed due to our hypothesis. The resulting grand averages were based on the correct trials and the averaging epoch was 1000 ms including a 200 ms pre− onset baseline. According to available findings and the grand averaging ERPs, high−energy food pictures, low−

energy food pictures and neutral pictures were significantly different in the P2 (130–205 ms), N2 (210–285 ms) and P3 (300–500 ms) components. The following cortexes were selected, frontal (F3, F4, Fz), frontal–central (FC3, FC4, FCz), central (C3, C4, Cz), central–parietal (CP3, CP4, CPz), parietal (P3, P4, Pz) and occipital cortex (O1,O2,O3). Repeated measures ANOVAs were conducted on the amplitude (baseline to peak) and peak latency (stimulus onset to peak), with group (S–REs, US–REs, UREs) as a between factor, and picture (high–energy food picture, low–energy food picture, neutral picture) and cortex as within factors. All the analysis was conducted by SPSS 17.0. P–value was computed for deviation in all analysis based on the Greenhouse–Geisser method. In addition, the Bonferroni method was used for the post–hoc comparisons.

3 Results

3.1 Behavioral data

Results showed there was no significant difference in levels of hungry and fullness pre– （$F_{(2,39)}$=0.889, p=0.419; $F_{(2,39)}$=0.640, p=0.533) and post–experiment （$F_{(2,39)}$=0.378, p=0.688; $F_{(2,39)}$=3.197, p=0.052) in S–REs, US–REs and UREs (See Table 2). Trials with incorrect responses, too

fast (RTs than 200 ms) or too slow (RTs more than 1000ms) were excluded. On the mean reaction times (RT), 3 (picture: High–energy food pictures, low–energy food pictures, neutral pictures) × 3(group: S–REs, US–REs, UREs) repeated measures ANOVA showed main effect of picture ($F_{(1.806, 70.439)}$= 18.437, p=0.001). Extra main or interaction effect was not significant. Post–hoc test showed that the RTs of high–energy food picture was longer compared to the low–energy food picture and neutral picture ($t_{(38)}$=5.577, p=0.001; $t_{(38)}$=4.524, p=0.001). Furthermore, no significant effects on the accuracy were found.

3.2 ERP data

Results of repeated measures ANOVA of 3 (picture: High–energy food pictures, low–energy food pictures, neutral pictures)× 3(group: S–REs, US–REs, UREs) ×4 (cortex: Frontal F3, F4, Fz; frontal–central FC3, FC4, FCz; central C3, C4, Cz; central–parietal CP3, CP4, CPz) on P2 amplitude showed main effects of picture ($F_{(1.089, 42.486)}$=7.531, p=0.007) and cortex ($F_{(1.580, 61.616)}$=7.220, p=0.003). The post–hoc test showed high–energy and low energy food pictures elicited smaller P2 amplitude compared to the neutral pictures ($t_{(38)}$=4.358, p=0.001; $t_{(38)}$=3.203, p=0.035), and high–energy food pictures were not significantly different from low–energy food pictures

Table 2. The descriptive and AVOVA results on the hungry/fullness level pre— and post—experiment

Hungry& fullness	Group	M ± SD	F value & p value
H_Pr	S–RE	38.93 ± 5.35	
	US–RE	49.14 ± 6.60	$F_{(2,39)}$=0.889, p=0.419
	URE	39.29 ± 6.42	
F_Pr	S–RE	45.79 ± 6.88	
	US–RE	46.43 ± 5.80	$F_{(2,39)}$=0.640, p=0.533
	URE	54.93 ± 6.48	
H_Po	S–RE	57.43 ± 2.26	
	US–RE	63.93 ± 5.80	$F_{(2,39)}$=0.378, p=0.688
	URE	63.57 ± 5.51	
F_Po	S–RE	56.43 ± 6.84	
	US–RE	37.50 ± 5.26	$F_{(2,39)}$= 3.197, p=0.052
	URE	37.50 ± 6.13	

Note: H_Pr=the hungry state pre–experiment; F_Pr=the fullness state pre–experiment; H_Po=the hungry state post experiment; F_Po=the fullness state post experiment.
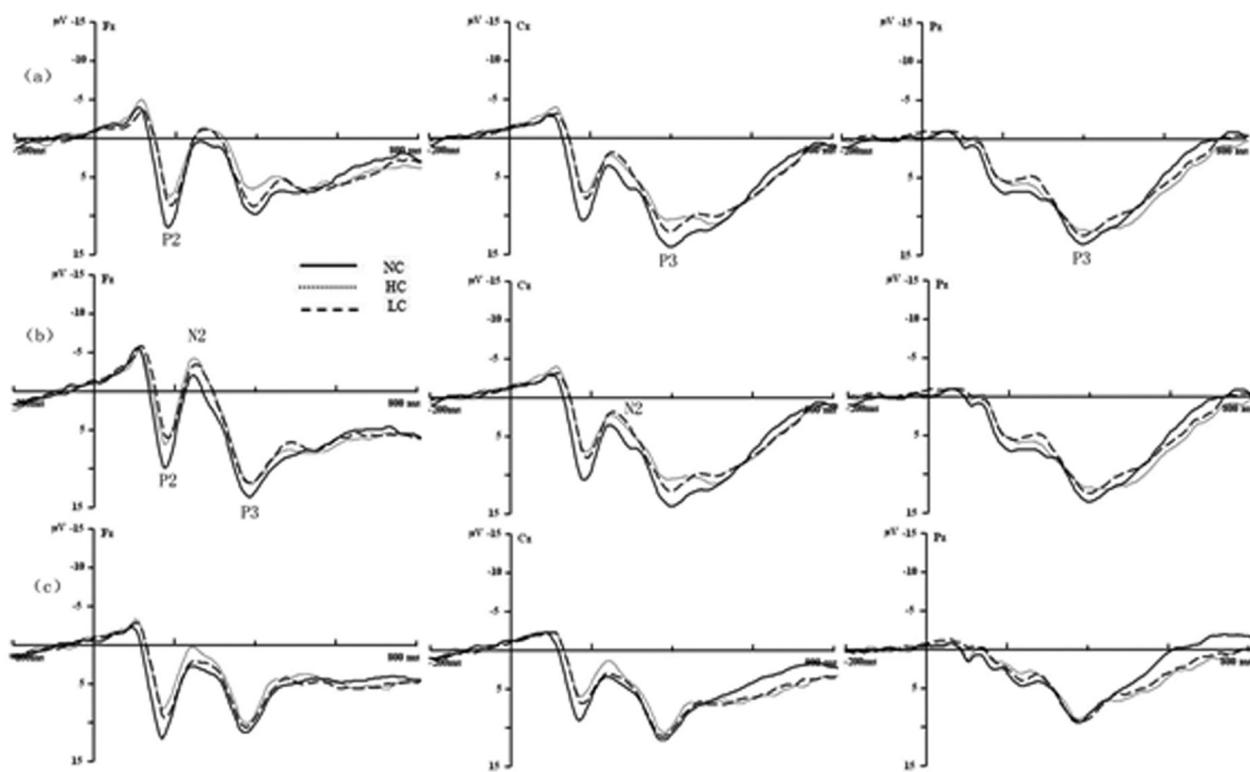
Fig. 1 Grand average waveform of the stimulus-locked ERPs for P2, N2 and P3 at Fz, Cz and Pz in S-RE (a), US-RE (b) and URE(c)

on the P2 amplitude. Post-hoc test on the cortex showed that P2 amplitude in the frontal cortex were larger than the central cortex ($t_{(37)}$=2.853, p=0.041), P2 amplitude in the frontal-central cortex was larger than the central and parietal cortex ($t_{(37)}$=4.132, p=0.001; $t_{(37)}$=3, p=0.028). Moreover, interaction of picture and cortex was also significant ($F_{(2.363,92.145)}$=61.452, p=0.001)(Fig1,Table 3). Simple effect analysis showed for the high-energy food pictures, P2 amplitude in the frontal-central cortex was larger than in frontal, central and central-parietal cortex ($F_{(3,37)}$=11.284, p=0.001; $F_{(3,37)}$=8.120, p=0.001; $F_{(3,37)}$=11.942, p=0.001). For the low-energy food pictures, P2 amplitude in the frontal-central cortex was larger than in frontal-central, central and central-parietal cortex ($F_{(3,37)}$=8.777, p=0.001; $F_{(3,37)}$=11.497, p=0.001; $F_{(3,37)}$=2.863,p=0.007). For the neutral pictures, P2 was larger in the central cortex than the frontal, frontal-central and central-parietal cortex ($F_{(3,37)}$=4.824, p=0.001; $F_{(3,37)}$=9.412, p=0.001; $F_{(3,37)}$=8.192, p=0.001).

Similar repeated measures ANOVA was conducted on the P2 latency, and results showed main effects of picture ($F_{(1.259, 49.099)}$=38.674, p=0.001) and cortex ($F_{(1.937, 75.546)}$=7.546, p=0.001). The interaction effects of picture and cortex ($F_{(3.518, 49.099)}$=2.528, p=0.047), picture and group($F_{(2.917,113.746)}$=18.049, p=0.001), were also significant. Post-hoc test showed P2 latency in frontal cortex was shorter than in frontal-central, central-parietal cortex ($t_{(37)}$=3.142, p=0.020; $t_{(37)}$=3.065, p=0.024; $t_{(37)}$=4.381, p=0.001), and P2 latencies of high-energy food and low-energy food pictures were shorter compared to the neutral pictures ($t_{(38)}$=6.284,p=0.001; $t_{(38)}$=7.034, p=0.001). Simple effect analysis on the interaction of picture and group showed P2 latency of high-energy food pictures was shorter compared to the neutral pictures ($F_{(2,38)}$=4.391, p=0.001) in S-REs. P2 latencies of high-energy food and low-energy food pictures were shorter in contrast to the neutral pictures($F_{(2,38)}$=4.765, p=0.001; $F_{(2,38)}$=4.744, p=0.001), and P2 latency of the high-energy food picture

Table 3. The differences of P2, N2 and P3 components in S–RE, U–RE and UE at Fz, Cz and Pz electrodes

Cortex	Picture	Group	P2（M±SD) Latency	Amplitude	N2（M±SD) Latency	Amplitude	P3（M±SD) Latency	Amplitude
Fz	HC	S–RE	173.60 ± 12.61	7.25 ± 4.97	254.67 ± 20.60	−3.32 ± 5.10	427.38 ± 62.82	9.47 ± 5.96
		US–RE	165.64 ± 6.74	7.42 ± 6.39	244.55 ± 23.60	−6.61 ± 6.81	401.69 ± 66.21	14.21 ± 8.44
		URE	170.92 ± 15.97	9.21 ± 5.99	252.00 ± 23.44	−2.24 ± 7.13	373.54 ± 23.68	11.42 ± 6.22
	LC	S–RE	176.27 ± 9.35	11.59 ± 5.62	251.60 ± 25.75	−1.88 ± 4.72	400.15 ± 48.56	11.26 ± 4.15
		US–RE	170.55 ± 8.63	11.08 ± 6.51	236.55 ± 21.45	−3.10 ± 7.62	388.15 ± 56.29	15.42 ± 9.65
		URE	176.77 ± 15.50	12.91 ± 6.33	251.85 ± 25.53	−.75 ± 6.83	373.38 ± 40.32	12.95 ± 7.21
	NC	S–RE	176.67 ± 17.88	8.24 ± 5.27	253.73 ± 17.42	−2.96 ± 4.69	409.85 ± 56.22	11.68 ± 6.47
		US–RE	169.64 ± 15.33	6.80 ± 5.22	247.82 ± 25.91	−4.91 ± 6.73	368.92 ± 29.83	10.59 ± 5.35
		URE	172.92 ± 19.25	10.38 ± 5.59	256.62 ± 22.17	−.98 ± 6.39	367.23 ± 19.55	13.45 ± 7.89
Cz	Hc	S–RE	172.27 ± 15.94	7.42 ± 5.17	242.53 ± 22.88	.82 ± 4.83	436.00 ± 55.13	11.98 ± 6.70
		US–RE	172.18 ± 22.35	7.89 ± 5.09	238.55 ± 24.46	−1.00 ± 4.33	398.00 ± 66.50	13.90 ± 6.76
		URE	169.54 ± 17.51	8.52 ± 4.79	237.08 ± 23.19	−.09 ± 6.13	378.15 ± 38.89	19.07 ± 4.38
	LC	S–RE	170.13 ± 15.46	11.53 ± 5.30	233.73 ± 20.04	2.45 ± 4.59	419.54 ± 47.87	15.03 ± 6.51
		US–RE	163.82 ± 17.54	11.68 ± 4.87	235.27 ± 26.48	1.95 ± 4.64	378.15 ± 46.99	16.23 ± 4.48
		URE	169.85 ± 18.81	11.68 ± 5.00	239.38 ± 26.31	1.31 ± 5.56	365.54 ± 19.94	20.48 ± 7.46
	NC	S–RE	171.60 ± 14.25	7.86 ± 5.26	243.33 ± 18.09	.75 ± 4.34	411.23 ± 47.86	16.67 ± 6.36
		US–RE	164.55 ± 12.84	7.67 ± 4.14	241.45 ± 28.14	−.04 ± 3.68	388.46 ± 58.58	14.29 ± 5.34
		URE	166.15 ± 13.99	9.30 ± 4.25	254.92 ± 28.43	.80 ± 5.24	387.85 ± 52.88	18.22 ± 6.41
	HC	S–RE	–	–	–	–	416.46 ± 63.59	15.08 ± 5.81
		US–RE	–	–	–	–	410.00 ± 66.49	14.08 ± 1.95
		URE	–	–	–	–	378.15 ± 42.16	20.23 ± 4.64
Pz	LC	S–RE	–	–	–	–	413.08 ± 49.86	14.94 ± 5.60
		US–RE	–	–	–	–	366.15 ± 36.42	14.88 ± 2.35
		URE	–	–	–	–	363.85 ± 18.27	21.25 ± 5.20
	NC	S–RE	–	–	–	–	388.00 ± 32.37	15.62 ± 5.71
		US–RE	–	–	–	–	383.54 ± 38.94	14.16 ± 2.65
		URE	–	–	–	–	366.46 ± 27.76	19.50 ± 4.69

Note: HC=high–energy food picture, LC=low–energy food picture, NC=Neutral picture.

was shorter than low–energy food picture in US–REs ($F_{(2,38)}$=2.492, p=0.017). P2 latency of the high–energy food was shorter than the neutral picture in UREs ($F_{(2,38)}$=2.256, p=0.030). Simple effect analysis on the picture and cortex showed that for the high–energy food pictures, P2 latency was shorter in frontal–central cortex compared to frontal, central and central–parietal cortex ($F_{(3,37)}$=2.897, p=0.006; $F_{(3,37)}$=8.499, p=0.006; $F_{(3,37)}$=3.466, p=0.001). For the low–energy food pictures, P2 latency was shorter in frontal cortex compared to frontal–central and parietal cortex ($F_{(3,37)}$=7.336, p=0.001; $F_{(3,37)}$=3.900, p=0.001). For the neutral pictures, P2 was shorter in the central cortex compared to the frontal, frontal–central and central–

parietal cortex ($F_{(3,37)}$=2.181, p=0.035; $F_{(3,37)}$=2.940, p=0.005; $F_{(3,37)}$=5.408, p=0.001).

For N2 amplitude, 3 (picture: High–energy food pictures, low–energy food pictures, neutral pictures) × 3(group: S–REs, US–REs, UREs) × 4(cortex: Frontal F3, F4, Fz; frontal–central FC3, FC4, FCz; central C3, C4, Cz; central–parietal CP3,CP4,CPz) repeated measures ANOVA showed main effects of picture and cortex ($F_{(1.226,47.806)}$=12.603, p=0.001; $F_{(1.226,47.806)}$=12.603, p=0.001), and interaction of picture and cortex ($F_{(1.226,47.806)}$=12.603, p=0.001) (Fig1 and Table 3). Post–hoc test showed the high– and low–energy food pictures elicited more negative N2 amplitude than the neutral pictures ($t_{(38)}$=6.779,

p=0.001; t$_{(38)}$=5.197, p=0.001), the high−energy food picture elicited larger N2 amplitude compared to the low− energy food pictures ($t_{(38)}$=7.940, p=0.001). N2 amplitude was smaller in the frontal than in the frontal−central, central and central−parietal cortex (t$_{(38)}$=7.114, p=0.001; t$_{(38)}$=7.069, p=0.001; $t_{(38)}$=6.711, p=0.001). Simple effect analysis showed that for the high−energy food pictures, N2 amplitude in frontal cortex was larger than central− parietal cortex (F$_{(3,37)}$=7.009, p=0.001). For the low−energy food pictures, N2 amplitude was larger in the central− parietal cortex than frontal cortex (F$_{(3,37)}$=5.891, p=0.001), frontal−central cortex(F$_{(3,37)}$=6.592, p=0.001) and central cortex (F$_{(3,37)}$=4.694, p=0.001). For the neutral pictures, N2 amplitude was larger in central−parietal cortex than frontal cortex, frontal−central and parietal cortex (F$_{(3,37)}$=4.694, p=0.001; F$_{(3,37)}$=2.181, p=0.035; F$_{(3,\ 37)}$=2.075, p=0.041). Results from N2 latency showed a main effect of picture (F$_{(1.226,\ 47.806)}$=12.603, p=0.001), and interaction effect of picture and cortex (F$_{(3.169,\ 123.589)}$=7.576, p=0.001). Post−hoc test showed that N2 latency of the high−energy food picture was longer than the low−energy food pictures and neutral pictures ($t_{(38)}$=4.932,p=0.001;$t_{(38)}$=3.736, p=0.002). Simple effect analysis showed that N2 latency of the high−energy food picture was shorter in frontal−central cortex than central cortex (F$_{(3,37)}$=3.292, p=0.002). For the low−energy food pictures, N2 latency was shorter in central−parietal cortex compared to frontal−central cortex (F$_{(3,37)}$=4.348, p=0.001). For neutral pictures, N2 latency was shorter in central cortex compared to frontal and central−parietal cortex (F$_{(3,37)}$=2.850, p=0.006;F$_{(3,37)}$=2.050, p=0.047).

On P3 amplitude, repeated measures of 3 (picture: High−energy food pictures, low−energy food pictures, neutral pictures) × 3(group: S−REs, US−REs, UREs) × 5 (cortex: Frontal F3, F4, Fz; frontal−central FC3, FC4, FCz; central C3, C4, Cz; central−parietal CP3, CP4, CPz; parietal P3, P4, Pz) showed main effects of picture and cortex (F$_{(1.118,43.596)}$=10.240, p=0.002;

F$_{(1.518,59.191)}$=8.435, p=0.002), interaction of picture and group (F$_{(2.762,43.596)}$=11.478, p=0.001). Post−hoc test showed P3 amplitude in the parietal cortex was larger than the frontal, frontal−central, central and parietal cortex ($t_{(36)}$=1.710, p=0.001; $t_{(36)}$=2.294, p=0.001; $t_{(36)}$=43.421, p=0.001; $t(36)$=5.214, p=0.001), and the low−energy food pictures and neutral pictures elicited larger P3 amplitude than the high−energy food pictures ($t_{(38)}$=3.735, p=0.002; $t_{(38)}$=3.263, p=0.007). Simple effect analysis showed that in S−REs, the low−energy food pictures and neutral pictures elicited larger P3 amplitude than the high−energy food pictures (F$_{(2,38)}$=2.398, p=0.021; F$_{(2,38)}$=2.701, p=0.010), the neutral pictures elicited larger P3 amplitude than the low−energy food pictures (F$_{(2,38)}$=2.578, p=0.014). For US− REs, the neutral pictures elicited larger P3 amplitude than the high− and low−energy food pictures (F$_{(2,38)}$=2.387, p=0.032; F$_{(2,38)}$=2.163, p=0.047). For UREs, there was no significant difference in the picture (Fig. 1 and Table 3).

Similar repeated measures ANOVA was conducted on the P3 latency, and results only showed interaction effect of picture and cortex (F$_{(3.403,132.724)}$=4.595, p=0.003). Simple effect analysis showed that for the high−energy food pictures, P3 latency was longer in central−parietal cortex compared to the frontal−central, central and parietal cortex (F$_{(4,36)}$=2.255, p=0.030; F$_{(4,36)}$=2.263, p=0.029; F$_{(4,36)}$=2.397, p=0.021). For the low−energy food pictures, P3 latency was longer in parietal cortex compared to central and central−parietal cortex (F$_{(4,36)}$=2.351, p=0.024; F$_{(4,36)}$=2.048, p=0.048). For the neutral pictures, P3 amplitude was longer in the central cortex than the frontal, frontal−central, central−parietal and parietal cortex (F$_{(4,36)}$=2.322, p=0.026; F$_{(4,36)}$=2.164, p=0.037; F$_{(4,36)}$=3.317, p=0.002; F$(4,36)$=2.918, p=0.006).

4 Discussion

4.1 Behavioral response

The present study shows that S−REs and US−REs

did not behaviorally differ in specific inhibition ability, which is not consistent with available findings that REs are deficient in terms of their specific inhibition ability . This finding, however, was consistent with the findings of Hachl, Hempel, and Pietrowsky (2003). The probable interpretation is that S–REs and UREs intentionally allow for the regulation of the inhibition ability.

4.2 Electrophysiological response

Results showed that the P2 latency of the high–energy food picture was shorter than that of the neutral pictures in S–REs. The P2 latencies of the high– and low–energy food pictures were shorter compared with that of the neutral pictures in US–REs, which tested our hypothesis. P2 indicated the perception analysis of the stimuli attributes in the brain. Thus, S–REs rapidly completed the perception analysis of high–energy food pictures, while US–REs easily finished the perception analysis of both high– and low–energy food pictures. As such, S–REs were found to be more sensitive to high–energy food and US–REs were more sensitive to food information regardless of energy due to their long dieting activity. Moreover, the P2 amplitudes of the high– and low–energy food pictures were smaller than the neutral pictures, which were related to the survival value of the food cues.

Moreover, results reflected that S–REs and US–REs show no difference in the N2 component, although the difference of N2 amplitude in S–REs, US–REs, and UREs reached a numerical edge. This finding does not support our hypothesis, but is consistent with Watson and Garvey's (2013) study, wherein food cues elicited larger N2 component than did neutral cues in both REs and UREs. N2 reflected the conflict detection, and the closer to the pre–frontal cortex, the stronger the conflict monitor, indicating that S–REs did not differ from US–REs in conflict monitor the target stimuli. This result was not consistent with previous findings, which was probably caused by the fact that S–REs and US–REs scored high

in the dimension of restraint dietary and presumably performed normally in daily eating behaviors except for some specific conditions. Another probable reason is that both S–REs and US–REs are chronic dieters and have developed an enhanced sensitivity to food cues because of the long–term diet activity.

Furthermore, the P3 amplitudes of the low–energy and neutral pictures were found to be larger than the high–energy food pictures, the P3 amplitudes of the neutral pictures were larger than the low–energy food pictures in S–REs, and P3 amplitude of the neutral pictures was larger than those of the high–energy and low–energy food pictures in US–REs, which is not consistent with previous research.

Given that the tasks were different from that in Watson and Garvey's (2013) study and similar to that in Babiloni et al. (2011), the current study's results are analyzed along with that of Babiloni et al. (2011). Babiloni et al. (2011) showed similar P3 amplitude evoked by the food pictures between normal–weight successful dieters and non–dieters, but larger frontal–parietal P3 component among successful dieters. Although there were some discrepancies, the visual oddball paradigm was also used in the current study. First, in Balboni et al.'s (2011), participants were asked to respond only to the deviant stimuli, but in the current study, participants were required to react to the standard and deviant stimuli. Second, the standard and deviant stimuli were the same in their study, with the deviant stimuli being the "fattening" standard stimuli. However, in the present study, we used different pictures for the standard and deviant stimuli, and the deviant stimuli included high–energy food pictures, low–energy food pictures, and neutral pictures. Finally, the current study utilized three groups (S–REs, US–REs, UREs), whereas the previous study used successful dieter and non–dieters.

In processing food cues, the P3 component was

associated with response inhibition. Therefore, only US–REs were deficit in the inhibit reaction to the nonfood and food cues (high– and low–energy food) and S–RES were only deficit in the response inhibition to the high–energy food cues. Previous studies demonstrated that S–REs and US–REs used different processing strategies in the food exposure condition . S–REs usually employed the flexible control strategy, depending on the specific food. However, US–REs often employed strict control strategies to treat all types of food. The flexible strategy benefited the successful dieting in the long–term, and the strict control strategy brought short–term effects and finally failed to overeat. Furthermore, REs were over–confident about successfully inhibiting their response when facing palatable food, but often failed in the end .

The current study has several limitations. First, the inhibition deficit was analyzed based on the temporal course of food cues processing. Future studies should thus use a combination of the technologies of ERP and functional magnetic resonance imaging to explore the inhibition ability among REs. Second, participants were grouped based on DEBQ and TFEQ measures. However, the actual eating behavior probably differed from the performances indicated by the items. Further studies should take into consideration the actual eating behaviors when grouping subjects. Finally, different people have different appetite or eating habits. Thus, future studies should use the individualized food pictures to control this effect.

5 Conclusion

This study was among the first to examine the neural correlates of inhibition ability in S–REs, US–REs, and unrestrained eaters. The results showed that S–REs were deficient in terms of inhibiting high–energy food pictures and US–REs were deficient in terms of inhibiting both high– and low–energy food pictures. The results suggest that S–REs and US–REs had different inhibition abilities, which is an important factor in successful dieting.

References

Alexander, J. E., & Polich, J. (1995). P300 differences between sinistrals and dextrals. *Cognitive Brain Res, 2*, 277–282.

Babiloni, C., Del, Percio, C., Triggiani, A. I., Marzano, N., Valenzano, A., Petito, A., et al. (2011). Attention cortical responses to enlarged faces are reduced in underweight subjects: An electroencephalographic study. *Clin Neurophysiol, 7*, 1348–1359. doi: 10.1016/j.clinph.2010.12.001 PMID: 21420903

Baker, R. C., & Kirschenbaum, D. S. (1993). Self–monitoring may be necessary for successful weight control. *Behav Ther, 3*, 377–394.

Beaver, J. D., Lawrence, A. D., van, Ditzhuijzen, J., Davis, M. H., Woods, A., & Calder, A. J. (2006). Individual differences in reward drive predict neural responses to images of food. *J Neurosci, 26*, 5160–5166. PMID:16687507

Blechert, J., Feige, B., Hajcak, G., & Tuschen–Caffier, B. (2010). To eat or not to eat? Availability of food modulates the electrocortical response to food pictures in restrained eaters. *Appetite, 2,* 262–268. PMID:21509240

Choo, V. (2002). WHO reassesses appropriate body–mass index for Asian populations. *Lancet, 9328*, 235.

Danhauer, S. C., Oliveira, B., Myll, J., Berra, K.,& Haskell, W. (2004). Successful dietary changes in a cardiovascular risk reduction intervention are differentially predicted by biopsychosocial characteristics. *Prev Med, 4*, 783–790. PMID: 15351546

Donkers, F. C., & Boxtel, G. J. (2004). The N2 in go/no–go tasks reflects conflict monitoring not response

inhibition. *Brain Cognition, 2*, 165–176. PMID: 15010083

Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1999). ERP components in Go/Nogo tasks and their relation to inhibition. *Acta psychol, 2–3*, 267–291.

Graziano, P. A., Calkins, S. D., & Keane, S. P. (2009). Toddler self-regulation skills predict risk for pediatric obesity. *Int JObe, 4,* 633–641.

Guerrieri, R., Nederkoorn, C., & Jansen, A. (2012). Disinhibition is easier learned than inhibition. The effects of (dis)inhibition training on food intake. *Appetite, 59,* 96–99. doi: 10.1016/j.appet.2012.04.006 PMID:22521403

Hachl, P., Hempel, C., & Pietrowsky, R. (2003). ERPs to stimulus identification in persons with restrained eating behavior. *Int J Psychophysiol, 2*, 111–121. PMID: 12919714

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: Ameta-analysis. *Psychol Bull, 4*, 495–525. doi: 10.1037/a0019486 PMID: 20565167

Herman, C. P., Polivy, J., Lank, C. N., & Heatherton, T. F. (1987). Anxiety, hunger, and eating behavior. *J Abnorm Psychol, 3*, 264–269. PMID: 3680767

Jansen, A., Nederkoorn, C., Baak, L., Keirse, C., Guerrieri, R., & Havermans. R. (2009). High-restrained eaters only overeat when they are also impulsive. *Behav Res Ther, 2,* 105–110. doi: 10.1016/j.brat.2008.10.016 PMID: 19038379

Kong, F. C. (2012). *The neural correlates of food cues processing in restrained eaters*. Southwest University: Doctor Thesis.

Kuijer, R., Ridder, D., Ouwehand, C., Houx, B., & Bos, R. (2008). Dieting as a case of behavioural decision making. Does self-control matter? *Appetite, 3*, 506–511. PMID: 18077056

Leitch, M. A., Morgan, M. J., & Yeomans, M. R. (2013).

Different subtypes of impulsivity differentiate uncontrolled eating and dietary restraint. *Appetite, 69*, 54–63. doi: 10.1016/j.appet.2013.05.007 PMID: 23702263

Lowe, M. R. (1993). The effects of dieting on eating behavior: A three-factor model. *Psychol Bull, 1,* 100–121.

Meule, A., & Kubler, A. (2014). Double trouble. Trait food craving and impulsivity interactively predict food-cue affected behavioral inhibition. *Appetite, 1,* 174–182. doi: 10.1182/asheducation–2014.1.174 PMID:25696852

Meule, A., Lukito, S., Vogele, C., & Kubler, A. (2011). Enhanced behavioral inhibition in restrained eaters. *Eating Behaviors, 2*, 152–155.

Meule, A., Vele, C., & Kübler, A. (2012). Restrained eating is related to accelerated reaction to high caloric food and cardiac autonomic dysregulation. *Appetite, 58,* 638–644. doi: 10.1016/j.appet.2011.11.023 PMID: 22142510

Meule, A., Westenhofer, J., & Kubler, A. (2011). Food cravings mediate the relationship between rigid, but not flexible control of eating behavior and dieting success. *Appetite, 57,* 582–584. doi: 10.1016/j.appet.2011.07.013 PMID: 21824503

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: does self-control resemble a muscle? *Psychol Bull, 2*, 247–259. PMID: 10748642

Musso, F., Brinkmeyer, J., Warbrick, T., & Winterer. G. (2011). Ketamine effects on brain function-simultaneous fMRI/EEG during a visual Oddball task. *Neuroimage, 9,* 269–269.

Nederkoorn, C., Houben, K., Hofmann, W., Roefs, A., & Jansen, A. (2010). Control yourself or just eat what you like? Weight gain over a year is predicted by an interactive effect of response inhibition and implicit

preference for snack foods. *Health Psych, 4*, 389–393.

Nederkoorn, C., Eijs, Y., & Jansen, A. (2004). Restrained eaters act on impulse. *Pers Indiv Differ, 8*, 1651–1658.

Nordgren, L. F., Harreveld, F., Pligt, J. (2009). The restraint bias: How the illusion of self–Restraint promotes impulsive behavior. *Psychol Sci, 20,* 1523–1528. doi: 10.1111/j.1467–9280.2009.02468.x PMID: 19883487

Ouwehand, C., & Papies, E. K. (2010). Eat it or beat it. The differential effects of food temptations on overweight and normal–weight restrained eaters. *Appetite, 1*, 56–60.

Papies, E. K., Stroebe, W., & Aarts, H. (2008). The allure of forbidden food: On the role of attention in self–regulation.*J Exp Soc Psychol, 5*, 1283–1292.

Polich, J., & Herbst, K. L. (2000). P300 as a clinical assay: Rationale, evaluation, and findings. *Int J Psychophysiol, 1*, 3–19.

Stroebe, W., Mensink, W., Aarts, H., Schut, H., & Kruglanski, A. W. (2008). Why dieters fail: Testing the goal conflict model of eating. *J Exp Soc Psychol, 1*, 26–36.

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 6582,* 520–522. PMID: 8632824

Van Koningsbruggen, G. M., Stroebe, W., & Aarts, H. (2013). Successful restrained eating and trait impulsiveness.*Appetite, 60*, 81–84. doi: 10.1016/j.appet.2012.09.016 PMID: 23017465

Van Strien, T., Cleven, A., & Schippers, G. (2000). Restraint, tendency toward overeating and ice cream consumption. *The Int J Eat Disorder, 3*, 333–338. PMID: 10942920

Watson, T. D., & Garvey, K. T. (2013). Neurocognitive correlates of processing food–related stimuli in a Go/No–go paradigm. *Appetite, 71*, 40–47. doi: 10.1016/j.appet.2013.07.007 PMID: 23892319

Yuan, J., He, Y., Qinglin, Z., Chen, A., & Li, H. (2008). Gender differences in behavioral inhibitory control: ERP evidence from a two–choice oddball task. *Psychophysiology, 6,* 986–993. doi: 10.1111/j.1469–8986.2008.00693.x PMID: 18778319

School connectedness and problematic Internet use in adolescents: A moderated mediation model of deviant peer affiliation and self-control

Dongping Li[1] , Xian Li [2] , Yanhui Wang [3] , Liyan Zhao[4], Zhenzhou Bao [5], Fangfang Wen[6]

([1] School of Psychology, Central China Normal University, Wuhan, China)

([2] Department of Educational and Counseling Psychology, University at Albany, State University of New York, Albany, New York, United States)

([3] School of Educational Science, Jiaying University, Meizhou, China)

([4] Arts and Sciences College of Sichuan Normal University, Chengdu, China)

([5] School of Psychology, South China Normal University, Guangzhou, China)

Abstract Although a growing body of research documents the negative association between school connectedness and adolescent problematic Internet use (PIU), little is known about the mediating mechanism (i.e., how school connectedness relates to PIU?) and moderating mechanism (i.e., when the protection is most potent?) underlying this relation. The present study examined whether affiliation with deviant peers mediated the relationship between school connectedness and PIU, and whether this mediating process was moderated by adolescent self−control. A total of 2,758 Chinese adolescents (mean age = 13.53 years) from 10 middle schools completed anonymous questionnaires regarding school connectedness, deviant peer affiliation, self−control, and PIU. After controlling for gender, age, socioeconomic status, and parental attachment, it was found that the negative association between school connectedness and adolescent PIU was partially mediated by deviant peer affiliation. Moreover, this indirect link was stronger for adolescents with poor self−control than for those with good self−control. These findings underscore the importance of integrating the social control theory and organism−environment interaction model to understand how and when school connectedness impacts adolescent PIU.

Keywords school connectedness; problematic internet use; deviantpeeraffiliation; self−control; protective−attenuating hypothesis

1 Introduction

The Internet has become an indispensable tool for information,communication, and entertainment among adolescents(Tsitsika et al. 2009). Despite many advantages, the Internetis not without its problems, especially when its use becomesexcessive or inappropriate. Problematic Internet use (PIU) re−fers to "use of the Internet that creates psychological, social,school and/or work difficulties in a person's life" [1] (Beard andWolf 2001). It consists of particular behaviors such as exces sive gaming, sexual preoccupations, and e−mail/text messagingthat share

the following characteristics: excessive use, withdrawal, tolerance and impaired social and psychological func-tioning (Block 2008; Ko et al. 2005). Ample research evidence has indicate that adolescent PIU is common (2 %~11 %) around the world (Aboujaoude 2010; Young and de Abreu2011), and is associated with a variety of physical, mental,interpersonal, and academic problems (Ko et al. 2012;Shapira et al. 2000; Young and de Abreu 2011). Therefore, itis essential to understand mechanisms that place adolescents atrisk for PIU and develop protective intervention programs.

Among many factors influencing adolescent PIU,the role of school connectedness has received increased attention over the past few years. School connectedness refers to a student's relationship to school and his/ her feelings about school(Libbey 2004). Students who experience positive school connectedness enjoy school life, believe that they are cared for and supported by their teachers and peers, and have better sense of belonging in school. There is considerable evidence suggesting that school connectedness is a powerful predictor for adoles-cents' physical, educational, and socio-emotional well-being(Anderman 2002; Catalano et al. 2004; Resnick et al. 1997). Research has also indicated that adolescents who feel connected to school are less likely to develop PIU (Sun et al. 2005; Wanget al. 2011; Yen et al. 2009). For instance, Jiang and Huang(2008) found that school connectedness was negatively associated with PIU among high school students.Similarly,Wang and colleagues (2011) concluded that school connectedness (positive relationships with teachers and classmates)was aprotective factor for adolescent PIU. These findings highlight the value of increasing school connectedness (an important

aspect of students' socio-emotional needs) in reducing adolescent PIU.

Nonetheless, prior research has focused primarily on the direct link between school connectedness and adolescent PIU. The mediating mechanism (i.e., how school connecedness relates to PIU?) and moderating mechanism (i.e.,when the protection is most potent?) underlying this relationremain largely unknown. Answers to these questions areessential for a better understanding of the etiology of adolescent PIU and the development of targeted intervention programs. In the present study, we aim to investigate two questions: first, whether school connectedness reduces deviant peer affiliation, which in turn reduces the likelihood ofadolescent PIU; second, whether the indirect association between school connectedness and PIU is moderated by important individual characteristics such as self-control.

1.1 Deviant peer affiliation as a mediator

Students spend increasingly more time with their peers inadolescent years, and their susceptibility to peer influence increases and peaks during adolescence (Steinberg and Monahan 2007). According to the social development model(Hawkins and Weis 1985), adolescent school connectednes smay influence the likelihood of deviant peer affiliation, which may in turn influence their deviancy. In other words, deviant peer affiliation may mediate school connectedness effects. Likewise, Oetting and Donnermeyer's (1998) primary social-ization theory postulates that the link between social connectedness (such as school connectedness) and deviant behaviors may not be adirect one;rather,it is deviant peer affiliation that mediates the relationship

---

[1]   Although some researchers used terms such as "Internet addiction","pathological Internet use", or "Internet dependency" to describe thismaladaptive behavior, we prefer the term PIU for two reasons. First,adopting the addiction perspective is inadequate because it remainsdebatable whether excessive Internet use constitutes an addiction (Anget al. 2012; Holden 2010). Second, compared to "pathological" or"dependency", the term "problematic" describes the behavior in a moreinclusive manner covering the entire rang of problematic behaviorfrom mild to severely disturbed behavior and thus can address a widesegment of the population (Ang et al. 2012).

between social connectedness and adolescent deviancy. Consistent with these theoretical frameworks, abundant empirical evidence has demonstrated the mediating effect of deviant peer affiliation in the association of school connectedness with adolescent substance use(Henry 2008; Swaim et al. 1998) and delinquency (Zhang and Messner 1996).For instance, in an American longitudinal study, Henry (2008) found adolescents who were poorly connected to school demonstrated a higher level of drug abuse, and this association was mediated by affiliating with drug-using peers. Although not yet tested, it is reasonable to expect that deviant peer affiliation will also mediate the association between school connectedness and adolescent PIU. In the following section, previous research findings would be reviewed to support our argument.

First,adolescents with strong school connectedness are less likely to affiliate with deviant peers. This is because they are consistently reinforced by maintaining good connection with school, and they are reluctant to compromise the good relationships with teachers and peers. In contrast, adolescents with low school connectedness have limited social reinforcment in school settings for positive conventional goal pursuit,and tend to engage in deviant peer relationships that maximize negative social reinforcement (Hawkins and Weis 1985;Oetting and Donnermeyer 1998). Previous research has also confirmed that school connectedness decreased the likelihood of adolescents' deviant peer affiliation (Hawkins and Weis1985; Murguia et al. 1998; Oetting and Donnermeyer 1998).

Second, when adolescents affiliate with deviant peers (including, but not limited to friends who engage in PIU), they are more likely to develop PIU. Specifically, adolescents are negatively influenced by deviant peers presumably through processes of peer norms, modeling, and pressure (Beard 2011), and are more likely to engage

in various socially disapproved behaviors and less likely to be accepted by the community. Therefore, they might indulge themselves in thevirtual world to escape from the stressful realities. For example, they might play massively multiplayer online games that are potentially addictive to distract themselves from the reall ives. A large body of empirical evidence supports the ideathat deviant peer affiliation plays an important role in shaping adolescent PIU (Ko et al. 2008; Wang et al. 2011). Yang et al. (2008) found that having computer game addicte dfriends was associated with higher tendency on computer game addiction among elementary students. Similarly, Yen and colleagues(2009) reported that high school students whose friends engaged in alcohol use or deviant behaviors (not including PIU) were at greater risk for PIU. These findings suggest that deviant peer affiliation is a significant risk factor for adolescent PIU.

In fact, two studies (Wang et al. 2011; Yen et al. 2009) had specifically examined the influence of school connectedness and deviant peer affiliation on adolescent PIU. However, both studies treated one variable (e.g., deviant peer affiliation) as a nuisance variable by statistically controlling for or removing its influence when they were interested in the impact of the other one (e.g., school connectedness). In other words, they assumed that the two variables operated independently. Unlike the above studies, we take into account the potential association of school connectedness with deviant peer affiliation. Based on the literature reviewed above, we propose the following hypothesis:

Hypothesis 1 deviant peer affiliation will mediate the relationship between school connectedness and adolescent PIU.

1.2 Self-Control as a moderator

Although school connectedness is generally protective against PIU, not all adolescents are equally influenced by school connectedness. Consequently, it is important

to examine variables that may moderate the relationship between school connectedness and PIU. The organism–environment interaction model (Cummings et al. 2002, p. 135) proposes that"individuals with certain intrapersonal attributes respond differently, or more specifically, with greater maladaptation, to similar environmental contexts". However, to our knowledge,no research to date has examined the moderating effect of adolescent intrapersonal characteristics on direct or mediatingpathways from school connectedness to PIU. In this study, we investigated whether the direct and/or indirect pathways would vary as a function of adolescent self–control.

Self–control refers to the ability to monitor, inhibit, persevere and adapt one's behavior, emotions, thoughts and desires in order to achieve a certain goal (Duckworth 2011). Adolescents with good self–control are "more adept than theirimpulsive counterparts at regulating their behavioral, emotional, and attentional impulses to achieve long–term goals" (Duckworth 2011, p. 2639). We have the following reasons for examining self–control as a moderator: (a) a substantive body of research has established self–control as a protective factor against adolescent PIU (Kim et al. 2008; LaRose et al. 2003; Li et al. 2010; Li et al. in press); and (b) it helps to examine how self–control and social control such as school connectedness (Hirschi 1969) interact to impact children's development, an important theoretical issue which attracts heightened interest among developmentalists (Baker 2010;Loukas et al. 2010; Wright et al. 2001).

The organism–environment interaction perspective generally emphasizes that the effect of schoolconnectedness will be moderated by adolescent characteristics such as self–control. Theorists disagree, however, about the nature of the School Connectedness × Self–Control interaction. The protectiveenhancing hypothesis holds that one protective factor enhances the positive effect of another protective factor(Fergus and Zimmerman 2005; Pluess and Belsky in press;Wang et al. 2009). According to this perspective, the beneficial effect of school connectedness would be stronger for adolescents with high rather than low self–control (Fig. 1a). In contrast, the protective–attenuating hypothesis maintains that one protective factor generally confers advantages, and this beneficial effect is particularly strong among individuals at low levels of
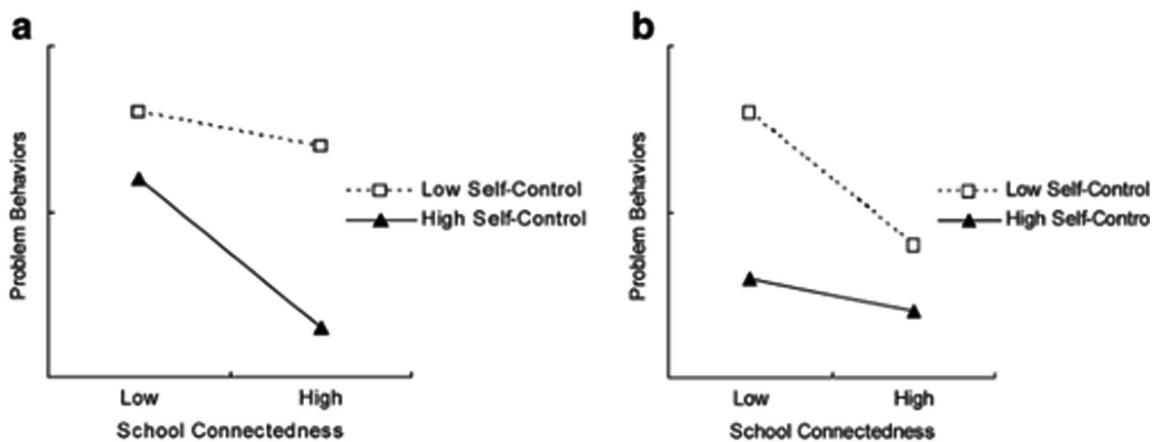
Fig. 1 Hypothetical School Connectednessx Self–Control interaction effects. a Protective–enhancing model, an interaction in which the beneficial effect of school connectedness on problem behaviors is more pronounced for adolescents with higher levels of self–control. b Protective–attenuating model, an interaction in which the beneficial effect of school connectedness on problem behaviors is more pronounced for adolescents with lower levels of self–control. Note that adolescent school connectedness and self–control were arbitrarily labeled as high or low depending on whether their scores on the school connectedness scale and the self–control scale were 1 standard deviation above or below the mean scale score in line with Dearing andHamilton (2006). As such, the labels "high" and "low" are used in relative terms to the overall mean scale score and not in absolute sense *per se*

another protective factor (Wang et al. 2009). From this perspective, the beneficial effect of school connectedness would be stronger for adolescents with low as opposed to high levels of self-control (Fig. 1b). These two hypotheses imply very different practical implications: the former suggests that intervention programs that seek to promote school connectedness will particularly benefit adolescents who have higher levels of self-control;whereas the latter suggests that this kind of intervention will particularly benefit those who score low in self-control.

In the current study, we examine both protective-enhancing and protective-attenuating patterns of the School Connectedness × Self-Control interaction. Empirical findings, although relatively limited and not directly related to PIU, have suggested the likelihood of a protective-attenuating pattern. Specifically, a small but important literature has confirmed that self-control attenuated the associations of school connectedness with alcohol use, antisocial behavior, and adolescent delinquency (Baker 2010; Loukaset al. 2010; Wright et al. 2001). Also, several studies have found that self-control attenuated the association betweend deviant peer affiliation and problematic behaviors (Gardneret al. 2008; Goodnight et al. 2006), although there are some negative results in the literature (Wills et al. 2002;Yarbrough et al. 2012). The Deviant Peer Affiliation × Self-Control interaction might be attributable to, at least in part, the reduced susceptibility of high self-control adolescents to deviant peers' negative reinforcement. Besides, to our knowledge, no research has examined whether the relationship between school connectedness and deviant peer affiliation was moderated by self-control. Based on the above theoretical analyses and empirical evidence, we propose the following hypothesis:

Hypothesis 2 adolescent self-control will moderate the negative indirect association between school connectedness and PIU. This indirect association will be significant among low self-control adolescents but much weaker among high self-control adolescents, which will support our protective-attenuating hypothesis.

1.3 The present study

In summary, the present study brings together two hitherto largely separate lines of research (the social development model and organism-environment interaction model) to account for the mechanisms by which school connectedness is associated with adolescent PIU. The first and second hypothesis, taken together, constitute a "moderated mediation" model whereby the mediating process that is responsible for the relationship between the predictor and the outcome depends on the value of a moderator variable(Edwards and Lambert 2007; Muller et al. 2005). Particularly, our purpose is twofold: (a) to examine whether deviant peer affiliation will mediate the relationship between school connectedness and adolescent PIU, and (b) to test whether this indirect relationship will be moderated by self-control. This integrated model can address questions about both mediation (i.e., how school connectedness relates to less PIU) and moderation (i.e., when the protection is most potent) in one model. Figure 2 illustrates the proposed research model. Given prior research showing that adolescents' gender, age, family socioeconomic status, and parent-adolescent attachment were correlated with PIU (Siomos etal. 2012), we include them as control variables in statistical analyses.

Fig. 2 The proposed moderated mediation model
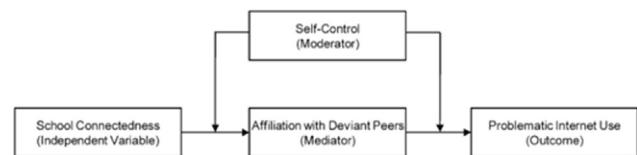
2 Method

2.1 Participants

We recruited participants from 10 middle schools in southern China through stratified and random cluster sampling. The sample was first stratified by region (urban versus rura lareas); then stratified by school type (selective schools versus regular schools, selective schools usually have more quality teachers, enjoy better facilities and resources, thus attract stronger students). Random cluster sampling was used to randomly choose two classes in each grade of each school. A total of 2,758 adolescents participated in this study. Forty−six percent of the participants were males. The mean age of the participants was 13.53 years (SD=1.06, range=10‒19). Research has shown that adolescents at this age have a relatively higher tendency to engage in PIU(Dong and Lin 2011). Reflecting the demographics of this area, 60 % of their fathers and 69 % of their mothers had less than a high school education; 35 % of their fathers and58 % of their mothers had an unskilled or semi−skilled occupation. In addition, we have less than 2 % missing data and the missing data were handled with mean imputation(Little and Rubin 2002).

2.2 Measures

School connectedness. adolescent school connectedness was measured by the School Connectedness Scale developed by Bao et al. (2013). This scale contains six items, which reflect students' relationships with their school (e.g., "I like my school", "I am proud to be part of the school"). Adolescents indicated how often they feel connected to their school on a 4−point scale ranging from 1=never to 4=always. Responses across the six items were averaged, with higher numbers representing greater school connectedness. Evidence for validity includes confirmatory factor analysis (one factor structure) and a nomological network of associations with other constructs (significant and positive correlations with perceptions of teacher support,student support,academic achievement,and a negative correlation with depression). For the current study,the measure demonstrated good reliability (α =0.85).

Deviant peer affiliation. adolescent affiliation with deviant peers was assessed with eight items adapted from prior published questionnaires (Fergusson and Horwood 1999; Kendleret al. 2007). Peers' deviant behaviors included smoking, alcohol use, cheating on school tests, stealing or shoplifting,misbehaving, Internet addiction, skipping or cutting school,and physical and verbal aggression[2] . Previous research on PIU often used only oneitem to measure peer deviancy(Wangetal. 2011; Yang et al. 2008), whereas this study used multiple items. Validity is enhanced because no one singular item adequately captures the full meaning of the construct of interest(Evans et al. in press). Adolescents indicated how many of their friends had shown each of the eight deviant behaviors during the prior year (e.g., "How many of your friends got drunk in the last year?") on a 5−point scale ranging from1=none to 5=almost all. Responses were averaged across the eight items, with higher scores representing greater deviant peer affiliation. For the current study, the measure demonstrat−ed good reliability (α =0.83).

Self−control. adolescent self−control was assessed by the Self−control Scale developed by Bao et al. (2013). Thisscale contained seven items drawn from prior similar instruments (e.g., Duckworth and Seligman 2005). Adolescents indicated how true each item (e.g., "I am veryself−disciplined") was of them on a 4−point scale rangingfrom 1=never to 4=always. Recent research suggests thatadolescent self−report is a relatively valid approach for measuring self−control (Duckworth & Kern 2011). Responses across the seven items were averaged,

with higher numbers representing greater self-control. Evidence for validity includes confirmatory factor analysis (two moderately correlated subscales of activation and inhibitory control) andsignificant associations with theoretically-relevant constructs such as effortful control, academic achievement,substance use, and delinquent behavior. In the present study,the Cronbach's α coefficient of this scale was 0.68 (< 0.70). However, the composite reliability for this measure was 0.71, which indicated that the reliability of this measure was acceptable (Wen and Ye 2011).

PIU. Adolescent PIU was assessed with 10 items from Young's (1996) questionnaire for screening of Internet dependency. This scale has demonstrated good reliability and validity in Chinese samples (Li et al. 2010) and similar items have been used in the National Children's Study of China project (Dong and Lin 2011). A representative item was: "Do you use the Internet as a way of escaping from problems or of relieving an unhappy mood?"Note that Young's PIU index is an index of self-reported addictive feelings/impulses to general online activities-global PIU(Ang et al. 2012; Davis 2001) rather than to specific types/categories of activities (e.g., gaming). Adolescents indicated how true each item was of them on a 6-point scale ranging from 1=*not at all true* to 6=*always true*. Responses across the 10 items were averaged, with higher scores representing greater PIU. For the current study, the measure demonstrated good reliability (α =0.92).

Control variables. This study controlled for adolescents'gender, age, family socioeconomic status, and parent-adolescent attachment. Family socioeconomic status was comprised of a single factor derived from principal component analysis of multiple indicators (parental education,family financial status, and parental occupation status),with higher scores representing higher socioeconomic status. In addition, adolescent attachment to their parents was assessed by the Chinese version of the Inventory of Parent and Peer Attachment-Short Version (Li et al. 2009). This inventory contains 13 items (e.g., "My parents respect my feelings"), which reflect the quality ofthe child-parents relationship. Adolescents indicated howtrue each item was of them on a 5-point scale rangingfrom 1=*never* to 5=*always*. Responses across the 13 items were averaged, with higher numbers representing greater attachment security. For the current study, the measure demonstrated good reliability (α =0.87).

2.3 Procedure

This study was approved by the research ethics committee of our institution. Because the protocol was judged to pose low risk and the data were collected and processed anonymously, oral consent was recommended and obtained fromschool administrators and participants before data collection. After a complete description of the study, participants were told that they could omit any uncomfortable questions and were free to withdraw at any time during data collection. They were also assured that their responses would be kept completely anonymous and confidential. The survey was conducted in classrooms. Trained data collectors administered the questionnaires using scripts and a manual of procedures so as to standardize the data collection process.

---

[2] Some of the friend behaviors (e.g., drinking) seem broadly knowable by adolescents, but cheating in school exams and Internet addiction seems relatively private and difficult for adolescents to report. However, we still included these items because: (a) many cheating behaviors of Chinese adolescents (e.g., copying from other students on a test or exam) require collaboration with others and many adolescents who have cheated in school exams are often criticized/punished "openly" by their teachers, and (b) unlike in the United States, where computers are often accessed from the home, in China adolescents often go together to Internet cafés and discuss what they have done there. These characteristics enable adolescents to provide reliable and valid information about their friends' behaviors.

Table 1. Univariate and bivariate statistics for all study variables

Variables	M	SD	1	2	3	4	5	6	7	8
1. Gender	0.46	0.50	—							
2. Age	13.53	1.06	0.00	—						
3. SES	0.00	1.00	0.05*	− 0.19***	—					
4. Parental attachment	3.45	0.72	− 0.02	− 0.18***	0.19***	—				
5. School connectedness	2.91	0.76	− 0.01	− 0.15***	0.17***	0.36***	—			
6. Deviant peer affiliation	1.55	0.54	0.19***	0.17***	− 0.20***	− 0.25***	− 0.32***	—		
7. Self−control	3.86	0.83	− 0.05**	− 0.10***	0.10***	0.40***	0.30***	− 0.22***	—	
8. PIU	2.47	1.02	0.28***	0.04*	− 0.02	− 0.30***	− 0.22***	0.32***	− 0.40***	—

Gender was dummy coded such that 0 = *male* and 1 = *female*. *SES socioeconomic* status; *PIU* problematic Internet use.

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

3 Results

3.1 Preliminary analyses

According to Young's (1996) diagnosistic criterion for Internet dependency, 6.3 % (*n*=174) of the participants in the current sample displayed severe PIU, which is consistent with the Chinese national data (Dong and Lin 2011) and recent literature (Aboujaoude 2010; Lei 2010; Young and de Abreu 2011). Note that this rate is for descriptive purposes only; all of the subsequent analyses maintained the continuous nature of the PIU variable. Table 1 contains univariate statistics and bivariate correlations for all study variables. Both school connectedness and self−control were negatively associated with adolescent PIU (*r*=−0.22, *p*<0.001;*r*=−0.40, *p*<0.001, respectively), suggesting that school connectedness and self−control are protective factors for PIU. In addition, deviant peer affiliation was positively associated with PIU, *r*=0.32, *p*<0.001, indicating that deviant peer affiliation is a viable risk factor for PIU. Finally,school connectedness was negatively associated with deviant peer affiliation, *r*=− 0.32, *p*<0.001, which is in linewith previous research.

Given that we were analyzing clustered data (i.e.,students are nested within schools), which violated the assumption of ordinary least squares regression that each observation is independent of all other observations in the data set, we chose to use the SAS PROC SURVEYREG to obtain robust standard errors for regression parameter estimates. This method explicitly took into account the clustered nature of the data(SAS Institute Inc 2011).

3.2 Testing for mediation effect

In Hypothesis 1, we anticipated that deviant peer affiliation would mediate the relationship between school connectedness and PIU. To test this hypothesis, we followed MacKinnon's (2008) four−step procedure to establish mediation effect, which requires (a) a significant association between school connectedness and PIU; (b) significant association between school connectedness and deviant peer affiliation; (c) significant association between deviant peer affiliation and PIU while controlling for school connectedness; and (d) significant coefficient for the indirect path between school connectedness and PIU via deviant peer affiliation. A Sobel test determines whether the last condition is satisfied. In all analyses, we included adolescents'gender, age, socioeconomic status, and parental attachment as covariates.

Regression analysis[3] indicated that, in the first step,school connectedness was significantly associated with PIU, *b*=−0.13, *p*<0.001 (see Model 1 of Table 2). In the second step, school connectedness was significantlyassociated with deviant peer affiliation, *b*=− 0.24, *p*<0.001 (see Model 2 of Table 2). In the third step, whenwe controlled for school connectedness, deviant peeraffiliation was significantly associated with PIU,

Table 2. Testing the mediation effects of school connectedness on adolescent *PIU*

Predictors	Model 1 (criterion PIU)		Model 2 (criterion deviant peer affiliation)		Model 3 (criterion PIU)	
	b	*t*	*b*	*t*	*b*	*t*
CO: Gender	0.55	12.31***	0.38	5.09***	0.47	11.24***
CO: Age	− 0.02	− 0.73	0.09	5.23***	− 0.04	− 1.53
CO: SES	0.03	1.00	− 0.13	− 4.43**	0.06	1.64
CO: Parental attachment	− 0.26	− 8.83***	− 0.12	− 5.11***	− 0.24	− 8.64***
X: School connectedness	− 0.13	− 11.58***	− 0.24	− 7.84***	− 0.08	− 5.67***
ME: Deviant peer affiliation					0.20	8.07***
R^2	0.18	0.18	0.22		0.22	
F	94.82***	47.98***	311.08***		311.08***	

Each column is a regression model that predicts the criterion at the top of the column. Gender was dummy coded such that 0=*male* and 1 =*female*. *SES* socioeconomic status; *CO* control variable; *X* independent variable; *ME* mediator; *PIU* problematic Internet use.

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

b=0.20, $p<0.001$. Finally, the Sobel test indicated that the indirect effect of school connectedness on adolescent PIU through deviant peer affiliation was significant, Z=−5.63, $p<0.001$. Overall, the above four criteria for establishing mediation effect were fully satisfied. Therefore, Hypothesis 1 was supported.

3.3 Testing for moderated mediation

As noted, Hypothesis 2 predicted that self−control would moderate the indirect association between school connectedness and adolescent PIU via deviant peer affiliation(Fig. 2). To test this moderated mediation hypothesis, we used the approach suggested by Muller et al. (2005). Specifically, we estimated parameters for three regression models. In Model 1, we estimated the moderation effect of self−control on the relationship between school connectedness and PIU. In Model 2, we estimated the moderation effect of self−control on the relationship between school connectedness and deviant peer affiliation. In Model 3, we allowed both the partial effect of deviant peer affiliation on PIU and the residual effect of school connectedness on PIU to be moderated by self−control. The specification of these models can be seen in Table 3. In each model, we also controlled for

relevant covariates. All the predictors were standardized to minimize multicollinearity (Dearing and Hamilton 2006).

For present purposes, moderated mediation was established if either or both of two patterns existed (Edwardsand Lambert2007; Muller et al. 2005): (a) the path from school connectedness to deviant peer affiliation was moderated by self−control, and/or (b) the path from deviant peer affiliation to adolescent PIU was moderated by self−control.

As Table 3 illustrates, in Model 1, there was an overall effect of school connectedness on adolescent PIU, b=−0.07,$p<0.001$. This effect was not moderated by self−control,b=0.01, $p>0.05$. In Model 2, the mediator, deviant peer affiliation, was the criterion. Here, there was a significan tmain effect of school connectedness, b=−0.21, $p<0.001$,and a significant School Connectedness × Self−Control interaction effect on adolescent PIU, b=0.07, $p<0.05$. For descriptive purposes, we plotted the predicted deviant peer affiliation against school connectedness, separately for lowand high levels of self−control (1 *SD* below the mean and 1*SD* above the mean, respectively) (Fig. 3). Simple slope test(Dearing and Hamilton 2006) indicated that for low self−control adolescents, higher

---

[3] Structural equation modeling (SEM) can also be used to test the theoretical model in the present study. However, (a) SEM often needs complex factor structure, but the measures in this study did not have complex factor structures (Cohen et al. 2003), and (b) SEM often increases standard errors and reduces power (Ledgerwood and Shrout2011). Therefore, we used multiple regression to estimate the model.

Table 3. Testing the Moderated mediation effects of school connectedness on adolescent PIU

Predictors	Model 1 (criterion PIU)		Model 2 (criterion deviant peer affiliation)		Model 3 (criterion PIU)	
	b	t	b	t	b	t
CO: Gender	0.52	10.67***	0.37	5.14***	0.46	9.75***
CO: Age	− 0.02	− 1.02	0.09	4.78***	− 0.04	− 1.81
CO: SES	0.03	1.11	− 0.13	− 4.70***	0.06	1.69
CO: Parental attachment	− 0.16	− 6.16***	− 0.09	− 3.57**	− 0.14	− 5.76***
X: School connectedness	− 0.07	− 5.73***	− 0.21	− 8.81***	− 0.03	− 2.36*
MO: Self−control	− 0.31	− 18.51***	− 0.09	− 3.78**	− 0.29	− 17.15***
XMO: School Connectedness × Self−Control	0.01	0.63	0.07	3.19*	− 0.00	− 0.24
ME: Deviant peer affiliation					0.18	8.15***
MEMO: Deviant peer affiliation × Self−Control					− 0.01	− 0.32
R^2	0.26	0.19	0.28		0.28	
F	429.59***	61.07***	455.31***		455.31**	

Each column is a regression model that predicts the criterion at the top of the column. Gender was dummy coded such that 0=*male* and 1= *female*. *SES*=socioeconomic status; *CO* control variable; *X* independent variable; *MO* moderator; *XMO* interaction between independent variable and moderator; *ME* mediator; *MEMO* interaction between mediator and moderator; *PIU*=problematic Internet use.

*$p < 0.05$. **$p < 0.01$. ***$p < 0.001$.

school connectedness was asso ciated with lower deviant peer affiliation, b_{simple}=−0.29, $p<0.001$. However, for high self−control adolescents, the effect of school connectedness on deviant peer affiliation was weaker, b_{simple}=−0.14, $p<0.001$. Finally, Model 3 showed that the effect of deviant peer affiliation on PIU was significant, b=0.18, $p<0.001$, and this effect was not moderated by self−control (the Deviant Peer Affiliation × Self−Control interaction effect was non−significant).

Overall, the indirect effect of school connectedness
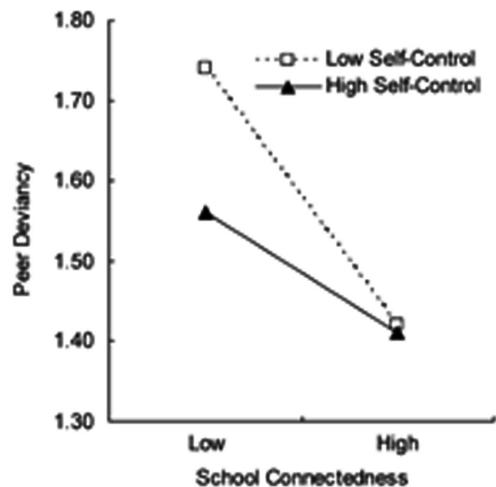
Fig.3 Deviant peer affiliation among adolescents as a function of school connectedness and self−control. Functions are graphed for two levels of self−control: 1 standard deviation above the mean and 1 standard deviation below the mean. Note that the graph is for descriptive purpose only. All inferential analyses maintained the continuous values of school connectedness and self−control

on adolescent PIU via deviant peer affiliation was moderated by self−control. For adolescents low in self−control, school connectedness had protective effect on PIU through reduced deviant peer affiliation, Z=−6.72, $p<0.001$. In contrast, the indirect effect was smaller for adolescents high in self−control, Z=−4.52, $p<0.001$. Given that self−control only affected the first stage of the mediation process, we called this type of model the first stage moderation model, which is one form of moderated mediation model (Edwards and Lambert 2007). Thus, Hypothesis 2 was supported.

4 Discussion

The beneficial effect of school connectedness on adolescent PIU has garnered considerable empirical support (Jiang and Huang 2008; Sun et al. 2005; Wang et al. 2011; Yen et al. 2009). However, the underlying mediating mechanism (i.e.,how school connectedness relates to PIU) and moderating mechanism (i.e., when the protection is most potent) remain as questions of inquiry. This study formulated and tested a moderated mediation model based on an integration of existing theories (i.e., the social development model and organism−environment

interaction model). Findings suggest that the protective effect of school connectedness on adolescent PIU is explained in part by decreased deviant peer affiliation. Moreover, this indirect link is stronger for adolescents with poor self-control than for those with good self-control. We discuss each of our research hypotheses in light of this moderated mediational model of school connectedness and adolescent PIU.

First, consistent with our hypothesis, we found that deviant peer affiliation is an important, underlying psychosocial mechanism that helps explain why strong school connectedness is associated with less PIU. When youth have positive experiences and are closely tied to school, they are less likely to involve with deviant peers, which in turn is associated with less PIU. This finding is congruent with the social development model (Hawkins and Weis 1985) and primary socialization theory (Oetting and Donnermeyer1998), as well as the previous research which showed that the protective effects of school connectedness were mediated by deviant peer affiliation (Henry 2008; Swaim et al. 1998; Zhang and Messner 1996). This finding also affirms an elementary principle in developmental psychology and sociology: social institutions such as school and peer contexts are interdependent, and the functioning of one has important implications for the operation of others (Zhang and Messner 1996). To our knowledge, our study is one of the first that apply the above theories to adolescent PIU research.

As noted, two studies (Wang et al. 2011; Yen et al. 2009) explicitly examined the influences of both school connect edness and deviant peer affiliation on adolescent PIU. However, they assumed that the two variables operated independently. Unlike these studies, the present study examined the impact of school connectedness on deviant peer affiliation, which some argue as an important mechanism that schools may protect students from problematic behav-iors (Hawkins and Weis 1985;

Oetting and Donnermeyer1998). Accordingly, we found that strong school connect edness decreased the likelihood of adolescents' deviant peer affiliation, which in turn decreased the likelihood of PIU. Therefore, the protective effects of school connectedness on adolescent PIU were probably underestimated as a result of controlling for deviant peer affiliation (the mediator in this study). Indeed, Newcombe (2003) pointed out that "some controls control too much" when she commented on such inappropriate use of statistical control.

Second, our results offered support for the moderating role of adolescent self-control on the indirect link between school connectedness and adolescent PIU. In line with the protective-attenuating hypothesis, we found self-control attenuating the relationship between school connectedness,deviant peer affiliation, and PIU through the link betweenschool connectedness and deviant peer affiliation. This protective-attenuating moderation effect has also been reported in previous research. For instance, the negative association between school connectedness and problematic behaviors such as alcohol use, antisocial behavior, and delinquency was substantially weaker among children with higher self-control (Baker 2010; Loukas et al. 2010; Wright et al. 2001). If this pattern of moderation effect could be successfully replicated in future studies, we are able to propose the following argument: for adolescents with poor self-control, social control such as school connectedness will be critical for their development; however, for adolescents with good self-control, the beneficial effect of socia lcontrol may not be very significant. Of course, these predictions are still preliminary and should be tested in future studies. It is worth noting that self-control "attenuates" the relationship between school connectedness and adolescent PIU, this moderation effect still suggests that self-control is a protective rather than a risk factor for adolescent PIU,because high self-control adolescents had

significantly lower levels of deviant peer affiliation than those low in sel−fcontrol (see Table 3 and Fig. 3).

Contrary to our expectations, we did not find evidence that the relationship between deviant peer affiliation and PIU was moderated by self−control. This finding indicates that, to some extent, self−control can protect low school−connected adolescent from affiliating with deviant peers. However, once they affiliate with deviant peers, self−control does not attenuate the relationship between deviant peer affiliation and PIU. Although some previous studies have found self−control attenuating the relationship between deviant peer affiliation and adolescent development (Baker 2010; Loukas et al. 2010; Wright et al. 2001), their outcome variables were externalizing behaviors or antisocial behaviors rather than PIU. Given that this is a negative result with other possible explanations, we should not draw strong conclusions at this stage.

Nonetheless, our findings suggest that the impact of school connectedness may not be constant across all levels of self−control. When we are interested in the effect of school connectedness on PIU, we should not ignore adolescent self−control and vice versa. The two control processes perhaps should not be viewed as separate but rather interdependent with each other. Therefore, this study makes important contributions to the social development model. By incorporating self−control into the meditation process of PIU, we discovered effects that would have been neglected without the moderation effect analysis. The moderated mediation model in this study can offer greater predictive power and conceptual fruitfulness than the original social development model alone. In fact, this theoretical integration has been highly valued by researchers and policy makers during the past decade for its power in offering in sights that cannot be obtained by separating these theories.

Limitation and future drection. Several limitations must also be considered when interpreting the results of the present study. First, we are not able to make any causal inferences from this cross−sectional data. Future studies should use longitudinal de−signs to seek evidence for the causal assumptions that are made in this study. Second, the data were based on adolescent self−reported measures. Although previous research has shown that adolescent self−report of risk behavior does not contain strong bias under conditions of anonymity andconfidentiality (Chan 2009; Tourangeau and Yan 2007), as was the case in our study, future studies should simulta−neously employ multiple informants and multiple methods to collect data. Third, given that tests of moderating effects in observational studies typically require large sample size to provide enough statistical power, we did not stratify the analyses by gender or other background variables. Such research is important for understanding whether these results can be applied to different subgroups. Finally, the model was tested based on a population sample of Chinese adolescent. Therefore, we should not generalize the current conclusion to other cultural and/or geographical settings.

Despite these limitations, this research has important practical implications. First, our results suggest that promoting school connectedness could be helpful in reducing the risk of PIU. Given the fact that satisfying students'socioemotional needs such as school connectedness is still an underemphasized aspect in school reform (Shochet et al. 2006), this finding is critical. Second, our findings can help practitioners understand pathways by which school connectedness is associated with adolescent PIU, suggesting a possible venue for targeted interventions. For example, reducing deviant peer affiliation may ameliorate some of thedetrimental effects of poor school connectedness on adolescent PIU. This is important because many of the current PIU interventions were delivered in peer group settings (i.e.,adolescents with PIU were aggregated into groups such as training camps), which under some circumstances may inadvertently

increase adolescent interaction with deviant peers (Dishion and Tipsord 2011). Third, given that the protective effect of school connectedness on PIU through deviant peer affiliation is stronger for adolescents with poor self-control than for those with good self-control, we should develop and conduct targeted interventions, such as increasing students' connectedness to school, especially for those with low selfcontrol.

In summary, although further replication and extension are needed, this study is an important step in investigating how school connectedness relates to adolescent PIU. It shows that deviant peer affiliation can serve as one potential mechanism by which school connectedness is associated with less PIU. Moreover, our finding suggests that this beneficial effect of school connectedness appears to be stronger for adolescents with poor self-control than for those with good self-control. These findings demonstrate the importance of moderated mediation models in understanding the mechanisms by which school connectedness is associated with adolescent PIU.

References

Aboujaoude, E. (2010). Problematic internet use: an overview. *WorldPsychiatry, 9*, 85 – 90.

Anderman, E. M. (2002). School effects on psychological outcomesduring adolescence. *Journal of Educational Psychology, 94*, 795 – 809.

Ang, R. P., Chong, W. H., Chye, S., & Huan, V. S. (2012). Lonelinessand generalized problematic internet use: parents' perceivedknowledge of adolescents' online activities as a moderator.*Computers in Human Behavior, 28*, 1342 – 1347.

Baker, J. O. (2010). The expression of low self-control as problematicdrinking in adolescents: an integrated control perspective. *Journalof Criminal Justice, 38*, 237 – 244.

Bao, Z., Zhang, W., Li, D., Li, D., & Wang, Y. (2013). School climate and academic achievement among adolescents: a moderated mediation model. *Psychological Development and Education, 29*, 61 – 70.

Beard, K. W. (2011). Working with adolescents addicted to the Internet. In K. S. Young & C. N. de Abreu (Eds.), Internet addiction: *A handbook and guide to evaluation and treatment* (pp. 173 – 189). Hoboken,: John Wiley & Sons Inc.

Beard, K. W., & Wolf, E. M. (2001). Modification in the proposed diagnostic criteria for internet addiction. *CyberPsychology & Behavior, 4*, 377 – 383.

Block, J. J. (2008). Issues for DSM-V: internet addiction. *American Journal of Psychiatry, 165*, 306 – 307.

Catalano, R. F., Haggerty, K. P., Oesterle, S., Fleming, C. B., & Hawkins, J. D. (2004). The importance of bonding to school for healthy development: findings from the social development research group. *Journal of School Health, 74*, 252 – 261.

Chan, D. (2009). So why ask me? Are self-report data really that bad? In C. E. Lance & R. J. Vandenberg (Eds.), *Statistical and methodological myths and urban legends* (pp. 309 – 336). New York:Taylor & Francis Group.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Appliedmultiple regression/correlation analysis for the behavioral sci ences* (3rd ed.). Mahwah: Lawrence Erlbaum Associates, Inc.

Cummings, E. M., Davies, P. T., & Campbell, S. B. (2002). *Developmental psychopathology and family process: Theory, research, and clinical implications*. New York: The Guilford Press.

Davis, R. A. (2001). A cognitive-behavioral model of pathologicalInternet use. *Computers in Human Behavior, 17*, 187 – 195.

Dearing, E., & Hamilton, L. C. (2006). Contemporary

advances and classic advice for analyzing mediating and moderating variables.*Monographs of the Society for Research in Child Development,71*, 88–104.

Dishion, T. J., & Tipsord, J. M. (2011). Peer contagion in child and adolescent social and emotional development. *Annual Review of Psychology, 62*, 189–214.

Dong, Q., & Lin, C. D. (2011). *Contemporary Chinese children and adolescents' psychological development: Report of the National children's study of China*. Beijing: Science Press.

Duckworth, A. L. (2011). The significance of self-control. *Proceedings of the National Academy of Sciences, 108*, 2639–2640.

Duckworth, A. L., & Kern, M. L. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality, 45,* 259–268.

Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline out does IQ in predicting academic performance of adolescents.*Psychological Science, 16*, 939–944.

Edwards, J. R., & Lambert, L. S. (2007). Methods for integrating moderation and mediation: a general analytical framework using moderated path analysis. *Psychological Methods*, *12*,1–22.

Evans, G. W., Li, D., & Sepanski Whipple, S. (in press). Cumulative risk and child development. *Psychological Bulletin.*

Fergus, S., & Zimmerman, M. A. (2005). Adolescent resilience: a framework for understanding healthy development in the face of risk. *Annual Review of Public Health, 26*, 399–419.

Fergusson, D. M., & Horwood, L. J. (1999). Prospective childhood predictors of deviant peer affiliations in adolescence. *Journal of Child Psychology and Psychiatry, 40*, 581–592.

Gardner, T., Dishion, T., & Connell, A. (2008). Adolescent self regulation as resilience: resistance to antisocial behavior withinthe deviant peer context. *Journal of Abnormal Child Psychology, 36*, 273–284.

Goodnight, J. A., Bates, J. E., Newman, J. P., Dodge, K. A., & Pettit,G. S. (2006). The interactive influences of friend deviance and reward dominance on the development of externalizing behavior during middle adolescence. *Journal of Abnormal Child Psychology, 34,* 573–583.

Hawkins, J. D., & Weis, J. G. (1985). The social development model: an integrated approach to delinquency prevention. *Journal of Primary Prevention, 6,* 73–97.

Henry, K. L. (2008). Low prosocial attachment, involvement with drug-using peers, and adolescent drug use: a longitudinal exam ination of mediational mechanisms. *Psychology of Addictive Behaviors, 22*, 302–308.

Hirschi, T. (1969). *Causes of delinquency. Berkeley: University* of California Press.

Holden, C. (2010). Behavioral addictions debut in proposed DSM-V.*Science, 327*, 935.

Institute Inc, S. A. S. (2011). *SAS/STAT® 9.3 user's* guide. Cary: SAS Institute Inc.

Jiang, Y., & Huang, G. (2008). The influence of family and school environment on the youth's use of the Internet. *Journal of Hunan Mass Media Vocational Technical College, 8*, 36–42.

Kendler, K. S., Jacobson, K. C., Gardner, C. O., Gillespie, N., Aggen, S. A., & Prescott, C. A. (2007). Creating a social world: a developmental twin study of peer-group deviance. *Archives of General Psychiatry, 64,* 958–965.

Kim, E. J., Namkoong, K., Ku, T., & Kim, S. J. (2008). The relationship between online game addiction and aggression, self-control and narcissistic personality traits. *European Psychiatry, 23*, 212–218.

Ko, C. H., Yen, J. Y., Chen, C. C., Chen, S. H., & Yen,

C. F. (2005). Proposed diagnostic criteria of internet addiction for adolescents.T*he Journal of Nervous and Mental Disease, 193*, 728–733.

Ko, C. H., Yen, J. Y., Yen, C. F., Chen, C. S., Weng, C. C., & Chen, C. C. (2008). The association between Internet addiction and problematic alcohol use in adolescents: the problem behavior model.*CyberPsychology & Behavior, 11*, 571–576.

Ko, C. H., Yen, J. Y., Yen, C. F., Chen, C. S., & Chen, C. C. (2012).The association between internet addiction and psychiatric disorder: a review of the literature. *European Psychiatry, 27*, 1–8.

LaRose, R., Lin, C. A., & Eastin, M. S. (2003). Unregulated Internet usage: addiction, habit, or deficient self–regulation? *Media Psychology, 5*, 225–253.

Ledgerwood, A., & Shrout, P. E. (2011). The trade–off between accuracy and precision in latent variable models of mediation processes. *Journal of Personality and Social Psychology*, *101*, 1174–1188.

Lei, L. (2010). Exploration of adolescents' internet addiction. *Psychological Development and Education, 26*, 554–560.

Li, X., Li, D., & Newman, J. (in press). Parental behavioral and psychological control and problematic Internet use among Chinese adolescents: The mediating role of self–control.*Cyberpsychology, Behavior, and Social Networking*.

Li, N., Zhang, W., Li, D., Mai, Y., Wang, J., & Xing, W. (2009). Parent adolescent attachment, emotional intelligence, and aggression in adolescence. *Psychological Development and Education, 25*, 91–96.

Li, D., Zhang, W., Li, X., Zhen, S., & Wang, Y. (2010). Stressful life events and problematic Internet use by adolescent females and males: a mediated moderation model. *Computers in Human Behavior, 26*, 1199–1207.

Libbey, H. P. (2004). Measuring student relationships to school: attachment, bonding, connectedness, and engagement. *Journal of School Health, 74,* 274–283.

Little, R. J., & Rubin, D. B. (2002). *Statistical analysis with missing data* (2nd ed.). New Jersey: Wiley.

Loukas, A., Roalson, L. A., & Herrera, D. E. (2010). School connectedness buffers the effects of negative family relations and poor effortful control on early adolescent conduct problems. *Journal of Research on Adolescence, 20*, 13–22.

MacKinnon, D. P. (2008). *Introduction to statistical mediation analysis*. New York: Taylor & Francis Group.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of Personality and Social Psychology*, *89*, 852–863.

Murguia, E., Chen, Z., & Kaplan, H. B. (1998). A comparison of causal factors in drug use among Mexican Americans and non–Hispanic Whites. *Social Science Quarterly, 79*, 341–360.

Newcombe, N. S. (2003). Some controls control too much. *Child Development, 74*, 1050–1052.

Oetting, E. R., & Donnermeyer, J. F. (1998). Primary socialization theory: the etiology of drug use and deviance. *I. Substance Use & Misuse, 33*, 995–1026.

Pluess, M., & Belsky, J. (in press). Vantage sensitivity: Individual differences in response to positive experience. *Psychological Bulletin*.

Resnick, M. D., Bearman, P. S., Blum, R. W., Bauman, K. E., Harris, K. M., Jones, J., et al. (1997). Protecting adolescents from harm: findings from the National longitudinal study on adolescent health. *Journal of the American Medical Association, 278*, 823–832.

Shapira, N. A., Goldsmith, T. D., Keck, P. E., Jr., Khosla, U. M., & McElroy, S. L. (2000). Psychiatric features of individuals with problematic Internet use. *Journal of Affective Disorders, 57*, 267–272.

Shochet, I. M., Dadds, M. R., Ham, D., & Montague, R.

(2006).School connectedness is an underemphasized parameter in adolescent mental health: results of a community prediction study.*Journal of Clinical Child and Adolescent Psychology, 35*, 170－179.

Siomos, K., Floros, G., Fisoun, V., Evaggelia, D., Farkonas, N.,Sergentani, E., et al. (2012). Evolution of Internet addiction inGreek adolescent students over a two-year period: the impact of parental bonding. *European Child and Adolescent Psychiatry, 21*,211－219.

Steinberg, L., & Monahan, K. C. (2007). Age differences in resistance to peer influence. *Developmental Psychology, 43*, 1531－1543.

Sun, P., Unger, J. B., Palmer, P., Gallaher, P., Chou, C., Baezconde-Garanati, L., et al. (2005). Internet accessibility and usage among urban adolescents in southern California: implications for web-based health research. *CyberPsychology & Behavior, 8*, 441－453.

Swaim, R. C., Bates, S. C., & Chavez, E. L. (1998). Structural equation socialization model of substance use among Mexican-American and White non-Hispanic school dropouts. *Journal of Adolescent Health, 23*, 128－138.

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys.*Psychological Bulletin, 133*, 859－883.

Tsitsika, A., Critselis, E., Kormas, G., Filippopoulou, A., Tounissidou, D., Freskou, A., et al. (2009). Internet use and misuse: a multi variate regression analysis of the predictive factors of internet use among Greek adolescents. *European Journal of Pediatrics, 168*,655－665.

Wang, Y., Zhang, W., Peng, J., Mo, B., & Xiong, S. (2009). The relations of attachment, self-concept, and deliberate self-harm in college students. *Psychological Exploration, 5*, 56－61.

Wang, H., Zhou, X., Lu, C., Wu, J., Deng, X., & Hong, L. (2011). Problematic internet use in high school students in Guangdong province, China. *PLoS ONE, 6,* e19660.

Wen, Z., & Ye, B. (2011). Evaluating test reliability: from coefficient alpha to internal consistency reliability. *Acta Psychologica Sinica, 43,* 821－829.

Wills, T. A., Sandy, J. M., & Yaeger, A. M. (2002). Moderators of the relation between substance use level and problems: test of a self regulation model in middle adolescence. *Journal of Abnormal Psychology, 111*, 3－21.

Wright, B. R. E., Caspi, A., Moffitt, T. E., & Silva, P. A. (2001). The effects of social ties on crime vary by criminal propen sity: a life-course model of interdependence. *Criminology, 39*, 321－352.

Yang, L., Cheng, X. Y., Wang, Y., & Wang, H. (2008). Analysis of the correlation factor of fascination tendency on computer game of urban primary pupils in 4 6 grades. *Modern Preventive Medicine, 35*, 4435－4438.

Yarbrough, A., Jones, S., Sullivan, C., Sellers, C., & Cochran, J. (2012). Social learning and self-control: assessing the mod erating potential of criminal propensity. *International Journal of Offender Therapy and Comparative Criminology, 56*, 191－202.

Yen, C. F., Ko, C. H., Yen, J. Y., Chang, Y., & Cheng, C. P. (2009).Multi-dimensional discriminative factors for Internet addictionamong adolescents regarding gender and age. *Psychiatry andClinical Neurosciences, 63*, 357－364.

Young,K.S.(1996,August).I*nternetaddiction:Theemerge nceofa newclinical disorder.* Paper presented at the 104th Annual Meeting of the AmericanPsychologicalAs sociation,Toronto, Ontario, Canada.

Young, K. S., & de Abreu, C. N. (2011). Internet addiction: *A hand book and guide to evaluation and treatment.* Hoboken: John Wiley & Sons Inc.

Zhang, L., & Messner, S. F. (1996). School attachment and official delinquency status in the People's Republic of China. *Sociological Forum, 11,* 285－303.

Multi-family group therapy for adolescent Internet addiction: Exploring the underlying mechanisms

Qinxue Liu[1,2], Xiaoyi Fang[3,4], * Ni Yan[5], Zongkui Zhou[1,2] , Xiaojiao Yuan[6], Jing Lan[3], Chaoying Liu[3]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan 430079, China)

([2] School of Psychology, Central China Normal University, Wuhan 430079, China)

([3] Institute of Developmental Psychology, Beijing Normal University, Beijing 100875, China)

([4] Academy of Psychology and Behavior, Tianjin Normal University, Tianjin 300387, China)

([5] Faculty of Psychology, Southwest University, Beibei 400700, China)

([6] School of Sociology and Psychology, Southwest University for Nationalities, Chengdu 610041, China)

Abstract   Objective: Internet addiction is one of the most common problems among adolescents and effective treatment is needed. This research aims to test the effectiveness and underlying mechanism of multi−family group therapy (MFGT) to reduce Internet addiction among adolescents. Method: A total of 92 participants consisting of 46 adolescents with Internet addiction, aged 12−18 years, and 46 their parents, aged 35−46 years, were assigned to the experimental group (six−session MFGT intervention) or a waiting−list control. Structured questionnaires were administered at pre−intervention (T1), post−intervention (T2) and a three−month follow−up (T3). Results: There was a significant difference in the decline both in the average score and proportion of adolescents with Internet addiction in MFGT group at post−intervention(M_{T1}=3.40, M_{T2}=2.42, p<0.001; 100 versus 4.8%, p<0.001) maintained for three months(M_{T3}=2.06, p<0.001; 100 versus 11.2%, p<0.001). Reports from both adolescents and parents were significantly better than those in the control group. Further explorations of the underlying mechanisms of effectiveness based on the changed values of measured variables showed that the improvement in adolescent Internet use was partially explained by the satisfaction of their psychological needs and improved parent−adolescent communication and closeness. Conclusions: The six−session multi−family group therapy was effective in reducing Internet addiction behaviors among adolescents and could be implemented as part of routine primary care clinic services in similar populations. As family support system is critical in maintaining the intervention effect, fostering positive parent−adolescent interaction and addressing adolescents' psychological needs should be included in preventive programs for Internet addiction in the future.

Keywords multi−family group therapy; Internet addiction; family relationships; need satisfaction; effectiveness mechanism

1 Introduction

With the rapid development of the Internet, Internet addiction has become a widespread and problematic phenomenon. Internet addiction, also known as Pathological Internet Use, Problematic Internet Use and Compulsive Internet Use, is characterized by excessive and compulsive Internet use and a preoccupation with and loss of control over this use that interferes with individuals' daily functioning (Caplan, 2002; Davis, 2001; Van den Eijinden, Meerkerk, Vermulst, Spijkerman, & Engles, 2008; Young & Abreu, 2011). Currently, it is one of the most common behavioral problems for adolescents, who are more exposed to Internet use and consequently more vulnerable than adults (Lortie & Guitton, 2013), with a prevalence rate higher than 8% in some countries (Cho, Kim, Kim, Lee, & Kim, 2008; Kuss, Grifths, & Binder, 2013; Van den Eijnden, Spijkerman, Vermulst, van Rooij, & Engels, 2010). In China, approximately 10% of adolesecents (approximately 20 million teenagers) reported a tendency towards or current diagnosis of Internet addiction (China Internet Network Information Block, 2008; Center, 2013). Internet addiction may cause psychological distress, personality development problems, social problems and poor school performance (Brezing, Derevensky, & Potenza, 2010; Young, Pistner, O'Mara, & Buchanan, 2000). In addiction, high comorbidity with effective disorders, impulse control disorders and substance abuse disorders have been reported (Petersen, Weymann, Schelb, Thiel, & Thomasius, 2009; Weinstein & Lejoyeux, 2010). There is significant research around the diagnosis, epidemiology, predicting factors and negative outcomes of Internet addiction, but little is known about treating it, which is an imperative for adolescents, families, schools and society, especially in China (King, Delfabbro, Griffiths, & Gradisar, 2011; Winkler, Drsin,

Rief, Shen, & Glombiewski, 2013).

Petersen et al. (2009) conducted a survey at the request of the German health department and argued that clinical recommendations are not possible due to the lack of studies and that further research is urgently needed. In a systematic review of Internet addiction treatment, only eight studies were included. Half of them were psychological approaches, and two utilized cognitive-behavioral therapy (King et al., 2011). Peukert, Sieslack, Barth, & Batra (2010) also indicate cognitive-behavioral and pharmacological approaches as potentially effective treatments in their review. They suggest that interventions with family members could be useful. Winkler et al. (2013) further examine the efficacy of different treatments for Internet addiction (13 studies included) in their meta-analysis, and their results show that CBT did not perform significantly better than other psychological treatments, even though it appears to be the predominant approach for treating Internet addiction. They also suggest that both individual counseling and group therapy have their shortcomings and that further research around different approaches and modalities is needed. However, there is no study that examines which factors contribute to the efficacy of treatment or what predictors cause the behavior change to happen, which is very important to evaluate and improve interventions (Liu, Fang, & Zhou, 2011).

Family plays a central role in the socializing process for adolescents, and parents provide emotional connection, behavioral constraints and modeling (Gray & Steinberg, 1999; Lau, Quadrel, & Hartman, 1990). Family-based intervention is the most thoroughly studied treatment modality for adolescent substance dependence and addiction, and there is a large body of research to support its efficacy (for a review, see Liddle, 2004). Previous research also proved that a good relationship and communication with parents are protective factors for adolescents from Internet addiction (Kim, Jeong,

& Zhong,2010; Van den Eijinden et al., 2010). Family members involved in interventions facilitate the process of recovery and help the addict maintain a lasting effect of intervention after sessions (Liddle, 2004; Zhong et al., 2011).Grounded in family system theory and integrated in family and group therapy, multi−family group therapy (MFGT) was proposed as a promising new approach to treat Internet addiction behaviors, but no empirical study was conducted (Liu et al., 2011). The effectiveness of MFGT has been empirically demonstrated among adolescents with psychological disorders (Chien & Chan, 2013; McDonell & Dyck, 2004), children at risk for special educational services (Kratochwill, McDonald, Levin, Scalia, & Coover, 2009) and in addiction related areas (Conner et al., 1998; Zubrick et al., 2005). In this field, Zhong et al. (2011) found that family−based intervention is more effective in reducing Internet addiction than group therapy that involved only the adolescents. The multi−family group offers both adults and adolescents the advantages of a peer group, which helps them to get support and learn from peer confrontation. Transferential reactions occur not only within one family but also across family lines, facilitating the group to serve as both as an arena for cross transferences−based on each person's introject and as a reality tester (Leichter, & Schulman, 1974). Connection within family members is also helpful for high treatment attendance (Nieter, Thornberry Jr., & Brestan−Knight, 2013). Moreover, family−oriented intervention might be particularly effective in Chinese culture, where the cohesion between family members is highly emphasized. Therefore, the present study aims to explore both the effectiveness of MFGT on Internet addiction and the underlying mechanisms of the effectiveness.

One mechanism through which MFGT may effectively reduce Internet addiction is improving parent−adolescent communication and closeness. Compared with non− addicts, adolescents with Internet addiction have poorer communication with their parents (Park, Kim, & Cho, 2008) and are more likely to receive rejection and negative feedback from their parents (Van den Eijnden et al., 2010). Poor parent−adolescent communication and low perceived parent−adolescent closeness, in turn, predicted adolescents' Internet addiction (Liu, Fang, Deng, & Zhang, 2012; Liu, Fang, Zhou, Zhang, & Deng, 2013). According to the Circumplex Model of Marital and Family Systems proposed by Olson (Olson, 2000; Olson, Sprenkle, & Russell, 1974;), family communication is critical in facilitating intimacy among family members and strengthening the family's adaptability to change. MFGT emphasizes improving family cohesion and motivation to change within the family; it not only focuses on the parent−adolescent interaction but also values the style and strength of attachment between family members (Dickerson & Crase, 2005). Therefore, it could be a well−suited approach to treat Internet addiction among adolescents.

The second mechanism through which MFGT may take effect in treating adolescents' Internet addiction is by fulfilling their psychological needs through strengthening their communication and relationship with their parents. Psychological need is considered one of the most important driving forces that promotes behavioral change. Fulfillment of psychological needs through Internet use has been proposed as an internal motive in adolescents' Internet addiction (Morris & Ogan, 1996; Suler, 1999). Adolescents' unfulfilled needs for competence and relatedness in life and perceived need satisfaction online are the major precursors of their excessive Internet use (Cai, Cui, & Li, 2007; Shen,Liu, & Wang, 2013; Wan, Zhang, Liu, Deng, & Fang, 2010). Compared with non−addicts, Internet addicts perceived higher need satisfaction online and lower need satisfaction in real life (Deng, Fang, Wan, Zhang, & Xia, 2012). Therefore, if parent−adolescent communication practices and relationships are improved, adolescents' psychological needs for relatedness or competence

might be more easily fulfilled through their daily life interactions with their parents, which, in turn, could be helpful to reduce their reliance on the Internet for fulfilling their needs. As fundamental as these two underlying mechanisms appear to be in affecting adolescents' Internet addiction, they have nevertheless been barely examined explicitly in prior Internet addiction intervention studies. In this study, we include these two underlying mechanisms as major intervening variables to examine whether the effectiveness of family group intervention for adolescent Internet addiction depends on them.

Based on a quasi-experimental design, the present study examines the effectiveness of the MFGT for adolescent Internet addiction among 46 pairs of adolescents and their parents. The study aims to examine three hypotheses: First, the intervention group shows a reduction in Internet addiction both at the end of the intervention and at three-month follow-up compared with the control group. Second, adolescents in the intervention group show improved communication and relationship with their parents and psychological needs satisfaction in real life. Third, the effectiveness of the intervention is partially explained by the improved parent-adolescent relationship and communication, and adolescents' psychological need satisfaction in real-life.

2 Method

2.1 Participants

Participants were recruited through advertisements on school websites in Baotou City of Inner Mongolia in China. Related information about the research and a simplified scale of Internet addiction, which is used for clinical diagnosis, were included in the advertisement. Families who were interested and matched the diagnosis were welcome to sign up and have a face-to-face interview one by one. Among the 55 families who signed up for the intervention study, 46 families were selected based on the Adolescent Pathological Internet Use Scale (APIUS; detailed information about APIUS is provided in the measurement section) and inclusion criteria. The body screening scale, SCL-90 and simplified addiction screening scale were used to exclude participants who possessed physical disabilities, mental disorders or other addictive behaviors. Only one boy was excluded for depression.Twenty-onefamilies were assigned to the intervention group because their schedules matched with the intervention arrangement and the other 25 families were included in the control group because they could not set up a continued intervention schedule. Families in the control group were added to the waiting-list for the intervention study after the informed consent from the parents and the adolescents.

The intervention group had a dominant proportion of male (n=17) over female (n=4) with an average age of 15 years old (SD=1.73). The female-to-male ratio of parents in the intervention group was 16: 5 with an average age of 40.9 years old (SD=2.85). Nine of them held a degree of college or above (42.9%); six of them obtained a high school degree (28.6%) and six of them did not obtain a high school degree (28.6%). The average monthly income ranges from 2,000 to 10,000 Yuan with an average of 4,685, which indicated middle-income families in the city. All parents were first-time married.

The demographic compositions in the control group resembled those in the intervention group. Adolescents in the control group had a male-to-female ratio of 21:4 and the average age was 15.7 years old (SD=1.2). Families in the control group did not did not show significantly differences from those in the intervention group.

2.2 Procedures

First, the Manual of Adolescents Internet Addiction Family Group Therapy was developed with precision based on the theoretic framework of family group therapy,

previous intervention practices, and empirical studies. Before the intervention was launched, a pilot study was implemented among six adolescents and their families to assess the operability of the intervention design, potential problems in administering the intervention, and smooth transitions between activity themes in each intervention session. With preliminary results, interview feedback from the pilot studyand consultation with experts on the intervention team, the *Manual of Adolescents Internet Addiction Family Group Therapy* was modified and finalized. Then, recruited parents provided informed consent for their adolescent children's and their own participation.

All participants were asked to complete assessments both before (T1) and after the intervention (T2), and at a three-month follow-up (T3) as well. Participants were assured of the confidentiality of their responses. Procedures were approved by the Institutional Review Board of the Institute of developmental Psychology, Beijing Normal University. The details of the procedures are presented in Figure 1.

2.3 Intervention

After the intervention began, the 21 families in the intervention condition were randomly divided into three intervention groups with seven families in each group. Two therapists were assigned to each group randomly and all therapists had the same clinical background under family and group therapy training. The intervention was given every three days, with each session lasting 2 h. Six sessions were administered for each grouped families.

The intervention was tailored to strengthen parent-adolescent communication and relationships and shift adolescents' fulfillment of psychological needs from the Internet to interactions and building relationships with family members. Specific topics and activities were designed for each intervention session and connected with each other across six sessions, each of which included five parts in 2h : a warm-up exercise, feedback on homework from the last session (except the first session), a main structured activity, a brief summary and the family
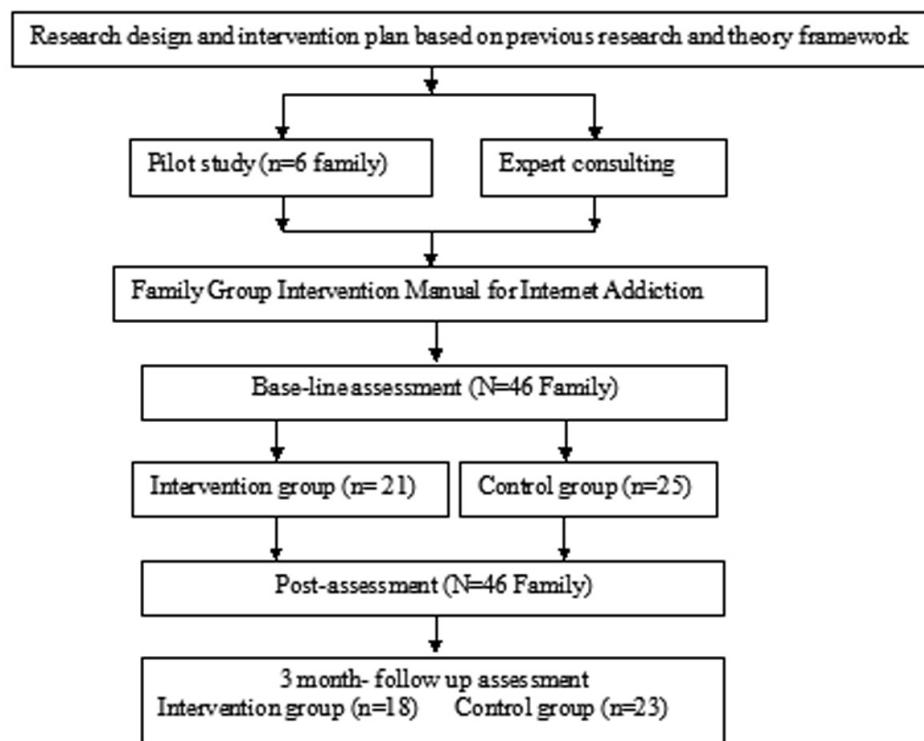
Fig. 1 The intervention process and participants flow diagram.

Table 1. Comparisons of measured variables between the Intervention group and control group at T1, T2 andT3

| | | T1 | T2 | T3 | | |
		M(SD)	M(SD)	M(SD)	F^1	F^2
Average APUIS	Intervention	3.40(.27)	2.46(.61)	2.06(.73)	38.31***	65.98***
	Control	3.38 (.20)	3.59(.31)	3.27(.26)	7.62**	2.50
	t	−0.54	−7.79***	−6.72***		
Internet use time	Intervention	26.38 (9.6)	11.43(5.75)	7.08(3.98)	40.16***	56.65***
	Control	27.08(11.1)	27.52(11.40)	22.29(6.0)	2.21	3.03
	t	−0.21	−4.73***	−9.39***		
Parental reports	Intervention	3.37(.48)	3.13(.66)	2.70(.44)	6.05**	21.10***
	control	3.36(.52)	3.45(.72)	3.20(.57)	1.43	1.97
	t	0.01	−4.12***	−5.26***		
Addiction rate[a]	Intervention	100%	4.8%	11.1%		
	Control	100%	96%	87 %		

Note. F^1 indicates F statistics from comparison among three assessments; F^2 indicates the F statistics from the linear test; [a]Five Parcitipants were missing at T3.
**p<0.01
***p<0.01.

assignment. During the sessions, the following topics were focused on: understanding a family with Internet addict (session 1), parent−adolescent communication skills training (session 2), parent−adolescent communication practices on Internet addiction (sessions 2 and 3), parent−adolescent relationship building skills training (session 4), associations between psychological needs and Internet use and how to satisfy the unfulfilled need in the family relationships (session 5) and setting up appropriate and healthy expectations for the family system (session 6). One additional session at the three−month follow up was designed to target potential relapse, discuss new issues and generate solutions to maintain the effectiveness of the intervention.

2.4 Measurement

2.4.1 Adolescent internet addiction.

Two indicators of Internet addiction were reported by adolescents themselves. First, they reported their average number of hours spent on the Internet per week over the past month. Second, they reported on their Internet addiction behaviors with the Adolescent Pathological Internet Use Scale (Lei & Yang, 2007). This scale contains 38 items with each item being rated from 1 to 5 (1=not true at all; 5=true all of the time). Six subscales were included: salience, social comfort, mood alteration, tolerance,

compulsive Internet use and negative outcomes. The average scores across the 38 items were used as indicators of Internet addiction with higher scores indicating more serious Internet addiction. This scale has high internal consistency (α =0.95 for the whole scale; α between .81 and .91 for all subscales) and high test−retest reliability (r=0.86). Based on the APIUS average scores, adolescents with scores below 3 were considered to be normal Internet users, adolescents with scores between 3 and 3.15 were considered to have a tendency towards Internet addiction, and adolescents with scores higher than 3.15 were defined as having Internet addiction (Lei & Yang, 2007). The Internet addiction rate was calculated based on the number of adolescents defined as having a tendency towards or having Internet addiction (APIUS > 3) and is considered as one of the intervention effectiveness indicators in the study.

Second, parents' reports of children's Internet use behaviors in the last month were used as a supplementary measure of adolescents' Internet use behaviors. Three items were reported: Internet use frequency (1="very rare"to 5 ="very frequently"), parents' observation about adolescents' Internet use appropriateness (1="very appropriate"to 5 ="not appropriate at all"), and parents'satisfaction toward children's Internet use

behaviors (1="very satisfied" to 5 ="not satisfied at all"). The average score across the three items was used to present the parents' evaluation of children's Internet use behaviors. A higher score indicates more Internet use and lower behavior control.

2.4.2 Parent–adolescent relationship.

Adolescents reported their relationships with the parent who participated in the intervention on nine items from the Closeness to Parents on a scale (Buchnan, Maccoby, & Dornbush, 1991) from 1 (not at all true) to 5 (very much true). Sample questions included "How openly do you talk with your mother (father)?" or"How close do you feel to your mother (father)?" A Chinese adaption of the scale has been used in previous study (Liu et al., 2013) and it has a high internal consistency in the study ($\alpha = .91$). Average scores across the nine items were used to represent the adolescent's relationships with the mother or the father during the intervention.

2.4.3 Parent–adolescent communication.

The Parent–Child Communication Scale (Barnes & Olson, 1985) was used to assess the adolescents' perception of their communication with the parent in the intervention. This scale contains 20 items on a scale from 1 (never) to 5 (always). It is composed of two dimensions that measure the degree of openness and the extent of problems in family communications. The responses were identified separately for fathers and mothers. The average score across both dimensions was used to represent the average level of parent–child communication in this study. A Chinese adaption of the scale has been used in previous study (Liu et al., 2012) and the α for the scale is 0.82 in the study.

2.4.4 Adolescent Psychological Needs.

Adolescents rated their psychological needs using a scale modified from the College Students' Psychological Needs and Fulfillment *Scale* (Wan et at., 2010). The scale is composed of three subscales: the degree of psychological needs; needs satisfaction from real life, and needs

satisfaction from the Internet. Each subscale includes 35 items that tap into eight dimensions of the targeted subscale:need for autonomy, need for entertainment, need for interaction, need for achievement, need for impact, need for acknowledgement, need for expression and need for information. The rating for each item in the degree of psychological needs subscale ranges from 1 (not strong at all) to 5 (extremely strong). Ratings in the other two need satisfaction subscales range from 1 (very low) to 5 (very high). Unsatisfied psychological needs was calculated by subtracting the subscale scores on need satisfaction in real life from the subscale scores on the degree of psychological needs. The advantage of the Internet in satisfying needs was calculated by subtracting the subscale scores on need satisfaction in real life from the subscale scores on need satisfaction from the Internet. These two scores were two major process variables of interest in the study. The α for the scale is 0.97 in the study.

2.4.5 Analysis plan

The data analysis proceeded in three steps. First, T tests and Repeated–Measures ANOVA analyses were conducted to test the effectiveness of the intervention based on the comparison among adolescents' Internet addiction measures in the intervention and control group at T1, T2 and T3. Second, comparisons on all intervening variables between the two groups at T1, T2 and T3 were conducted, and changed values of the Intervening variables from T1 to T2 and T3 were created. Third, hierarchical multiple regression analyses were conducted to examine whether the change in adolescents' Internet addition behaviors is explained by the change in the measured intervening variables.

3 Results

3.1 Effectiveness of the Intervention

Before examining the effectiveness of the intervention,

Table 2. Comparisons on intervening process variables

| | | T1 | T2 | T3 | | |
		M(SD)	M(SD)	M(SD)	F[1]	F[2]
P–A relationship	Intervention	2.99(0.45)	3.72(0.81)	3.79(0.64)	14.86***	29.94***
	Control	2.96(0.44)	2.99(0.49)	3.05(0.48)	1.45	2.84
	t	0.21	2.89**	2.96**		
P–A communication	Intervention	2.96(0.71)	3.71(0.49)	3.83(0.62)	13.78***	30.57***
	Control	2.94(0.44)	2.98(0.47)	3.03(0.42)	0.08	
	t	0.19	4.08***	3.97***		
Unsatisfied needs	Intervention	0.67(0.58)	0.19(0.56)	−0.01 (0.45)	11.56***	19.85***
	Control	0.65(0.37)	0.47(0.25)	0.48(0.29)	3.48	4.65
	t	0.18	−2.57*	−3.87**		
Advantage of Internet in satisfying need	Intervention	0.59(0.57)	−0.24(0.59)	−0.81(0.58)	23.68***	39.15***
	Control	0.58(0.44)	0.46(0.25)	0.47(0.29)	3.41	2.61
	t	0.09	−4.83***	−7.78***		

Note. P–A stands for parent–adolescent; F[1] indicates F statistics from comparison among three assessments ; F[2] indicates the F statistics from the linear test.
*$p<0.05$,**$p<0.01$,***$p<0.001$;

the Internet addiction measures at T1 were compared between the intervention and control groups and no significant difference was detected, indicating that the two groups were at the same or similar level of Internet addiction at the baseline of the study (see F values at T1 in Table 1). The effectiveness of the intervention was manifested in three aspects: First, adolescents in the intervention group significantly reduced the time they spent on the Internet by the end of the intervention, spending about half of the time as adolescents in the control group did (see F values at T2 in Table 1). Second, comparisons of APIUS scores demonstrated that the intervention group experienced a decrease in their average APIUS scores from T1 to T2. Third, parents in the intervention group reported more satisfaction with adolescents' Internet use behaviors at the end of the intervention compared with both their satisfaction at the baseline and parents' satisfaction in the control group. In sum, based on reports from both adolescents and their parents, the intervention was effective in terms of reducing adolescents' Internet addiction behaviors by the end of the intervention.

Results from the repeated measures ANOVA (see Table 1) showed that the differences in APIUS scores across the three measurements were significant(F (2, 78) =

38.31, $p < .001$) and also displayed a linear decrease over time from Time 1 to Time 3 ($F(1,39)= 65.98$, $p < .001$), which indicates that the intervention effects remained at T3. Time spent on the Internet also displayed a significant decrease from T1 to T3 (F (2, 78) = 40.16, $p <.001$) with a linear decrease from T1 to T3 as well (F (1, 39) =56.65, $p < .001$). In the control group, significant differences were found among the three APIUS assessments and a further paired sample t test showed that APIUS scores at T2 were significant higher than at T1($t=-4.15$, $p<0.001$), but there was no linear tendency of APIUS scores from T1 to T3 (F (1, 39)=2.50, $p=0.13$). Parents' reports on adolescent Internet addiction at the three-month follow-up also substantiated the effect of the intervention. At 3 months after the intervention, parents still reported a significant reduction in their perception of their children's Internet addiction, revealing a significant linear decrease from T1 to T3.

As an important indicator of intervention effectiveness in the whole intervention group, the Internet addiction rate yielded from the APIUS average scores was also compared across the intervention and control groups over time. As displayed in Table 1, all adolescents in both groups were either addicted to the Internet or had the tendency towards Internet addiction before the intervention. By

the end of the intervention, only 1 out of 21 adolescents (4.8%) in the intervention group was still addicted to the Internet, compared with as many as 24 out of 25 (96%) in the control group. At the three−month follow−up, two adolescents in the intervention group showed a relapse. However, the intervention effects remained as only 11.1% of adolescents in the intervention group remained addicted after the intervention ended, compared with 87% of their counterparts in the control group.

3.2 Examining mediating effects of intervening variables

Because we measured the perception of parent−adolescent interaction and psychological need that is deeply inside of adolescents, we used the reports from the adolescents themselves to examine the change process of adolescent Internet behaviors. To examine the second hypothesis (that the intervention effects are mediated through major process variables), we first examined the change in parent−adolescent relationships, parent−adolescent communication and adolescents' psychological needs. We measured these attributes at their baseline, at the end of the intervention, and at a three−month follow−up for both the intervention and control groups.

As displayed in Table 2, the measured variables did not differ significantly between the intervention and control groups at T1, but did differ significantly at T2 and T3. Adolescents in the intervention group demonstrated improvement in their relationship and communication with their parents, an increase in their fulfillment of needs, a decrease in Internet use' advantage in fulfilling their needs in a linear function.

With the demonstration of improvement on all process variables in the intervention group, we continued to examine whether the change in those process variables accounted for change in adolescents' Internet addiction behaviors. Before examining those mediating effects, change values (ΔX) were created for adolescent reported Internet addiction behaviors, parent−adolescent communication and relationship, and psychological needs by subtracting the post−intervention assessment values from the baseline values ($\Delta X = X_{T2} - X_{T1}$) .

To examine the contribution of those process variables in the explanation of intervention effectiveness on adolescents' Internet addiction, a hierarchical multiple regression analysis was performed. Process variables were entered in three steps: In step 1, $\Delta APIUS$ was the dependent variable and Δ *parent−adolescent relationship* was the independent variable. In step 2, Δ *parent−adolescent communication* was entered into the step 1 equation. In step 3, two indicators of $\Delta Adolescents'$ *psychological* needs were added in the equation in step 2. The results of step 1 (Table 3) indicated that 31% of the variance in the change in adolescents' Internet addiction ($R^2 = 0.31$) was accounted for by improvements in the parent−adolescent relationship. In step 2, after entering changes in mother−adolescent communication, an additional 34% of the variance in the change in adolescents' Internet addiction was explained($\Delta R^2 = 0.34$) and this change in R^2 was significant. In step 3, the addition of psychological needs did not explain additional variance in the dependent variable.

To examine the mediating effects of process variables from T1 to T3, change values (ΔX) were created by following the same procedure: subtracting the three−month follow−up assessment values from the baseline values ($\Delta X = X_{T3} - X_{T1}$). A hierarchical multiple regression analysis was again performed (see the right half of Table 3). First, the results of the hierarchical analysis revealed that changes in parent−adolescent communication and relationships explaining significant variance in adolescents' Internet addiction from T1 to T3. Moreover, when predicting longer term effects in reducing Internet addiction, changes in adolescents' psychological needs also added significant prediction of variance in the Internet addiction change

Table 3. Regressions for changed values of measured variables at T2 and T3

	$\triangle X = X_{T2} - X_{T1}$						$\triangle X = X_{T3} - X_{T1}$					
	β	t	R^2	F	$\triangle R^2$	$\triangle F$	β	t	R^2	F	$\triangle R^2$	$\triangle F$
Step 1												
ΔP–A relationship	−.56	−2.93**	.31	8.55**	.31	8.55**	−.61	−2.62*	.36	6.76*	.36	6.76*
Step 2												
Δ P–A relationship	−.03	−.13	.65	16.99***	.34	17.87**	−.03	−.11	.55	6.79*	.19	4.74*
Δ P–A communication	−.78	−4.22**					−.73	−2.19*				
Step 3												
Δ P–A relationship	.03	.17	.70	7.07**	.05	.82	.00	.01	.83	10.01**	.27	6.49*
Δ P–A communication	−.51	−1.98*					−.53	−1.98*				
Δ Unsatisfied Needs	.02	.13					.34	1.00				
ΔAdvantage of Internet in satisfying needs	.39	1.33					.88	3.44**				

Note. P–A stands for parent–adolescent;

*p<0.05, **p<0.01, ***p<0.001.

(ΔR^2 =0.27). This was in large part due to changes in adolescents' satisfaction of their psychological needs through the Internet (β =0.88, p< .01).

4 Discussion

The current study represents a practical clinical trial of treating adolescents' Internet addiction using the multi–family group approach.Based on prior family group intervention practices in treating adolescent psychopathology (Chien & Chan, 2013; McDonell & Dyck, 2004; Zhong et al., 2011), the current study is the first to our knowledge to apply the approach of MFGT in treating adolescents' Internet addiction. The adolescents' Internet addiction rate dropped from 100% at the baseline assessment to 4.8% at the end of the intervention and remained at 11.1% at the three–month follow–up assessment. Time spent on Internet in the intervention group also significantly declined throughout the intervention until the three–month follow–up. Analyses of the value changes in measured variables indicated which factors were associated with the decrease of adolescent Internet addiction. Improved parent–adolescent communication and need satisfaction in real life were associated with decrease in Internet addiction. If adolescents in the intervention group perceived an improvement in their communication and relationships with parents, learned alternative ways to fulfill their needs and felt less reliance on the Internet,this might promote their motivation to sustainably change their behavior. The results are consistent with prior evidence suggesting the role of feeling supported and trusted in improving the effectiveness of family group intervention (Dickerson & Crase, 2005).

These findings corroborate the idea that intervention programs for adolescents need to get parents actively involved and include them as part of the solution. The family system approach shifts the emphasis on individual family members to the entire family as a unit and the dynamic interactions between family members (Dickerson & Crase, 2005). Further, multi–family group is very helpful in the lasting the effective of intervention. In the multi–family group, each family represents a subsystem with a shared history and current life situation which makes for an enriching and complex process. Then, the multi–group serves both the family system and individual as an arena for cross transferences and as a reality tester (Leichter & Schulman, 1974). Moreover, the participation of other family members in the intervention can create a more supportive environment in which the participants' behavioral changes are valued, encouraged,and maintained even after the intervention

ends (McDonell & Dyck, 2004).

The present study takes a further step in unraveling the underlying mechanism through which MFGT took effects. Parent–adolescent interactions and relationship quality and adolescents' increasing satisfaction of psychological needs in real life partially accounted for the effectiveness of the intervention in reducing adolescents' Internet addiction, further supporting prior findings (Liu, et al, 2012, 2013; Olson et al., 1974). Consistent with previous studies, perceived positive interaction with parents protected adolescents from Internet addiction (Liu et al, 2013; Van den Eijnden et al., 2010). Positive communication facilitated emotional connections between family members and helped them to understand and clarify their needs (White, 2000). The alternative ways and skills of need satisfaction that adolescents and parents learned from the sessions would also expand the degree of communication and interaction within the family. All of these changes could be helpful to keeping children away from Internet addiction. However, the decrease in advantage of Internet in satisfying need did not predict change in Internet use behaviors until the three–month follow–up assessment. This delayed effect has been reported before in interventions targeted to treating adolescents' affective disorders (Goldberg–Arnold, Fristad, & Gavazzi, 1999). It suggests that improvement in adolescents' satisfaction of psychological needs through real life interactions may take time to occur. The positive and effective interaction skills and patters between adolescents and parents need to be practiced before adolescents gradually feel comfortable and natural enough to get used to them. It also may take time for adolescents to change and adapt their need satisfaction habits, which may be a potential advantage of MFGT as it may have lasting intervention effects. Once a benign interaction and relationship pattern is established, the family obtains a built–in force to sustain

the intervention effects (McDonell & Dyck, 2004).

Some factors limit the conclusions that can be drawn from these data. First, the study did not take other competing intervention models into consideration. It has been recommended in intervention studies to compare across multiple intervention paradigms to better evaluate the effectiveness of certain intervention approach. For example, a family therapy approach or group intervention approach could be incorporated in a comparison group to further evaluate the effectiveness of the MFGT paradigm. Second, the gender ratio of parents was not balanced, as most of the participating parents were mothers. This might be an obstacle in clarifying the unique role of the mother–adolescent relationship or father–adolescent relationship in preventing Internet addiction among adolescents, as the impact of their relationship with their father and mother differed with adolescent genders on Internet addiction (Liu et al., 2013). It would be clearer if further clinical studies endeavored to recruit parents of both genders. Third, based on a quasi–experimental design, participants were not randomly assigned to an intervention or control group. However, the baseline comparison between the groups did not show a significant difference and indicate that the results of the study are reliable. Fourth, the data were mainly from questionnaires, and social desirability might influence adolescents' reports of their Internet use behavior. Future studies should assess Internet addiction through psychological interviews. Additionally, the present study only measured general Internet use and addiction, limiting the generalization of results to specific Internet addiction. It would be helpful to improving intervention if further studies paid attention to the treatment of subtypes of Internet addiction.

Given the different characteristics of adolescents and their families, multiple dimensions of multi–family group therapy might be more complex and require exploration in future studies.The existence of equifinality

in adolescents' Internet addiction behaviors also indicates that multiple factors, including personal, interpersonal and environmental factors, could lead to the emergence of Internet addiction among adolescents. For example, individual differences among children in basic need satisfaction in daily real life were associated with the way in which they engage with the Internet (Shen et al., 2013). Therefore, multi-family group interventions that tailor to different etiologies of Internet addiction might improve its effectiveness.

In the addiction field, prevention is more important than intervention. Future studies might explore the possibility and feasibility of an integrated preventive intervention framework including school, community and family, based on evidence from other related adolescent health-related domain (Dodge & Godwin, 2013; Lovato et al., 2013). In consideration of the particularity of Internet addiction behaviors, namely that Internet serves as a necessary part of life and Internet use cannot be completely cut off but only be guided to a rational and controllable level, it is very important to discern how to guide youth use of the Internet appropriately and optimize the benefits of the Internet for adolescents and children.

References

Barnes, H. L., & Olson, D. H. (1985). Parent - adolescent communication and circumplexmodel. *Child Development, 56*, 438 - 447.

Block, J. J. (2008). Issues for DSM-V: Internet addiction. *The American Journal of Psychiatry,165*, 306 - 307.

Brezing, C., Derevensky, J. L., & Potenza, M. N. (2010). Non-substance-addictive behaviorsin youth: Pathological gambling and problematic internet use. Child and *AdolescentPsychiatric Clinics of North America, 19*(3), 625 - 637.

Buchnan, C. M., Maccoby, E. E., & Dornbush, S. M. (1991). Caught between parents:Adolescents' experience in divorced homes. *Child Development, 62*, 1008 - 1029.

Cai, Y. Y., Cui, L. J., & Li, X. (2007). A research on the psychological needs of teenagers'online game behaviors. *Psychological Science, 30*(1), 169 - 172 (in Chinese).

Caplan,S. E. (2002). Problematic internet use and psychosocial wellbeing: Development ofa theory-based cognitive-behavioral measurement instrument. *Computers in HumanBehavior, 18*, 553 - 575.

Chien, W. T., & Chan, S. W. C. (2013). The effectiveness of mutual support group interven-tion for Chinese families of people with schizophrenia: A randomized controlledtrial with 24-month follow-up. *International Journal of Nursing Studies, 50*,1326 - 1340.

China Internet Network Information Center (2013r). Statistical report on the develop-ment of internet in China. (2013, Jan).

Cho, S. C., Kim, J. W., Kim, B. N., Lee, J. H., & Kim, E. H. (2008). Biogenetic temperamentand character profiles and attention deficit hyperactivity disorder symptoms inKorean adolescents with problematic internet use. *Cyberpsychology & Behavior,11*(6), 735 - 737.

Conner, K. R., Shea, R. R., McDermott, M. P., Grolling, R., Tocco, R. V., & Baciewicz, G. (1998).The role of multifamily therapy in promoting retention in treatment of alcohol andcocaine dependence. *The American Journal on Addictions, 7*(1), 61 - 73.

Davis, R. A. (2001). A cognitive-behavioral model of pathological internet use (PIU).*Computers in Human Behavior, 17*(2), 187 - 195.

Deng, L. Y., Fang, X. Y., Wan, J. J., Zhang, J. T., & Xia, C. C. (2012). The relationship ofpsychological needs and need gratification with internet addiction among collegestudents. *Journal of Psychological Science, 35*(1), 123 - 128 (in Chinese).

Dickerson, A. D., & Crase, S. J. (2005). Parent - adolescent

relationships: The influence ofmulti-family therapy group on communication and closeness. *The American Journalof Family Therapy, 33*(1), 45–59.

Dodge, K. A., & Godwin, J. (2013). The conduct problems prevention research group.Social-information-processing patterns mediate the impact of preventive interven-tion on adolescent antisocial behavior. *Psychological Science, 24*(4), 456–465.

Gray, M. R., & Steinberg, L. (1999). Unpacking authoritative parenting: Reassessing a mul-tidimensional construct. *Journal of Marriage and Family, 61*, 574–587.

Goldberg-Arnold, J. S., Fristad, M. A., & Gavazzi, S. M. (1999). Family psychoeducation: Giv-ing caregivers what they want and need. *Family Relations, 48*, 411–417.

Kim, D. H., Jeong, E. J., & Zhong, H. (2010). Preventive role of parents in adolescent prob-lematic Internet game use in Korea. *Korean Journal of Sociology, 44*(6), 111–133.

King, D. L., Delfabbro, P. H., Griffiths, M. D., & Gradisar, M. (2011). Assessing clinical trials ofinternet addiction treatment: A systematic review and CONSORT evaluation. *ClinicalPsychology Review, 31*, 1110–1116.

Kratochwill, T. R., McDonald, L., Levin, J. R., Scalia, P. A., & Coover, G. (2009). Families andschools together: An experimental study of multi-family support groups for childrenat risk. *Journal of School Psychology, 47*(4), 245–265.

Kuss, D. J., Griffiths, M. D., & Binder, J. F. (2013). Internet addiction in students: Prevalenceand risk factors. *Computers in Human Behavior, 29*, 959–966.

Lau, R. R., Quadrel, M. J., & Hartman, K. A. (1990). Development and change of young adults'preventive health beliefs and behavior: Influence from parents and peers. *Journal ofHealth and Social Behavior, 31*, 240–259.

Lei, L., & Yang, Y. (2007). The development and

validation of adolescent pathologicalinternet use scale. *Acta Psychologica Sinica, 39*(4), 688–696 (in Chinese).

Leichter, E., & Schulman, G. (1974). Multi-family group therapy: A multidimensionalapproach. *Family Process, 13*, 95–110.

Liddle, H. A. (2004). Family-based therapies for adolescent alcohol and drug use: Researchcontributions and future research needs. *Addiction, 99*(Suppl. 2), 76–92.

Liu, Q. X., Fang, X. Y., Zhou, Z. K., Zhang, J. T., & Deng, L. Y. (2013). Perceived parent–ad-olescent relationship, perceived parental online behaviors and pathological internetuse among adolescents: Gender-specific differences. *PloS ONE*, 8(9), e75642.

Liu, Q. X., Fang, X. Y., Deng, L. Y., & Zhang, J. T. (2012). Parent–adolescent communication,parental internet use and internet-specific norms and pathological internet useamong Chinese adolescents. *Computers in Human Behavior, 28*, 1269–1275.

Liu, Q. X., Fang, X. Y., & Zhou, N. (2011). A review of the research on internet addictionamong Chinese adolescents. *Journal of South China Normal University (Social Scienceedition), 3*, 65–70.

Lortie, C. L., & Guitton, M. J. (2013). Internet addition assessment tools: Dimensional struc-ture and methodological status. *Addiction, 108*, 1207–1216.

Lovato, C., Watts, A., Brown, K. S., Lee, D., Sabiston, C., Nykiforuk, C., et al. (2013). Schooland community predictors of smoking: A longitudinal study of Canadian high school.*American Journal of Public Health, 103*(2), 362–368.

McDonell, M. G., & Dyck, D. G. (2004). Multiple-family group treatment as an effective in-tervention for children with psychological disorders. *Clinical Psychology Review, 24*,685–706.

Morris, M., & Ogan, C. (1996). The internet as mass medium. *Journal of Communication,46*(1), 29–50.

Nieter, L., Thornberry, T., Jr., & Brestan-Knight, E. (2013). The effectiveness of group par-ent - child interaction therapy with community families. *Journal of Child and FamilyStudies, 22*, 490 - 501.

Olson, D. H. (2000). Circumplex model of marital and family systems. *Journal of FamilyTherapy, 22*(2), 144 - 167.

Olson, D. H., Sprenkle, D. H., & Russell, C. S. (1974). Circumplex model of marital and familysystems: I. Cohesion and adaptability dimensions, family types, and clinicalapplications. *Family Process, 18*, 3 - 28.

Park, S. K., Kim, J. Y., & Cho, C. B. (2008). Prevalence of internet addiction and correlationswith family factors among South Korean adolescents. *Adolescence, 172*(43), 895 - 907.

Petersen, K. U., Weymann, N., Schelb, Y., Thiel, R., & Thomasius, R. (2009). Pathologicalinternet use — Epidemiology, diagnostics, co-occurring disorders and treatment.Fortschritte Der Neurologie *Psychiatrie, 77*, 263 - 271.

Peukert, P., Sieslack, S., Barth, G., & Batra, A. (2010). Internet- and computer game addic-tion: Phenomenology, comorbidity, etiology, diagnostics and therapeutic implica-tions for the addictives and their relatives. *Psychiatrische Praxis, 37*, 219 - 224.

Shen, C. X., Liu, R. D., & Wang, D. (2013). Why are children attracted to the internet? Therole of need satisfaction perceived online and perceived in daily real life. *Computersin Human Behavior, 29*, 185 - 192.

Suler, R. (1999). To get what you need: Healthy and pathological internet use.*Cyberpsychology & Behavior, 5*(2), 385 - 393.

Van den Eijinden, R., Meerkerk, G. J., Vermulst, A. A., Spijkerman, R., & Engles, R. (2008).Online communication, compulsive internet use, and psychosocial well-beingamong adolescents: A longitudinal study. *Developmental Psychology, 44*(3), 655 - 665.

Van den Eijnden, R. J. J. M., Spijkerman, R., Vermulst, A. A., van Rooij, T. J., & Engels, R. C. M. E.(2010). Compulsive internet use among adolescents: Bidirectional parent - childrelationships. *Journal of Abnormal Child Psychology, 38*(1), 77 - 89.

Wan, J. J., Zhang, J. T., Liu, Q. X., Deng, L. Y., & Fang, X. Y. (2010). Development of college stu-dents' psychological need internet gratification questionnaire. *Studies of Psychologyand Behavior, 8*(2), 118 - 125 (in Chinese).

Weinstein, A., & Lejoyeux, M. (2010). Internet addiction or excessive internet use.*The American Journal of Drug and Alcohol Abuse, 36*, 277 - 283.

White, F. A. (2000). Relationship of family socialization processes to adolescent moralthought. *The Journal of Social Psychology, 140*(1), 75 - 91.

Winkler, A., D?rsin, B., Rief,W., Shen, Y., & Glombiewski,J.A.(2013). Treatment of internetaddiction: *A meta-analysis. Clinicial Psychological Review, 33*, 317 - 329.

Young, K. S., Pistner, M., O'Mara, J., & Buchanan, J. (2000). Cyber-disorders: The mentalhealth concern for the new millennium. *Cyberpsychology & Behavior, 3*(5), 475 - 479.

Young, K. S., & Abreu, C. (2011). Internet addiction. A handbook and guide to evaluation andtreatment. Hoboken, NJ: John Wiley & Sons.

Zhong, X., Zu, S., Sha, S., Tao, R., Zhao, C., & Yang, F. (2011). The effect of a family-basedintervention model on internet-addicted Chinese adolescents. *Social Behavior andPersonality, 39*(8), 1021 - 1034.

Zubrick, S. R., Ward, K. A., Silburn, S. R., Lawrence, D., Williams, A. A., Blair, E., et al. (2005). Prevention of child behavior problems through universal implementation of agroup behavioral family intervention. *Prevention Science, 6*(4), 287 - 303.

Sentential complements and false belief understanding in Chinese mandarin-speaking preschoolers: A training study

Shuliang Mo[1,2,3,*] , Yanjie Su[3] , Mark A. Sabbagh[4], Jiaming Xiu[3]

([1] School of Psychology, Central China Normal University, Wuhan, 430079, China)

([2] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU) of Ministry of Education, Wuhan, China)

([3] Department of Psychology, Peking University, Beijing 100871, China)

([4] Department of Psychology, Queen's University, Kingston, Ontario, Canada)

Abstract We conducted a training study to better understand how Chinese Mandarin−speaking preschoolers' facility with sentential complement grammatical constructions affects performance on false belief tasks. Eighty−four Mandarin−speaking Chinese 34 year−olds who were initially unsuccessful on false belief tasks were randomlyassigned to four training conditions. Two involved training on sentential complement structures, one involved training on understanding of false representations, and one was a control conditionthat involved no specific training. Participants who received trainingon sentential complements with communication verbs performed significantly better on false belief posttests than thos ein the control group. Children in the false representation training group did not show improvement in the sentential complement tests. The findings suggest facility with sentential complement grammatical structures can promote false belief reasoning. How−ever,explicit false belief understanding can emerge even when children have little competence with sentential complement constructions.

Keywords False belief understanding; Sentential complement; False representation; Training, language

1 Introduction

False belief understanding is a milestone in representational theory of mind development(Wellman, Cross, & Watson, 2001). Children's performance on false− belief tasks is associated with language ability (Astington & Jenkins, 1999; Hughes & Dunn, 1998). A recentmeta−analysis (Milligan,Astington,& Dack, 2007) concluded that children's early−developing language skills predict subsequent development of theory of mind during the

preschool years. The goal of this study is to examine Chinese Mandarin−speaking preschoolers to investigate the role of one aspect of language development−facility with the sentential complement grammatical construction− in the development of false belief understanding.

The hypothesis that facility with sentential complement grammatical constructions is critical for false belief understanding was first advanced in a series of influential papers by de Villiers and colleagues (de Villiers & Pyers, 2002; de Villiers & de Villiers, 2000). Noting that mental state verbs (e.g.,think) take sentential complements meaning that the clause that complements the mental verb is itself a complete sentence (e.g., Mary thinks that the chocolate is in the cupboard). De Villiers and colleagues argued that the syntax provides a formal mechanism for contrasting the truth value of two clauses within a single sentence. With reference to the above example, Mary may truly think that the chocolate is in the cupboard, but the content of Mary's belief may or may not be true depending on the preceding events. de Villiers and de Villiers (2000) suggested that this class of syntactic con struction provides a format for thinking about epistemic mental states, and that until facility with this construction is acquired, an understanding of belief may be seriously limited.

There is some evidence that facility with sentential complement constructions is important for the development of preschoolers' false belief reasoning. In one study, de Villiers and Pyers (2002) aske dchildren to report the propositional content of a story character's mistake, lie, or false belief (e.g., He thought he found his ring, but it was really a bottlecap. What did he think?). Ability to report the propositional contents and produce sentential complements themselves showed rapid gains between the ages of 3 and 4, about the time false belief understanding emerges. Indeed, false belief understanding could be uniquely predicted by the earlier onset of sentential complements (de Villiers & Pyers,2002).

A second source of evidence in support of the association between sentential complement facility and theory of mind development comes from studies of deaf children born into hearing families. Schick, de Villiers, de Villiers, and Hoffmeister (2007) reported that deaf children from hearing families showed a significant delay on theory of mind tasks. Importantly, vocabulary and sentential complement comprehension were independent predictors of success on both typical verbal and low−verbaltheory of mind tasks.

In addition to these correlational studies, several studies have shown that training to increase children's facility with sentential complements leads to improvements in false belief performance. For example, Hale and Tager−Flusberg (2003) found that children trained on sentential complements not only acquired the linguistic knowledge fostered by the training, but also significantly increased their scores on a range of theory of mind tasks, including false belief tests. In a typical training session,an experimenter acted out a story with characters (e.g., Big Bird, Grover and a boy). In the story, the boy does one thing, but says that he does another. For instance, in one story, the boy is shown kissing Big Bird, but the boy says, "I kissed Grover." The experimenter then asks the child: "What did the boy say?" Correct responses were responded to with, "That's right. The boy said, 'I kissed Grover,' but he really kissed Big Bird." If the child made an incorrect response, the examiner acted out again and said, "But remember, the boy says, 'I kissed Grover,' but he really kissed Big Bird." The results showed that those who participated in the sentential complement training also improved in false belief performance relativeto a control group.

Hale and Tager−Flusberg's (2003) training protocol was specifically designed to determine whether it was the grammatical features of sentential complement structures, rather than other kinds of more semantic content, that

affected children's theory of mind development. Thus, these researchers used "communication"verbs (i.e., said that). However, it is not entirely clear whether the aim of leaving out semantic content was fully achieved. Their training involved a kind of deception in which participants were told that although a story character intentionally said one thing, something else was in fact true. The deceptive aspect of the story was not emphasized in the training protocol, but its presence makes it unclear as to whether the training promoted children's understanding of deception (which is relevant to theory of mind), sentential complement constructions, or both.

With this in mind, Lohmann and Tomasello (2003) developed a training protocol that avoided deceptive communication and obtained the same results, thereby suggesting that the deceptive content of the training was not the sole factor promoting false belief understanding in Hale and Tager−Flusberg's study. Yet, Lohmann and Tomasello's protocol had interpretive difficulties of its own. Specifically, their training involved the use of terms that could, arguably, be interpreted as having mental content (e.g., discussion of what story characters "feel" or "know"). Thus, it remains unclear whether the association between facilitating sentential complement understanding and false belief reasoning is specific to the cognitive−structural "template" that sentential complements provide, or the semantic content of the training.

Other concerns about the empirical relation between sentential complement facility and false belief have been raised by researchers working in languages other than English. For instance, Perner, Sprung, Zauner, and Haider (2003) noted that in German, sentences involving the mental verb want obligatorily take sentential complements; yet, German children (like others) appear to understand the entailments of want well before they understand the entailments of think and believe. Also, in studies with Cantonese−speaking preschoolers that are more analogous

to those conducted by de Villiers and colleagues, two groups have shown that once relevant factors are controlled (e.g., general language ability, age, prior theory of mind development), understanding of sentential complements does not make a unique contribution to false belief reasoning (Cheung et al., 2004; Tardif, So, & Kaciroti, 2007,study 2). Taken together, these findings suggested that the relation between sentential complement and false belief understanding may not extend across languages or cultures, which calls into questionits validity as a strong explanatory theory of acquisition of false belief understanding.

With these concerns in mind, the goal of the present study was to provide some clear evidence regarding whether facility with sentential complements is necessary for false belief understanding. We hoped to achieve this goal by conducting a training study with Mandarin speaking Chinese preschoolers. There are important differences between English and Mandarin with respect to the use of mental terms and sentential complements. Relative to English speakers, the use of mentalterms with sentential complements is rare among Mandarin− speaking parents and children (Tardif & Wellman, 2000; Snedeker & Li, 2000). Yet Mandarin−speaking parents and children use sentential complement constructions for communication verbs (e.g., say, in Mandarin) more commonly and earlier in development than their English− speaking counterparts (Tardif & Wellman, 2000).

This comparison raises two interesting possibilities. The first is that perhaps Mandarin−speaking preschoolers trained on sentential complements with the more common communication verbs will show more efficient acquisition of the construction relative to those trained on sentential complements with mental verbs. The second possibility is that training with sentential complements involving mental verbs may be less effective in facilitating children's acquisition of the structure because theuse of mental verbs with sentential complements is non−canonical and

potentially confusing. If so,this may allow for a fairly compelling test of the role that sentential complement understanding plays in false belief reasoning. Namely, if sentential complement understanding per se contributes to false belief reasoning, children who receive sentential complement training with communication verbs may perform better on false belief tasks than do children who receive sentential complement training with mental verbs. Investigating this hypothesis is the main goal of the present study.

A second goal is to gain evidence regarding the mechanisms by which sentential complement training may affect false belief understanding. Some have suggested that tests of false belief understanding and tests of sentential complement understanding both rely on an understanding of misrepresentationmore generally–that is, they both require one to consider that the propositional contents of a representation (either a communicative utterance or a belief) may not be consistent with the true state of affairs it is meant to represent (Ruffman, Slade, Rowlandson, Rumsey, & Garnham, 2003). Ifso, sentential complement training may be one way to increase children's understanding of misrepresentation, but perhaps not the only way (Ruffman et al., 2003). In the present study, we included for comparison a condition that trained children on concepts of misrepresentation without the use of sentential complements. Specifically, we borrowed a procedure first used by Wellman, Hollander, and Schult (1996), who investigated preschoolers' understanding of thought bubbles. They found that vast majority of 3–4–year–olds understand that thought bubbles can be used as graphical depictions of"what a person is thinking", and thus can reveal the contents of characters' thought in a variety of situations. More important, thought bubbles can be used to graphically depict a person's misrepresentation of some states of affairs (that is, the propositional content of the thought bubblecan be discrepant with some true state

of affairs). If facility with sentential complements per se is critical to false belief understanding,the misrepresentation training should have little effect on false belief reasoning independent of whatever effects it might also have on sentential complement facility. Alternatively, if sentential complement training has its effects through the more general mechanism of developing a broader understanding of misrepresentation, we might expect a relatively robust effect of misrepresentation training,irrespective of whether it also affects facility with sentential complements.

2 Methods

2.1 Participants

A total of 120 kindergarten children participated (mean age = 46.3 months, SD = 4.9months,range = 40–55 months). They came from kindergartens affiliated with two universities in Beijing,China. Most came from working– and middle–class families. All spoke Mandarin. According to parents'or teachers' reports, no children had linguistic or psychological abnormalities. Eighty–four children from this original sample failed at least one of two questions in false belief pre–test (see below) and thus were considered eligible for the training study. Post–test data from three participants were missing due to children's illness (n = 2) or the family relocating (n = 1). Thus, a full set of data was acquired for 81 children (33 boys). Each child received a gift for participation.

2.2 Design

We employed a between–subject design in which participants were randomly assigned to one of the four training groups: (1) sentential complement–communication verb (SC–COMM), (2) sentential complement–mental verb (SC–MENTAL), (3) false representation (FR), and (4) control.

2.3 Pre and posttest measures

2.3.1 False belief pretest

A standard unexpected contents false belief task was

Table 1 Mean ages and pretest performance across groups.

	Groups			
	SC−COMM(n = 20)	SC−MENTAL(n = 20)	FR(n = 20)	Control (n = 21)
Age (months)	45.35 (4.40)	47.80 (5.22)	45.45 (5.79)	46.70(4.20)
FB prediction (0−1)	0.20 (0.41)	0.15 (0.36)	0.15 (0.36)	0.14(0.35)
FB ignorance (0−1)	0.50 (0.51)	0.60 (0.50)	0.60 (0.50)	0.52(0.51)
MCwMV(0−1)	0.50 (0.51)	0.45 (0.51)	0.45 (0.51)	0.61(0.49)
MCwCV(0−1)	0.52 (0.50)	0.55 (0.51)	0.50 (0.51)	0.47(0.50)

Note: Sentential complement−communication verb = SC−COMM, sentential complement−mental verb = SC−MENTAL, false rep−resentation= FR.

used to assess false belief understanding(Perner, Leekam, & Wimmer, 1987). A child was shown a candy box and asked what he or she thought was inside. After children responded saying that they believed the box contained candy, the box was opened to reveal a ball−point pen. After closing the box, the experimenter asked a false belief question:When another child sees this closed box for the first time, what would he say is inside the box? This was followed by an ignorance question: Does he know what is in the box before it is opened? Children received a point for each correct answer.

2.3.2 *Sentential complement pretests*

Two tasks of memory for sentential complements were administered using the procedure from previous studies (de Villiers & Pyers, 2002; Lohmann & Tomasello, 2003). Children were told a story involving sentential complement with a mental verb (e.g., think) or communication verb (e.g., say) accompanying with line drawings relevant to the sentence. One story involving a mental verb was as follows, "A little rabbit was playing at home, and a wolf was knocking at the door; however, therabbit though it was his/her Mom knocking at the door". The test question on memory for sentential complement for the mental verb think was, "Who did the little rabbit *think* was knocking at the door?"

A story to assess understanding of sentential complements with communication verbs was as follows: Xiaohong was asked to go buy milk by her mother, but she bought a tin of Cola instead of milk. When she went home, her mother asked her, "Xiaohong, what did you buy?" Xiaohong said, "I bought milk". The test question on

memory for sentential complement for the communication verbsay was, "What did Xiaohong say that she bought?"

Children received 1 point for correctly answering the content of complement or for using the entire complement structure.

Children performed better than expected on the tests for sentential complements. Specifically,although four children in each training group did not pass either of the two questions, some children in all groups (4 children in SC−COMM, 4 in SC−MENTAL, 3 in the FR group and 6 in control) answered both correctly. Thus there were children in our final group who performed poorly on false belief tasks although they already showed facility with sentential complements. This pattern would be unexpected if sentential complement understanding were itself sufficient for false belief understanding. Nonetheless, average performance on the sentential complement pretest was equivalent across the different training groups (see Table 1), which allowed us to assess the effects of different kinds training on false belief understanding at the group level.

2.3.3 *False belief posttests*

Twostandard location−change tasks (Wimmer& Perner, 1983) and two standard unexpected content tasks (Hogrefe, Wimmer, & Perner, 1986; Perner et al., 1987) were administered to children following training.

In the location−change tasks, children were told stories. In one, a boy Xiaogang put his cake into a cupboard and then went out to play. Then his mother transferred the cake from the cupboard to the refrigerator. Children were asked two control questions (Where did Xiaogang put the cake before he left? Where is the cake now?). If a child did not

answer correctly the story was repeated up to three times. The false belief questions included a standard behavior prediction question (When Xiaogang comes home, where will he first look for the cake?) and a justification question (Why will he look for it there?). A justification was judged as appropriate if it included references to location (e.g., He put it there before he went out), knowledge (e.g., he does not know it was moved), or belief (e.g., He thinks it still is there). All other justifications were coded as inappropriate (e.g., I do not know; the cake is in the cupboard; he wants to find the ball; he cannot find the ball).

A second story involved a girl who put a ball in a box and in her absence, her classmate moved the ball to a drawer. The two unexpected contents tasks were similar to the false belief pretest except that the false belief prediction question was, "When another child comes in, before having opened it, what does he/she think is inside the box?"

In the false belief tasks, each correct answer to the behavior prediction question was awarded 1 point. Across the four trials, total scores for the prediction question ranged from 0 to 4.

2.3.4 Sentential complement posttests

The sentential complement posttests were identical in structure to the pretests. Children were orally presented ten brief stories in random order, each accompanied by line drawing pictures. Each story involved a character who made a mistake, told a lie, or held a false belief. Of the ten stories, five involved mental state verbs (e.g., think) and five involved communication verbs (e.g., say). The test questions assessed children's memory for a sentential complement presented with each of the verbs. Based on previous work (de Villiers & Pyers, 2002), our Mandarin version was revised slightly for Chinese children (Cheung, Chen, & Yeung, 2009). Two of the stories were as follows:

Story 1: He thought he found his ring, but it was really a bottle cap.

Test question: What did he think he found?

Story 2: Mom asked Limei to buy some milk. But Limei bought some orange juice. Mom asked Limei, "Did you buy milk?" Limei said, "Yes, I bought milk."

Test question: What did Limei say she bought?

As with the pretest, children's responses were scored as correct if they included the content of the complement or use of the entire complement structure (de Villiers & Pyers, 2002; Perner et al., 2003). Children received 1 point for each correct answer, resulting in a maximum score of 5 each for sentential complements involving either mental or communication verbs.

2.4 Training protocols[1]

Children interacted with one of three adult experimenters alone in a quiet room in their preschools. The two training sessions took place within two weeks of one another, with each session lasting about 25 min (range: 20–30 min), with a 7–day interval between sessions. Posttests were administered about 4 days after the second training session.

2.4.1 Sentential complement training

The training procedure was modeled on that used by Lohmann and Tomasello (2003) and was revised slightly to make it culturally appropriate for Chinese children. In addition, in order to stimulate children's interest in talking about some topics and to improve the ecological validity of the study, the sentential complement training was conducted in an elaborated, conversational manner.

Each of the two sessions consisted of four trials and lasted about 25 min. In the first session, trials involved the experimenter presenting an object (e.g., apple) which was wrapped in a piece of thin paper and letting children touch it. Children were then asked about characteristics

---

[1] Mandarin materials are available on request from the anthors

of the objects with questions that included a sentential complement whose verb depended on condition (mental vs. communication). The experimenter repeated or corrected children's responses to the questions. For example, if the question was, "What do you think/say this is?" the feedback was, "Okay, you think it is an apple. Yes, I also think it is an apple."

In the second session, trials involved the experimenter telling children two short stories accom panied by relevant pictures. One story was as follows. "Xiaoming wants to get a toy airplane for his birthday gift. On his birthday, his mother bought him a toy airplane. Xiaoming thought/said that his mother bought him a toy airplane. He thought/said that his mother loved him very much. After each story, the experimenter asked the child questions regarding the parts of the story that involved sentential complements (e.g., "Can you tell me, what did Xiaoming think/say his mother bought him?" The experimenter repeated or corrected children's answers. For example, the experimenter said, "Right, Xiaoming thought/said that his mother bought him a toy airplane."

2.4.2 False representation training

A training paradigm based on the "thought bubble" experimental paradigms discussed earlier was adapted for use with Mandarin−speaking children (Flavell, Everett, Croft, & Flavell, 1981; Wellman et al., 1996).

To begin, an experimenter presented children with three black and white line drawings, two of which depicted a familiar object (e.g., horse and turtle); the third was a "thought bubble" picture, i.e., a set of bubbles appearing above the right side of the pictured child's head. The experimenter first presented children with the thought bubble picture and said, "Look at the large bubble that contains a horse. It shows what is in the character's mind. Have you ever seen this kind of thought bubble in cartoons?" Children typically demonstrated understanding of the relation between thought bubbles and mental states

following this description. Children were then shown scenarios in which the contents of thought bubbles did not match some true state of affairs.

The experimenter presented the first picture (e.g., of a horse) and asked children to name the object. Then, presenting the "thought bubble" picture again, the experimenter said to the child, "When this boy/girl saw this picture, does he/she know there is a horse in this picture?" After children answered correctly, the experimenter continued, "Well now this boy/girl turned away from us." (The thought bubble picture was turned away.) The experimenter then showed the second picture, saying, "Look at this picture (showing the second object, e.g., a car), replacing the first picture with this one and asking, 'What is the object in this picture now?' " The experimenter then continued, "Well, the boy/girl turned and does not see the object in this picture." Children were then asked, "Now, what is the object in this picture in his/her mind?"

If children answered correctly, the experimenter asked them to explain why. If not, the experimenter explained that the boy/girl had not seen the picture being replaced and then asked the question again. If children did not answer correctly (i.e., by naming the object in the second picture), the experimenter asked, "Is he/she right (or wrong)? Why?" If children answered incorrectly, the experimenter asked, "If you do not see the picture, do you know what it is in the picture (the experimenter turned the picture away)?"

The false representation training included two sessions, each consisting of four trials, with each trial lasting about 5−7 min.

2.4.3 Control group

The control training condition was based on the procedure used by Lohmann and Tomasello (2003) and the same objects as in the sentential complement training condition. However, the experimenter did not ask

Table 2 Performance on posttest tasks across groups (mean and SD).

	Groups			
	SC−COMM(n = 20)	SC−MENTAL(n = 20)	FR(n = 20)	Control (n = 21)
FB prediction(0–4)	2.45 (1.50)	2.20(1.47)	2.45(1.38)	1.28 (1.27)
MCwMV (0–5)	3.25(1.21)	3.55 (1.14)	2.45(1.43)	2.48 (1.28)
MCwCV (0–5)	4.25 (0.96)	3.80 (0.95)	3.50 (1.27)	3.61 (1.07)

Table 3 Percentage of participants giving appropriate or inappropriate reasons in cake story task.

	Groups			
	SC−COMM(n = 20)(%)	SC−MENTAL(n = 20)(%)	FR(n = 20)(%)	Control (n = 21)(%)
Appropriate justification				
Location or knowledge	40	35	20	9.5
Mental state	10	15	20	9.5
Inappropriate justification	50	50	60	81

children questions involving sentential complement with communication or mental verbs,and only asked simple questions with clear answers to keep their attention. Children's responses were given neutral feedback, although an effort was made to keep children engaged. Finally, an effortwas made to ensure that the control training episodes were roughly as long as in the other training conditions.

3 Results

3.1 Preliminary analyses

Children's mean ages and average performance on the various pretests in each group are presentedin Table 1. A one−way analysis of variance (ANOVA) and Kruskal−Wallis tests showed no significant age differences or any of the pretests across the four groups. Nor were there significant gender differences. These preliminary analyses demonstrated that the four groups were indeed equivalent before training.

3.2 Training effect on false belief posttests

3.2.1 Prediction measures

Performance on the prediction questions in the two false belief tasks (contents and location) was significantly correlated, $r(81) = 45$, $p<.001$. Thus, scores on the two tasks were summed to obtain a total score for each child, ranging from 0 to 4 (see Table 2).

A one−way ANOVA, with training condition as

independent variable and the aggregated false belief post−test scores as dependent variable, revealed a significant training condition effect, $F(3, 77)=3.14$, $p=0.03$, partial $\eta^2=0.10$. Post hoc tests revealed that the SC−COMM group and the FR group performed significantly better than the control group (Tukey HSD, $ps<.05$). The SC−MENTAL group did not sig nificantly outperform the control group (Tukey HSD, $p>.1$). No other significant differences were found.

Did improvement reach a level of above−chance responding? Tests comparing posttest performance in each group to chance (i.e.$\mu = 2.00$) showed that the control group performed significantly belowc hance, $t=2.57$, $df=20$, $p=0.018$. Performance in the focal training groups did not exceed chance.

Tables 3 and 4 showed the percentage of children by group who gave appropriate explanations for their responses in the two location−change tasks. In the "cake story" task, both sentential complement training groups outperformed the control group, $x^2 (1, N=41)=4.36$, $p=0.037$, for both comparisons. In the "ball story" task, however, only the SC−COMM group gave more appropriate reasons than thecontrol group, $x^2 (1, N=41)=5.53$, $p=0.019$.

3.3 Training effect on sentential complement posttests

The scores of the two sentential complement tests (mental and communication verbs) were significantly

Table 4 Percentage of participants giving appropriate or inappropriate reasons in ball story task.

	Groups			
	SC—COMM(n = 20)(%)	SC—MENTAL(n = 20)(%)	FR(n = 20)(%)	Control (n = 21)
Appropriate justification				
Location or knowledge	40	40	35	19.1
Mental state	20	5	10	4.7
Inappropriate justification	40	55	55	76.2

correlated, $r(81)$ =0.25, p=0.02 (see Table 2) and therefore combined. Thus, composite scores ranged from 0 to 10. A one—way ANOVA with condition as independent variable showed a significant main effect of condition, $F(3, 77)$= 3.99, p=0.01, partial η^2=0.11. Post hoc comparison revealed that the SC—COMM group and SC—MENTAL group outperformed the FR group and control group (Tukey HSD,the SC—COMM group > control group, p= .016; SC—MENTAL group > the control group, p=0.03; SC—COMMgroup>FR group, p=0.009; SC—MENTAL group >FR group, p=0.018). No other significant differences were found (see Fig1). Thus, sentential complement training promoted facility with complement syntax.

3.4 Reduced sample analyses

As noted, several children in each group passed the sentential complement pretests. This made it difficult to test the hypothesis that the sentential complement training involving communication verbs would be more effective than complement training involving mental verbs at promoting both general sentential complement understanding and false belief reasoning. To address this difficulty,we performed some exploratory analyses to determine whether the above patterns would remain when children who performed perfectly on sentential complement pretests were removed from the analysis. A one—way ANOVA with false belief prediction score across all four false belief tasks as dependent variable demonstrated a significant effect of group, $F(3, 60)$=3.09, p=0.03. Tukey HSD testshowed that the SC—COMM group (n=16, M= 2.25, SD =1.57) outperformed the control group (n=15,M= 0.93, SD=1.03, p=0.05), and FR group (n=17, M =2.29, SD=1.49) performed better than the controlgroup (p<.05).
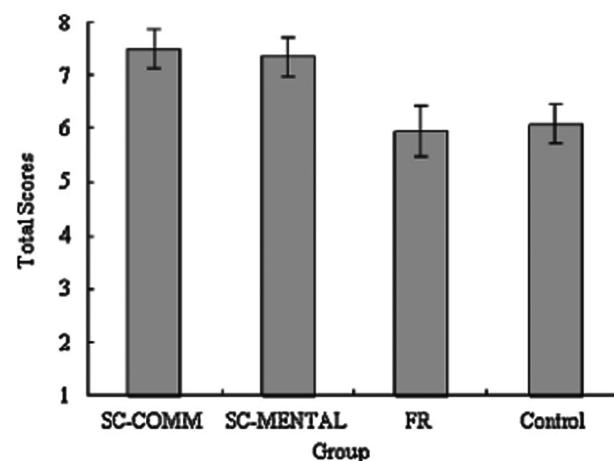
Fig. 1 Total posttest score on sentential complements by group. Note: Error bars represent standard error.

There was no significant difference between the SC—MENTAL group (n=16, M=2.12,SD=1.58) and the control group (p>.10). A similar one—way ANOVA with sentential complementposttest scores as dependent variable showed a significant effect of condition, $F(3, 60) = 3.56$, p=0.02. Post hoc tests showed significant differences between the SC—COMM group (M=7.37, SD=1.70) and FRgroup (M=5.76, SD=2.27) (p=01), between the SC—MENTAL group ($M =$ 7.37, SD=1.78) and FR group(p=0.01), between the two sentential complement training groups and the control group (M=6.00,SD=1.46, ps<.05). There was no significant difference between FR and control groups.

4 Discussion

This study explored the extent to which sentential complement and false representation comprehension training affected false belief understanding among Mandarin—speaking children. First, we found that training with sentential complement with communication verbs and

false representations significantly improved performance on false belief posttests. Second, only the sentential complement training groups performed better than the control group on the sentential complement posttests.

Children who were trained with sentential complements involving communication verbs showed greater improvement on false belief posttests than did children trained on sentential complements with mental verbs. This asymmetry was predicted for Mandarin−speaking children because of the differences in the kinds of verbs that typically take sentential complements in Mandarin. Mandarin−speaking children hear and use sentential complements with the communication verb "*Shuo*1" [/say/] earlier and more commonly than they do with mental verbs (Snedeker & Li, 2000; Tardif & Wellman,2000). Thus, sentential complementtraining with a communication verb may have provided a more straightforward opportunity for children to capitalize on whatever structural benefits acquiring the sentential complement construction has for reasoning about false beliefs.

These findings have implications for understanding of the mechanisms by which sentential comple ment training might promote false belief understanding. The literature to date left open the possibility that the sentential complement training protocols used by past researchers unintentionally provided informationrelevant to mental state reasoning (such as using deceptive objects or talking about story characters' preferences and feelings). Here, we designed our training to avoid this possibility and exam− ined a group of children who showed a greater benefit of sentential complement training involving verbs that were explicitly non−mental. These findings clarify that sentential complement training can confer benefits on false belief reasoning even when there is no (or minimal) mental state conten twithin the training.

Thus, these findings are generally in line with the view that mastery of sentential complements may provide an important template for reasoning about others' mental states (de Villiers& Pyers,2002). This view, sometimes called the "syntactic enrichment" view, is that as children acquire the syntacticstructures that allow for embedding one thought in another (e.g., embedded propositions),they gain a format that can facilitate reasoning that requires the explicit separation of two different perspectives on a single situation (de Villiers & Pyers, 2002; Harris, de Rosnay, & Pons, 2005). Our findings here are, at least in part, consistent with this view,given that our training protocols that involvedcommunication verbs were carefully designed toavoid false−belief relevant content but they still promoted false belief understanding.

The enhanced effectiveness of sentential complement training with communicative verbs relative to mental verbs was predicted specifically for Mandarin−speaking preschoolers given the relative frequency with which sentential complements are used for the two verbs in everyday language. An open question is whether, if investigated more directly, a similar advantage might be seen for children who speak other languages. de Villiers and her colleagues argued that an early emerging facility with sentential complements involving communication verbs might provide children with a sort of syntactic bootstrap to developing a more sophisticated (i.e., representational) understanding of mental verbs that are used in sentential complement formats. Evidence from Lohmann and Tomasello (2003) suggests that training with both communication and mental verbs is equally effective and thus that the advantage for communication verb training is specific to Mandarin for the reasons mentioned. Nonetheless, given that our training was designed with the explicit goal of separating the contents of communication and mental verb training, it may be worthwhile to apply our paradigm to other languages to see whether the advantage for communication verbs may be more general, for the reasons described by de Villiers (2003).

A final note about our training protocols concerns the particular mental verbs we used in the training and false belief task protocols. Like some other languages, Mandarin includes several verbs for "think" and "believe". One such verb is/yi3−wei2/which is usually translated as "think falsely". Lee,Olson, and Torrance (1999) found that Mandarin−speaking children performed significantly better in the false belief tests when this false belief verb was used, compared with more neutral verbs (see Cheung et al., 2009, for a similar finding in Cantonese). In the present study, we used the neutral mental verbs/ren4−wei2/and/jue2−de/during the sentential complement training and false belief tasks,respectively. Our choice to use these rarer, semantically neutral verbs may have weakened the extent to which children capitalized on the training protocols. Yet, this caveat does not negate the main finding, which is that sentential complement training involving communication verbs can benefi tfalse belief reasoning, even when that training has no explicit or implicit mental content.

We also found that children in the false representation group showed evidence of improved false belief performance relative to the control group, although the false representation training group did not show any measurable gains in their facility at processing sentential complements. At some level,this finding was expected. As previous work had shown (Wellman et al., 1996), training paradigms that use thought bubbles to emphasize the representational nature of mental states can improve even 3−year−olds' performance on standard false belief tasks. Similar findings have been shown with individuals with autism, a neurodevelopmental disorder in which false belief reasoning is particularly affected (Wellman et al., 2002). Our results extend these findings and clarify that the benefits children receive from thought bubble training likely do not have their effects through promoting facility with sentential complement constructions. In doing so,

our findings are consistent with those of Hale and Tager−Flusberg (2003), who showed that training on false belief only (i.e., without exposure to sentential complement training) improved false belief performance but not sentential complement understanding. We see these findings as having two important implications. First, they show that the ability to reason about false representation is not sufficient for processing sentential complement syntax (Ruffman et al., 2003). Second, they show that explicit false belief understanding can emerge in the absence of sentential complement understanding (Hale & Tager−Flusberg, 2003; Lohmann &Tomasello, 2003).

It should be noted that the false representation training protocol was not fully non−linguistic; we did use a neutral Mandarin epistemic mental verb that we felt was appropriate to the situation (i.e.,/zhi1dao4/). Thus, we do not wish to argue against the possibility that naturalistic exposure to mental verbs plays some role in children's false belief development across cultures. Our argument, rather, is that making explicit the representational nature of mental states was effective in promoting false belief understanding, although it apparently did not affect children's facility with sentential complements.

A limitation of the present study is that we have no assessments of children's cognitive and lin−guisticcapacities beyond those investigated. Children's general linguistic abilities arerelated to both sentential complement understanding and false belief performance in Chinese− and English−speaking preschoolers (Tardif et al., 2007). By knowing more about how the training protocols that we used interact with general linguistic abilities, we mightgain greater insight into the specific mechanisms by which these interventions affected false belief development. Also, we did not include measures of representational understanding other than false belief understanding, such as measures of false photograph or false sign performance (Sabbagh, Moses, & Shiverick,

2006). Doing so might have allowed us to determine whether the training protocols that promoted false belief understanding had their effects on representational understanding more broadly or more specifically on mental state understanding.

In conclusion, the present study supports the view that training protocols aimed at improving children's facility with sentential complements can promote false belief understanding, even when the training protocols are structured so as not to include explicit or implicit mental state content. Indeed,there was some evidence that with Mandarin−speaking children, sentential complement training that involved communication verbs was more effective than training with mental state verbs. Yet the findings also showed that false belief performance can be improved by a training protocol that involved no training with sentential complements. Together, these findings suggest that sentential complement understanding may be one path that contributes to false belief understanding, the importance of which may depend on particular sociocultural and linguistic factors.

References

Astington, J. W., & Jenkins, J. M. (1999). A longitudinal study of the relation between language and theory−of−mind development.*Developmental Psychology, 35*, 1311 – 1320.

Cheung, H., Chen, H.−C., Creed, N., Ng, L., Wang, S. P., & Mo, L. (2004). Relative roles of general and complementation languagein theory−of−mind development: Evidence from Cantonese and English. *Child Development, 75*, 1155 – 1170.

Cheung, H., Chen, H.−C., & Yeung, W. (2009). Relations between mental verb and false belief understanding in Cantonese−speakingchildren. *Journal of Experimental Child Psychology, 104*, 141 – 155.

de Villiers, J. G. (2003). Getting complements on your mental state (verbs). In J. Van Kampen, & S. Baauw (Eds.), *Proceedings of2003 GALA conference* (pp. 13 – 26). Utrecht: LOT.

de Villiers, J. G., & de Villiers, P. A. (2000). Linguistic determinism and the understanding of false beliefs. In P. Mitchell, & K. J.Riggs (Eds.), *Children´s reasoning the mind* (pp. 191 – 228). Hove, England: Psychology Press.

de Villiers, J. G., & Pyers, J. E. (2002). Complements to cognition: A longitudinal study of the relationship between complexsyntax and false−belief−understanding. *Cognitive Development, 17,* 1037 – 1060.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidencefor the level 1 – level 2 distinction. *Developmental Psychology, 17*, 99 – 103.

Hale, C. M., & Tager−Flusberg, H. (2003). The influence of language on theory of mind: A training study. *Developmental Science,6*, 346 – 359.

Harris, P. L., de Rosnay, M., & Pons, F. (2005). Language and children's understanding of mental states. *Current Directions inPsychological Science, 14*, 69 – 73.

Hogrefe, G.−J, Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemicstates. *Child Development, 57*, 567 – 582.

Hughes, C., & Dunn, J. (1998). Understanding mind and emotion: Longitudinal associations with mental−state talk betweenyoung friends. Developmental *Psychology, 34*, 1026 – 1037.

Lee, K., Olson, D. R., & Torrance, N. (1999). Chinese children's understanding of false beliefs: The role of language. *Journal of ChildLanguage, 26*, 1 – 21.

Lohmann, H., & Tomasello, M. (2003). The role of language in the development of false belief understanding: *A training study.ChildDevelopment, 74*, 1130 – 1144.

Milligan, K., Astington, J. W., & Dack, L. A. (2007). Language and theory of mind:Meta-analysis of the relation between languageability and false-belief understanding. *Child Development, 78*, 622–646.

Perner, J., Leekam, S., & Wimmer, H. (1987). Three-year olds' difficulty with falsebelief: The case for a conceptual deficit. *BritishJournalof Developmental Psychology, 5*, 125–137.

Perner, J., Sprung, M., Zauner, P., & Haider,H. (2003). Want That is understood well before Say that Think that, and false belief:A test of de Villiers's linguistic determinism on German–Speaking children. *Child Development, 74*, 179–188.

Ruffman, T., Slade, L., Rowlandson, K., Rumsey, C., & Garnham, A. (2003). How language relates to belief, desire, and emotionunderstanding. *Cognitive Development, 18*, 139–158.

Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive functioning and preschoolers' understanding of false beliefs, falsephotographs, and false signs. *Child Development, 77*, 1034–1049.

Schick, B., de Villiers, P., de Villiers, J., & Hoffmeister, R. (2007). Language and theory of mind: A study of deaf children. *ChildDevelopment, 78*, 376–396.

Snedeker, J., & Li, P. (2000). Can the situations in whichwords occur account for cross-linguistic variation in vocabulary com-position? In J. Tai, & Y. Chang (Eds.), *Proceedings of the seventh international symposium on Chinese languages and linguistics.*

Tardif, T., So, C. W.-C., & Kaciroti, N. (2007). Language and false belief: Evidence for general, not specific, effects in Cantonese-speaking preschoolers. *Developmental Psychology, 43*, 318–340.

Tardif, T., & Wellman, H. M. (2000). Acquisition of mental state language in Mandarin- and Cantonese-speaking children.*Developmental Psychology, 36*, 25–43.

Wellman, H. M., Baron-Cohen, S.,Caswell, R., Gomez, J. C., Swettenham, J., Toye, E., et al. (2002). Thought-bubbles help childrenwith Autism acquire an alternative to a theory of mind. *Autism, 6*(4), 343–363.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *ChildDevelopment, 72*, 655–684.

Wellman, H. M., Hollander, M., & Schult, C. A. (1996). Young children's understanding of thought bubbles and thoughts. *ChildDevelopment, 67*, 768–788.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in youngchildren's understanding of deception. *Cognition, 13*, 103–128.

Psychological development and educational problems of left-behind children in rural China

Xiaojun Sun[1,*], Yuan Tian[1,*], Yongxin Zhang[1], Xiaochun Xie[1], Melissa A. Heath[2], Zongkui Zhou[1]

([1] Central China Normal University, China)

([2] Brigham Young University, USA)

Abstract　With China's rapidly developing economy and increasing urbanization, many adults from rural areas migrate to urban areas for better paid jobs. A side effect of this migration is that parents frequently leave their children behind (left–behind children). This research investigated left–behind children's and non–left–behind children's psychological, behav–ioral, and educational functioning. Survey participants included 1,708 adolescents (54.8%female; mean age=15.03 ± 1.93 years) from rural areas in Central China. Additionally,32 left–behind children and 32 head teachers were interviewed. Data indicated that in comparison to non–left–behind children, left–behind children were at a disadvantage in regard to emotional adjustment (i.e. lower life satisfaction, lower self–esteem, and higher depression), but fared better in educational adjustment (greater school engagement). Mitigating factors which positively influenced outcomes of certain subgroups of left–behind children included the presence of one parent, increased parental contact,and shorter length of time since parental migration. Information gathered from interviews with LBC also indicated adverse effects of parent absence on children's develop–ment. Teachers identified education measures and support offered to left–behind children and reported difficulties in communicating with parents. Based on this study's findings, and considering the perspective of educators, implications for school–based interventions are explored.

Keywords　Behavior problems; education; left–behind children; migration; psychosocial development; relationships; rural China; school engagement; teacher perspective

1 Introduction

Since the end of the 1970s,China's economy has rapidly expanded due to policy reform and opening up to international trade markets (Jaggi, Rundle, Rosen, & Takahashi,1996; Morrison, 2014). This economic growth has also promoted the process of urbanization, creating a significant influx of rural surplus laborers who pour into cities looking for better jobs. These individuals leave their

rural homes and become migrant workers. This massive rural–to–urban migration constitutes the largest migration in human history (Zhang, 2004). Unfortunately, because of long work hours in the city,destitute living conditions, insufficient income, and the restrictions associated with the binary system between China's rural and urban areas, many migrant workers are forced to leave their children behind in their rural hometown (Luo, Wang, & Gao, 2009).

Parents leaving children behind while one or both of the parents migrate for work is also very common in other countries, such as Japan (Carandang, Sison, &Carandang, 2007; Edillon, 2008; Melgar & Borromeo, 2002), the Philippines(Reyes, 2008), and Mexico (Bryant, 2005; Reyes, 2008; Tarroja & Fernando,2013; Yeoh & Lam, 2007). In China, many children are left in their rural hometown by their parents (one or both) who are hunting for more lucrative work in urban areas. These children are defined by Chinese scholars as left–behind children (LBC) (Luo et al., 2009; Zhao & Shen, 2010).

In recent years, China's migrating population has increased dramatically. Simultaneously, the number of LBC has also increased dramatically. Based on the results of China's Sixth National Population Census, nationwide, there were approximately 69.7 million LBC. The vast majority of China's LBC, 61 million, live in rural areas. In fact, rural LBC account for 87.5% of China's total number of LBCand 21.9% of China's total number of children (Duan, Lu, Guo, & Wang, 2013).

Beyond sharing some common vulnerabilities with unattended and disadvantaged children living in other countries around the world, China's rural LBC also face additional challenging situations. China's dual rural–urban system makes it especially difficult for children to migrate with their parents. Because of the serious imbalance of regional economic development, rural LBC are commonly faced with difficulties associated with low socioeconomic status and the shortage of educational resources.

However, similar to other countries' LBC, the large number of China's LBC continues to be considered a fairly small proportion of the massive population base. This topic has not been adequately researched, nor have these children's needs been adequately addressed (Liu & Wang, 2010; Luo et al., 2009).

By placing this topic in a world–wide context, understanding the developmentand relevancy of China's LBC may contribute to advancing research in the broad overarching field of understanding disadvantaged children's needs. Ultimately this knowledge will assist in identifying and providing appropriate support to address LBC's needs and perhaps the needs of children in other contexts who have been'left behind'. The present study aims to describe China's rural LBC in terms of basic characteristics of children's psychological development and how LBC's present situations relate to and affect their education and social–emotional well–being.

Additionally, research is needed to address serious concerns associated with LBC. For example, due to the lack of parental supervision and care, elevated rates of suicide and sexual abuse are reported for China's rural LBC (Zhou,Wang, & Hong, 2010). The media's criticism of these situations not only exposes LBC's frequent exposure to malignant events, but also to the family's and com– munity's failure to provide adequate supervision of LBC. The media's spotlight on these issues brings the topic of LBC to the visible forefront, fueling criticism aimed directly at Chinese society (Shen, 2009).

In addition to concerns related to failed supervision of LBC, psychological research is keenly interested in the effect of long–term parent–child separation and the ensuing maladjustment of children's social development. Previous studies in other countries and cultures have indicated that paternal absence is correlated with children's maladjustment (Amato & Keith, 1991; Carandang et al., 2007;Cronk, Slutske, Madden, Bucholz, & Heath, 2004).

For example, one research study conducted in China found that left-behind girls were more likely to be unhappy, to contemplate suicide, and to consider leaving home (Gao, Li, Kim,Congdon, Lau, & Griffiths, 2010). Researchers in other countries, such as Indonesia and Mexico, who explored the impact of transnational migration, also found impeded social and psychological development among LBC and a higher incidence of mental disorders (Aguilera-Guzman, de Snyder, Romero, & Medina-Mora, 2004; Hugo, 2002).

Therefore, regardless of whether parents' migration occurs within-country or transnationally, research indicates that separation from parents appears to negatively impact LBC's development and adjustment (Wen & Lin, 2012; Zhao, Liu, &Zhang, 2013). Hence, from an applied frame of reference, it is of practical significance to explore LBC's development from a psychological and educational perspective, focusing on the relationship between the child's individual development and their environment. Investigating these aspects of child development will assist educators in better understanding LBC's needs, helping guide intervention efforts and helping identify and implement teaching strategies aimed at strengthening both healthy psychological development and academic achievement.

In China, the family migrant statuses include non-migrant family, one-parent migrant family, and both-parent migrant family. Another aspect to consider is *who* offers supervision and care in the absence of one or both parents. Caregivers of LBC are divided into five groups: (a) parent who did not migrate—typically the mother; (b) grandparents; (c) kith and kin (extended family); (d) older sibling; and(e) no caregiver in situations where LBC care for themselves (typically adolescent LBC). Other factors which are commonly considered by researchers include the length of time since parental migration (Wei & Chen, 2010), frequency of parental contact (Chen & Xie, 2007), and economic support from parents (Wang, Hu, Shen,2011).

2 Research questions

The following questions will be explored in this study: (a) when taking into account the migrant statuses and demographic characteristics of China's rural LBC, what are the differences in how parental migration impacts these groups?; (b) what aspects of child development are affected by parental migration?; (c) how do LBC describe their living situation and challenges?; and (d) what is the role of school-based education in regard to LBC's development?

Previous research indicates that key areas of children's development include emotional adaption; behavioral development; participation in school activities(school engagement; Fredricks, Blumenfeld, & Paris, 2004); and social relationships (Attili, Vermigli, & Roazzi, 2010; Lin, Fan, Li, & Pan, 2010; McLanahan &Sandefur, 1994; Xiang, 2007; Zhou, Sun, Liu, & Zhou, 2005). In order to examine aspects of rural LBC's emotional adjustment, this study investigated the following emotions: Loneliness, depression, and happiness. Additionally, self-esteem was examined to determine whether being left behind influences children's feelings of self-worth.

Children's problem behaviors were also investigated because difficulties with behavioral adjustment negatively impact both physical and mental health (Fang, Zhen, & Lin, 2001; Qu & Zou, 2009). Previous research indicates that limited parental monitoring is closely related to an increase in children's problem behaviors (Beck, Boyle, & Boekeloo, 2003; Diclemente, Wingood, & Crosby, 2001;Unnever, Cullen, & Travis, 2003), so this study investigated whether LBC whose parent(s) migrated exhibit more problem behaviors than non-LBC. In addition,parent-child relationships and peer relationships—identified as important social relationships—are the main source of adolescents' daily interaction and have a profound influence on adolescent

psychological adaptation (Attili et al., 2010;Zhao et al., 2013; Zou, 1998).

In a recent nationwide study of Chinese children and adolescents the above indicators were used to describe the basic characteristics of psychological development (Dong & Lin, 2011). These indicators are well represented in the existing research base. The present study, using the method of psychological measurement, examined the similarities and differences between LBC and children living in non-migrant families. More specifically similarities and differences were investigated in the following areas: Psychological aspects, behavioral aspects, educational outcomes, and social relations.

In order to further understand perceptions of LBC's personal lived experiences and education status, 32 LBC and 32 head teachers were interviewed. The quantitative and qualitative data gathered from this research were considered and used as a basis for making recommendations to assist educators in more adequately meeting LBC's social-emotional, behavioral, and educational needs.

3 Method

3.1 Participants

A total of 1,708 adolescents (54.8% female, mean age=15.03 years; SD=1.93) from two junior high schools (contributing 42.3% of total participants) and two high schools (contributing 57.7% of total participants) were asked to complete an anonymous questionnaire. Participating schools were located in rural areas of Central China.

Of the adolescents who participated, 600 were non-left-behind children (non-LBC) and 1,108 (64.9%) were LBC. Of the LBC participants, 547 (49.4%) were from families of both-parent migration and 561 (50.6%) were from families with one-parent migration.

To supplement information gathered from the questionnaire data, 32 LBC and 32 head teachers who worked with this study's participants were interviewed. These 32 students were selected according to their gender, migrant status, and school performance: 16 boys and 16 girls; 16 from one-parent migration family and 16from both-parents migration family; and 11 with good school performance, 11 with moderately good school performance, and ten with poor school performance.

The 32 head teachers were selected from participating classrooms; their mean age was 38.3 years; and their mean number of years teaching was 17.2 years. However, six teachers reported having less than six years of teaching experience. Minimally, all interviewed teachers had an undergraduate college degree.

3.2 Measures

Loneliness was assessed with Children's Loneliness Scale (Asher & Wheeler, 1985;Bagner, Storch, & Roberti, 2004). The scale consists of 24 items: 16 scored items and eight additional items related to participants' interest. These eight items were intended to help relax the participants when answering questions and were not included in the scored items. Ten of the 16 items were reverse scored, and were transformed before the data were analysed. Participants recorded their responses on a five-point Likert scale, ranging from 1 (always true) to 5 (not at all true). The scale's eight items are internally consistent (Cronbach's α =0.90; Asher & Wheeler,1985; Bagner et al., 2004).

Children's depression was assessed with the Chinese version of the Children's Depression Inventory (CDI; Kovacs, 1992; Yu & Li, 2000). Children were asked to endorse one of three descriptions that best applied to him or her during the last 2 weeks (e.g. 'I feel like crying every day'; 'I feel like crying many days'; 'I feel like crying once in a while'). In this study, the scale contained 26 items, and participants' responses were scored on a three-point

Likert scale ranging from 0 (option which described the least often occurring) to 2 (option which described the most often occurring). Based on previous data describing the CDI, the 26 items are internally consistent (Cronbach's α =0.91).

In addition to a single question ('Do you feel your life is happy in general?'), life satisfaction was measured with the revised Student's Life Satisfaction Scale (SLSS;Huebner, 1991; Huebner, Suldo, & Valois, 2003). Items were rated on a six–point Likert scale ranged from 0 (low) to 5 (high). This scale's internal consistency is considered adequate (Cronbach's α =0.76; Kovacs, 1992; Yu & Li, 2000).

Self–esteem was measured on a ten–item scale, the Self–Esteem Scale (SES;Rosenberg,1965,1979;Schmitt& Allik,2005).However,the eighth item was deleted because of its low validity for Chinese participants (Tian, 2006). The participants' responses were reported on a four–point Likert scale, ranging from 1 (strongly disagree) to 4 (strongly agree). This scale's internal consistency is considered adequate(Cronbach's α =0.84; Baumeister, Campbell, Krueger, & Vohs, 2003; Rosenberg,1965, 1979; Rosenberg, Schooler, Schoenbach, & Rosenberg, 1995).

Problem behaviors were captured by asking the participants to rate each statement as to how often they participated in the described behavior. Questions included asking the student whether or not they participated in socially unacceptable behaviors (e.g. smoking, bingeing on alcohol, or other socially unacceptable behaviors). On this measure, response options range from 1 (not at all) to 4 (very often). This measure included 16 items and has an acceptable reliability(Cronbach's α =0.76; Fang, Li, & Dong, 1996).

School engagement was assessed with a scale of five items. To each of the five items, participants responded with either a 1 (Yes) or 2 (No). Adolescents reported whether or not they followed school rules, followed classroom rules, enjoyed homework, enjoyed going to school, and answered questions in class. Internal reliability of this measure is considered low (Cronbach's α =0.46; Wen & Lin, 2012).

Peer relationships were measured by peer nomination of peer rejection and acceptance. Students were asked to nominate three of their classmates that they liked most and three of their classmates that they liked least. Then, the total number of 'like most' and 'like least' obtained for each child was divided by the total number of students in the class. Proportion of positive nomination minus proportion of negative nomination indicates social preference and the higher the score, the greater the child's popularity. This method of assessing peer relationships is commonly used in Chinese research studies (Zhao et al., 2013).

Parent–child cohesion was measured by the father–child cohesion (Cronbach's α =0.65) and mother–child cohesion (Cronbach's α =0.65) subscales of the Family Adaptation and Cohesion Evaluation Scale (Joh, Kim, Park, & Kim, 2013; Olson,Sprenkle, & Russell, 1979). This portion of the test consisted of two parts (father scale and mother scale) with similar items. Responses were on a five–point Likert scale and ranged from 1 (mostly disagree) to 5 (mostly agree). Examples of the statements included, 'My father/mother are supportive to each other in difficulties';'My father/mother feel intimate to each other'.

Demographic variables included gender (male or female) and education level(junior high school or high school). Additionally adolescents were described as either an only child or one of several children. The participant's birth order was also designated. Finally, to assess each LBC's specific family situation, children were asked to respond to four questions. These questions included parental migration status (one–parent migration, both–parent migration, and no–parent migration); length of time since parent's migration; relationship to primary caregiver;and frequency of parental contact.

3.3 Procedure

Data were collected during spring 2013. Data were collected from two age groups:(a) middle school students from the 7th and 8th grades and (b) high school students from the 10th and 11th grades. More specifically, four classrooms from each of the identified grades in each rural school were randomly selected. Altogether, 32 classrooms were selected. Cluster surveys were used and participants answered anonymously. All questionnaires were administered in school classrooms and were collected on site. Prior to participants agreeing to participate (assent), information about the purposes of the survey and confidentiality of responses were explained. Students were informed that there were no correct or incorrect answers. Extra support was provided for students who had difficulty completing the surveys.

To gather additional information, 32 LBC, one student from each of the 32 classrooms, participated in an interview. The interview consisted of questions regarding the participating adolescent's perception of current life and the changes since their parent's or parents' migration. Additionally, 32 teachers, one from each participating class, were asked to participate in an interview about their education methods when instructing LBC, and the difficulties teachers and LBC faced in regard to schooling.

Table 1. Descriptive statistics reported in means and percentages

	Both-parent migration (N=547)	One-parent migration (N=561)	No-parent migration (N=600)	$F/\chi 2$
Demographic information				
Age	15.07 ± 1.94	15.05 ± 1.95	14.96 ± 1.92	0.57
Gender				
Males	55.20%	56.70%	52.70%	
Females	44.80%	43.30%	47.30%	
Only child				23.76***
Yes	23.10%	28.50%	36.20%	
No	76.90%	71.50%	63.80%	
Education level				
Junior high school	44.10%	43.30%	39.70%	
High schoo	55.90%	56.70%	60.30%	
Left-behind characteristics				
Length of parental migration[a]				21.54***
Short-term	23.50%	36.30%	—	
Medium-term	36.60%	29.60%	—	
Long-term	39.90%	34.10%	—	
Frequency of contact[b]				8.37*
Rarely	8.60%	12.50%	—	
Sometimes	36.80%	41.00%	—	
Often	54.50%	46.50%	—	
Primary caregiver				935.42***
Non-migrating parent[c]	—	91.10%	—	
Grandparents	68.70%	—	—	
Others[d]	31.30%	8.90%	—	

Note: *p<0.05; **p<0.01; ***p<0.001.

[a] Short-term means the length of parental migration is less than 1 year, medium-term means the length of parental migration is between 1 to 5 years. Long-term means the length of parental migration is more than 5 years.

[b] Frequency of contact is measured by subjective assessment of their communication with their parents.

[c] Most of non-migrating parents are mothers (more than 90.0%).

[d] Others include four categories: cared for by relatives, cared for by LBC's sibling, cared for by themselves, and cared for by others.

4 Results

4.1 Demographics

Comparing LBC's demographics with non–LBC's demographics, there were no significant differences on the distribution of age, gender, and education level. Table 1 displays the sample's descriptive statistics on selected variables. In this sample (N=1,708), 61.79% of participants reported that at least one parent migrated to an urban area for a job. More than 60% of children in this sample reported being from a one–only–child family. This phenomenon was especially common for LBC's families.

In regard to migrant status, there were significant differences between children of both–parent migration and children of one–parent migration on the following variables: Length of time since parental migration, frequency of contact, and identity of primary caregiver. Children of both–parent migration experienced a longer length of time since parental migration than children of one–parent migration. Among LBC, adolescents generally reported having contact with their migrating parent(s). Most LBC reported they 'often' had contact with their parents. In particular, children of both–parent migration reported often having contact with their parents. The majority of both–parent migration children were raised by their grandparents. Most children of one–parent migration were raised by the remaining parent who did not migrate.

4.2 Psychological development of LBC and their counterparts

There were significant correlations between the majority of variables. In terms of adolescents' loneliness, depression, life satisfaction, self–esteem, social preference,father–child relationship, mother–child relationship, problem behaviors, and school engagement, multivariate analysis of variance (MANOVA) was performed to compare group differences based on (a) migrant status (both–parent

migration, one–parent migration, no–parent migration); (b) education level (junior high school,high school); and (c) gender. Findings indicated significant group differences based on migrant status, education level and gender, but no interaction among the variables was found (see Table 2).

Further ANOVA revealed that, based on LBC's migrant status, significant group differences were noted in participants' reports of loneliness, depression,life satisfaction, self–esteem, and school engagement. Based on LBC's education level (junior high versus senior high), differences were noted in social preference,father–child relationship, mother–child relationship, and problem behavior. Based on LBC's gender, significant differences were found in life satisfaction, self–esteem,social preference, father–child relationship, problem behavior, and school engagement (see Table 3).

Results of post–hoc tests indicated that children with both parents who migrated had lower scores on life satisfaction ($p<0.001$) and self–esteem ($p=0.008$) than non–LBC. Similarly, children with one parent who migrated had lower scores on life satisfaction than non–LBC ($p=0.009$). Children with both parents who migrated reported higher levels of depression than non–LBC ($p=0.006$). However, LBC with one parent who migrated ($p=0.001$) or both parents who migrated ($p=0.003$) engaged in more school activities than non–LBC.

Post hoc tests indicated age differences in children's perceptions. Overall,students in high school reported higher scores on social preference ($p<0.001$),father–child cohesion ($p=0.006$), mother–child cohesion ($p=0.018$), and problem behaviors ($p=0.019$) than students in junior high school.

Table 2. MANOVA of all children on migrant status, education level, and gender (N=1,708)

Independent variable	Wilk's Λ	F	η_p^2
Migrant status	0.97	2.71***	0.014
Education level	0.97	5.30***	0.027
Gender	0.92	16.91***	0.082

Note. *$p<0.05$ **$p<0.01$ ***$p<0.001$.

Table 3. ANOVA of variables of all children on migrant status, education level, and gender(*N*=1,708)

Independent variable	Dependent variable	F	η_p^2
Migrant status	Loneliness	4.03*	0.005
	Depression	3.95*	0.005
	Life Satisfaction	8.72***	0.010
	Self–esteem	3.51*	0.004
	School Engagement	5.32**	0.006
Education level	Social Preference	13.62***	0.008
	Father–child Relationship	7.52**	0.004
	Mother–child Relationship	5.59*	0.003
	Problem Behavior	5.49*	0.003
Gender	Life Satisfaction	5.81**	0.016
	Self–esteem	7.61**	0.004
	Social Preference	5.17*	0.003
	Father–child Relationship	8.24**	0.005
	Problem Behavior	89.45***	0.050
	School Engagement	9.35**	0.005

Note. *p<0.05, **p<0.01, ***p<0.001.

Post hoc tests revealed gender differences. Boys reported higher self–esteem(p=0.006), stronger father–child cohesion (p=0.004), and more problem behaviors(p<0.001). Girls reported higher levels of life satisfaction (p=0.016), social pref–erence (p0.006), and school engagement (p=0.002).

Comparisons among LBC subgroups. In order to more thoroughly investigate the effects of migration status on LBC, multivariate analysis of covariance (MANCVOA) was employed simultaneously to assess group differences in all dependent variables,considering caregiving type, length of parental migration, and frequency of parental contact while controlling two key demographic characteristics, school and gender. No interaction effect was found between independent variables and covariates.

In regard to LBC's caregiving type, more than 85% of LBC were cared for either by their grandparents or their non–migrating parent. The remaining 15% of LBC children were categorized into four caregiving types, with the subgroups' sizes ranging from 4.1% to 9.5% of the total LBC group. Thus in this study, we merged these four subgroups into a new category ('other') and this subset was not included in the remaining analyses. Further, after testing for differences between the non–analysed category ('other') and the analysed category (those chidren brought up by a single parent or grandparents) on all dependent variables, no differences were found.

In addition, there were minimal missing data (<5%) on the two variables–length of parental migration and frequency of parental contact. When comparing the two groups, those children with missing data and those children without missing, data were homogeneous on all dependent variables. Therefore,those LBC with missing data were included in the data analyses. Important to consider when interpreting Table 4, the actual number of LBC included in this particular data analyses was 861.

The findings suggest that the main effects by caregiving type, length of parental migration, and frequency of parental contact were significant, as well the interaction between the length of parental migration and the caregiving type (Table 4). Further analyses revealed significant differences between one–parent caregiving and grandparent–caregiving when considering LBC's self–esteem, F(1, 851)=4.00,p=0.046, η_p^2=0.005; and mother–child cohesion, F (1, 851)=15.58, p<0.001, η_p^2= 0.018.

Table 4. MANCOVA of differences among left behind children (LBC) based on differences in migrant Status.

Independent variable	Wilk's Λ	F	η^2_p
Caregiving type	0.97	2.45**	0.026
Length of parental migration	0.96	1.84*	0.019
Frequency of parental contact	0.87	6.64***	0.066
Duration × Caregiving type	0.96	1.87**	0.020
Covariates			
Education level	0.97	3.36***	0.035
Gender	0.90	9.93**	0.096

Note. *p<0.05; **p<0.01; ***p<0.001.

N=861

Continuing with the results listed in Table 4, when considering the length of parental migration, significant group differences were noted on life satisfaction, $F(2, 851)=5.19$, $p=0.006$, $\eta^2_p=0.012$; and social preference, $F(2, 851)=3.32$, $p=0.037$, $\eta^2_p=0.008$.

When considering the amount of parental contact, there were significant group differences reported on the following variables: Loneliness, $F(2, 851)=12.69$, $p<0.001$, $\eta^2_p=0.029$; life satisfaction, $F(2, 851)=28.45$, $p<0.001$, $\eta^2_p=0.063$; self-esteem, $F(2, 851)=14.35$, $p<0.001$, $\eta^2_p=0.033$; depression, $F(2, 851)=30.67$, $p<0.001$, $\eta^2_p=0.067$; father-child cohesion, $F(2, 851)=21.87$, $p<0.001$, $\eta^2_p=0.049$; mother-child cohesion, $F(2, 851)=16.41$, $p<0.001$, $\eta^2_p=0.037$; school engagement, $F(2, 851)=4.00$, $p=0.019$, $\eta^2_p=0.009$. Additionally, an interaction between the length of parental migration and caregiving type was found on school engagement, $F(2, 851)=4.12$, $p=0.016$, $\eta^2_p=0.010$.

Post-hoc test comparisons revealed that LBC who were brought up by the nonmigrating parent reported higher scores on self-esteem ($p=0.046$) and mother-child relationship ($p<0.001$) than those who were brought up by grandparents. Long-term LBC reported lower life satisfaction scores than medium-term LBC($p=0.001$). However, Owever, on social preference, medium-term LBC reported lower scores than short-term LBC ($p=0.016$).

In regard to the contact frequency, results showed that LBC who often had contact with their migrant parent(s) felt less lonely and depressed than LBC who sometimes ($p<0.001$) and rarely ($p<0.001$) had contact with migrant parent(s). LBC who sometimes had contact with their migrant parent(s) were less depressed than LBC who rarely had contact ($p<0.001$).

On life satisfaction, self-esteem, father-child cohesion and mother-child cohesion, LBC who often had contact with their migrated parents fared better than those LBC who sometimes ($p<0.001$) and rarely ($p<0.001$) had contact. On father-child cohesion ($p=0.029$) and mother-child cohesion ($p=0.049$), LBCwho sometimes had contact with their parents fared better than LBC who rarely had contact. In comparison to LBC who rarely had contact with their migrant parents, school engagement was significantly better for children who often had contact with their migrant parent ($p=0.005$) and for LBC who sometimes had contact ($p=0.046$). Simple effect analysis for the interactive effects revealed that long-term LBC performed better on school engagement than those medium-term LBC ($p=0.035$) and short-term LBC ($p=0.004$), particularly if they were brought up by their grandparents.

4.3 LBC´s perceptions of life

Based on interviews, on the whole, LBC reported holding negative perceptions of life changes after parental migration. Most of them (69%) perceived that their current life was different from life prior to their parents' departure. The main reported change was the harmful effect on the parent-children relationship. For example, care from and communication with migrated parent(s) decreased. Estrangement from migrated parent(s) was either initiated or exacerbated following their departure. On the contrary, after the parents migrated a few LBC reported increased communication with parent(s).

During their interviews, several LBC indicated that they become tougher and more sensible and thoughtful after their parent(s) migrated. However, whether LBC perceived parental migration as positive or negative, the vast majority

reported prominent feelings of missing parents and being lonely. LBC commonly reported not wanting their parents to migrate for work and also reported feeling greater happiness when living together with their parents.

In regard to perceptions of their current situation, 56% of LBC indicated that learning was the most important aspect of their current life, followed by family relationships (28%) and friendships with peers (16%). However these aspects of life were both the source of their support and also the source of their dissatisfaction. When asked if LBC preferred to stay at school or at home, half preferred to stay at home and half preferred to stay at school. Each environment had particular benefits and drawbacks. At home, they felt freer, they felt emotionally supported, and they did not feel academic pressure. At school, with their fellow classmates and teachers, LBC reported feeling less lonely. Most LBC reported being dissatisfied with their current life because of interpersonal relationship problems, parent absence, and academic difficulties.

In the interviews almost all of LBC reported having at least one friend. LBC stated that friends made them feel happy. More specifically, friends helped each other with daily life issues and with learning activities; friends shared feelings; friends accompanied and encouraged each other when experiencing difficulties.

From LBC's point of view, teachers were very kind and supportive. Almost all LBC realized the important role teachers played in regard to students' academic guidance. Almost one-third of LBC said that their teachers not only took care of students' daily academic needs, but also taught them how to be 'human'.

4.4 Teachers´ perceptions of LBC´s education

According to information gathered from 32 teachers' interviews, schools have engaged in action to help LBC. Almost half of teachers (41%) reported that each term their schools archive records for LBC. Nearly one-third of teachers reported feeling an obligation to take responsibility for helping LBC in class and, beyond school, making more home visits for LBC. A few teachers (16%) reported that their schools provided a living subsidy for LBC. Additionally, one-fourth of teachers indicated that specific campus activities were developed for LBC, such as themed activities, safety and security education, and psychological counseling. Teachers also indicated that schools provided phone cards for LBC to promote LBC-parent communication.

In regard to how teachers educate LBC, almost half of teachers (44%) reported keeping close contact with students' parents (including grandparents). Teachers indicated that they 'communicate with left-behind students more' and take more responsibility for 'LBC's extracurricular life'. About one-third of teachers indicated that they paid extra attention to LBC. However, in opposition to the majority of teachers' input, one teacher emphasized that all students should be treated equally without discrimination or favoritism.

For most teachers, the greatest difficulty in the process of educating LBC is communicating with LBC's parents. Another major concern for teachers includes expressing concerns about LBC's challenging personal characteristics, such as being intractable and having a mistrustful attitude towards the teacher. Teachers expressed feeling powerless to address LBC's challenging attitudes and behaviors. Moreover, a few teachers pointed out the difficulty posed by the insufficient integration of and cooperation between the LBC's school education and family life.

Thus, based on these educational challenges, teachers also offered some advice. Nearly one-in-four teachers suggested that children should not be left behind, or minimally one parent should stay at home to take care of the children. A similar number of teachers reported the need for teachers to offer LBC more supportive care and encouraging praise. Teachers also stressed the importance of increasing communication between parents and

teachers, as well as between parents and children.

Additionally, approximately one-in-five teachers reflected on the healthy development of LBC and the necessary ingredients to support this health development. Teachers cited the need for a combination of LBC's family support, adequate school education, and support from community/government welfare.

5 Discussion

In rural China, parents frequently migrate to urban areas in search of employment. This move often necessitates leaving their children behind. This study conducted in two junior high schools and two senior high schools found significant developmental differences between rural left-behind children (LBC) and non-LBC. LBC experienced less life satisfaction than non-LBC. LBC with both-parent migration report being more depressed and having a lower self-esteem than non-LBC. These findings indicated that rural parents' migration had an overall negative impact on LBC's mental health, particularly when both parents migrated. Findings based on this study's quantitative data (surveys) and qualitative data (interviews) are consist ent with previous studies (Jia & Tian, 2010; Sun, Zhou, Wang, & Fan, 2010; Zhou et al., 2005).

However, in contradiction with the previous findings of Wen and Lin's (2012) study, in the current study LBC reported a higher level of school engagement and adaptation. LBC reported enjoying school activities, doing their homework, and following school rules. This finding was somewhat unexpected. Possibly in this study LBC were more disadvantaged and underwent more stressful life events than non-LBC (Liu & Wang, 2010), and to counter these challenges, LBC may have had more intense motivation to do well in school, to earn good marks, and to become more involved in school activities. Additionally,

these findings may be related to the participating schools' quality of care provided by classroom teachers and the quality of peer relationships unique to the participating schools. These supportive qualities may not be similar across China's rural schools and communities. Therefore, data collected in different schools and communities may not align with the findings of this study.

5.1 Age and gender

The role of two variables—education level (age) and gender—was noteworthy, but an interaction between these two variables and migrant status was not found. However, in comparison to younger LBC (junior high school students), older LBC (high school students) reported better peer relationships and parent-child relationships. On the other hand, problem behaviors of high school students were significantly greater in comparison to younger LBC. This may be related to older LBC's enhanced self-awareness and increased academic frustration (Greene, Krcmar, Walters, Rubin, & Hale, 2000; Siegel & Scovill, 2000). Additionally, older students' school boarding life further weakened parental monitoring and also gave adolescents certain freedoms to independently navigate the cost of living. Moreover, with increasing age and awareness, adolescents naturally become more fully engaged and involved in the complexities of society. This independence and increased freedom might also underlie older LBC's escalating behavior problems (Shaffer & Kipp, 2010).

Gender differences were found in LBC's problem behavior and social relations. In comparison to girls, boys exhibited more problem behaviors, lower levels of school engagement, and worse peer relationships. On the other hand, boys reported better relationships with their fathers and higher levels if self-esteem. Overall, these gender differences were consistent with previous studies (Fan, Fang, Liu, & Liu, 2009; Fan, Su, Gill, & Birmaher, 2010; Sun et al., 2010; Wen & Lin, 2012).

This study's findings were also related to parenting

style and social expectations in Chinese culture (Shaffer & Kipp, 2010). Therefore, educators should take into consideration the different developmental stages of youth in educational practice and the type of challenges which are specific to the gender of LBC.

5.2 Migrant status and caregivers

This study also found that different migrant conditions might cause different effects in LBC. Overall, LBC who were cared for by one parent fared better than those who were cared for by grandparent(s). LBC brought up by one parent had a more intimate relationship with their mother and reported higher levels of self-esteem than children brought up by grandparent(s). As previously indicated, more than 90% of LBC who were cared for by the remaining non-migrating parent were brought up by their mother. Additionally, previous research conducted by Zhao, Shen, and Liu (2008) found that mother's support had a positive influence on LBC's self-esteem.

LBC who were cared for by their mothers reported receiving more of their mother's support. The mother's care of LBC was positively related to the individual child's self-esteem and also with a stronger mother-child relationship. Therefore, in the process of raising a child the important role of parent (especially mother) is not easily replicated by a substitute caregiver, such as a grandparent.

5.3 Duration of separation and parental contact

Medium-term LBC had the highest score on life satisfaction and peer acceptance (social preference). Possibly, in comparison to short-term and long-term LBC, medium-term LBC may better cope with and adapt to separation from their parents. In addition, there was an interesting interaction between length of parental migration and caregiving type. For adolescents who were brought up by a grandparent, LBC who had a longer separation from the migrated parent actually reported better school engagement.

Generally speaking, the long-term LBC with

both-parent migration (and who were brought up by grandparents) struggled with the most disadvantages and often appeared to show the weakest adaptation to healthy development. However, in this study, LBC who experienced long-term separation from their parents and who were brought up by a grandparent presented the highest school engagement. This might reflect a positive coping style of these LBC who seemingly demonstrate greater resilience in a school environment.

Among the left-behind characteristics investigated in this study, the frequency of parental contact had the broadest impact on LBC's adaption. Parental contact was beneficial to LBC's mental health. Children who had the most frequent contact with their parents suffered less with loneliness and depression and reported the highest life satisfaction and self-esteem. In regard to social relationships, father-child cohesion and mother-child cohesion were strongest in the group of LBC who had the most frequent contact with their parents. However, based on frequency of contact with LBC's parents, in regard to peer relationships, there was no significant difference between groups.

Sufficient contact with parents may also improve LBC's school engagement. Adolescents who rarely had contact with their migrated parents performed the worst on school engagement. These results were consistent with previous research (Su, Li, Lin, Xu, & Zhu, 2013). LBC's communication with parents and children's level of disclosure with parents appears to facilitate LBC receiving social support, which enhances healthy emotional adaptation (Chaudoir & Fisher, 2010). Su et al. (2013) also found that high levels of parental communication is one of the strongest protective factors of satisfaction with school and life in general, ultimately related to children's overall happiness and well-being. Thus, frequent contact with parents helps promote LBC's positive mental health and social adaptation.

In the children's interviews, based on the LBC's

perception of their own lives most children reported becoming alienated from their parents following their parents' migration for work. This alienation appears to have the greatest effect on children impacted by parental migration, and subsequently a direct cause of children's loneliness in their parents' absence. According to the 'ecological model of rural LBC's psychological development,' suggested by Zhao and Shen (2010), the proximal factor (such as single parent's migration or both parents' migration) and the distal cause/outcome (such as parent–children relationship) in left–behind environment were differentiated. The distal factors may, through certain proximal factors, contribute to LBC's development. As a result, in the condition where parents migrate for work and leave their children behind, interventions must focus on improving the LBC–parent relationship, helping enhance parent–child communication and building family cohesion.

5.4 School engagement and relationships

Ultimately, educational intervention efforts must focus on encouraging and promoting LBC to actively adapt to their challenging situation. Primarily, LBC's dissatisfaction in life was associated with academic pressure, problems in family relationships, and difficulties in peer relationships. Therefore, school educators and mental health workers should consider these three aspects as they identify school–based strategies to improve LBC's life satisfaction and academic progress. Interestingly, this research study's data indicate that half of LBC like staying at home, while the other half prefer staying at school where they can associate with their peers.

Home and school are the two most important places where children and adolescents develop relationships and personal skills related to life satisfaction. Indeed, peers play an important role in children's socialization. Even when bullied, children's loneliness can be alleviated by associating with a supportive friend (Sun et al., 2010).

Adolescents who had no 'best friend' reported feeling greater loneliness than those who had a 'best friend' (Zou, 1998). Research indicates that the quality of adolescents' friendship is highly correlated with increased self–esteem and sense of self–value, lower levels of depression and anxiety, and social adjustment (Berndt & Keefe, 1995; Hartup, 1996). Friends help care for LBC on a daily basis and accompany them during school and outside of school. Additionally, friends are available to assist LBC with school assignments. Many LBC reported being happy when they were together with friends. This study's findings were in line with previous research which found that good peer relationships help alleviate children's loneliness (Sun et al., 2010; Zhou, Zhao, Chen, Jiang, & Hundley, 2003). The compensation of peer relationship is very important, especially because these relationships act as a protective factor for LBC's adaptation to the school environment.

5.5 School support of LBC

All LBC expressed having a positive attitude towards their teachers. They considered their teachers to play an important role in guiding and improving student learning, providing a daily lift, and helping students develop good character. Above all, it is of critical importance for adolescents to develop emotional intimacy and warmth that ensures good communication between peers and with caring adults. Peers and teachers provide companionship for LBC and serve as guides in day–to–day school life. Based on this study's data, school is an important place for LBC's social development.

According to the teacher interviews, teachers' frontline role in education allowed them to understand the difficulties in educating LBC. The greatest challenge was in bridging the gap between the school's role in education and the family's role in education. For example, teachers reported having difficulty communicating with LBC's parents and perceived that LBC did not have sufficient parent support. Thus, teachers stressed the importance of increasing the communication between parents, teachers,

and children. In this way, schools' and parents' combined efforts would better support LBC's educational and social—emotional needs, creating a school environment that nurtured LBC's learning and healthy development.

5.6 Limitations of current research

Considering that Cronbach's α was low for the measurement of School Engagement (Cronbach's α =0.46), we should explain whether it affects our confidence in interpreting data from this measurement. First of all, Cronbach's α is a special case of intra—class correlation coefficient (ICC), comparable to a two—way random model ICC (McGraw & Wong, 1996; Weir, 2005). Thus, when the correlation of items' correlation was low, would be weakened. School Engagement was a variable that measured the degree of students' school activity engagement (e.g. 'Do you like being obedient to school discipline?'). These events often had high incidence rate, so the distribution of these data was skewed, which weakened the item correlation. Second, although these items all belong to one category, each presented item may occur independently from the other items, therefore this traditional estimate of internal consistency may not have been an appropriate measure to judge this measure's reliability and whether these items measured what they purported to measure (Kim, Conger, Elder, & Lorenz, 2003). Previous researchers have also addressed similar concerns with using Cronbach's α (Kim et al., 2003; Straus & Kantor, 2005). Third, Cronbach's α typically increases with an increased number of items. In our research, there were only five items in this scale, thus limiting the scale's α.

Last but not least, Cronbach's α has been criticized as underestimating true reliability (Peterson & Kim, 2013). In fact, Turner and Wheaton (1995) declared that a measure may be valid even if α is zero. Thus we have reason to believe that the school engagement measure's low reliability is not an indication of the measure's validity.

However, we urge caution in drawing conclusions from this aspect of our data.

5.7 Implications for improving prevention and intervention efforts

In line with previous research, our findings support the need for assistance and intervention systems for LBC. Educators and parents must play an important role in providing interventions to address LBC's educational and social—emotional needs. The following seven interventions are recommended and further discussed in the following sections: (a) improve school leadership and management; (b) improve education on the school and classroom level; (c) strengthen school mental health education services; (d) build stronger coalitions between family and school; (e) provide a supportive and caring foundation; (f) support current and future education and (g) increase outside support of LBC. In order to bring about positive change, these recommended interventions should be integrated and complementary.

Improve school leadership and management. In view of the high proportion of left—behind students in rural schools, school leadership must consider and provide accommodations for LBC and strengthen the school's management to ensure the daily care and personal safety of LBC. In addition, according to the needs of some migrated family with good income, the school might consider charging a suitable fee to improve the school's boarding dietary standard and accommodations for children. Moreover, if it is possible, arrange life coaching teachers and teaching tutors to assist the management and education of LBC. In the long run, school administrators should pay close attention to the quality of education, try to improve teacher quality, and provide training opportunities (such as psychological counseling training). Strong academic courses, caring teachers, and a supportive school environment help LBC enjoy school and adapt to their challenging family circumstance.

Improve education on the school and classroom level. First, optimize the educational content. Increase the content such as agriculture, science, technology and other practical technology to help children grasp survival skills prior to graduating from junior high school. Second, adjust the course structure. Strengthen the LBC's survival, safety, and legal education, and improve their self−respect, self−reliance, discipline consciousness and legal knowledge. Third, enrich the extracurricular life of students, especially for the LBC. Entertainment activities increase peer interactions, and offer opportunities to strengthen shape healthy personality. Fourth, mobilize teachers, youth−league members, young pioneers and other students to help LBC by pairing LBC with supportive students to ease stress and encourage psychological adjustment.

Strengthen school mental health education services. As mental health and psychological well−being of LBC are often neglected (Luo, Wang, & Gao, 2009), it is necessary to publicly promote psychoeducational activities and to regularly provide psychological consultation activities. These efforts will be helpful in preventing potential psychological crises in disadvantaged LBC. It is important to establish a psychological counseling room and to encourage students to accept psychological counseling services which will help provide additional social support and enhance LBC's emotional adjustment. For children with serious psychological problems, psychological teachers should conduct tracking observations and offer supportive services to address challenging issues, track progress, and then summarize the effect of the relevant intervention measures.

Build stronger coalitions between family and school. Family supports (both from parents and the primary caregiver) are very important for LBC. Programs should be well designed to strengthen family ties before and following migration. We suggest offering supportive training for parents. Parent−teacher and parent−child communication is critical. Parents should strengthen the exchange and communication with their children and keep in touch with their children's temporary caregiver. Likewise, in a coordinated effort parents and caregivers must regularly communicate with LBC's teachers.

Provide a supportive and caring foundation. Remind parents, caregivers, and teachers to frequently offer LBC more supportive sentiment. Demonstrate concern for LBC's well−being and provide character education to further strengthen social relationships and social skills.

Support current and future education. Even though migrating parents are not present to give learning guidance during their absence, parents must always encourage their children's studying and build children's confidence in their ability to learn and study. Parents should avoid extreme reactions to their children's educational progress— negligence and limited expectations for the children's academic achievement or, on the other hand, exerting too much pressure for children to achieve unrealistic academic goals. Head teachers may take more responsibility for LBC's school involvement by providing special activities, establishing routine home visits, and holding teacher− parent meetings when parents are home during the holidays.

Increase outside support of LBC. Finally, it is worth noting that interventions organized and funded outside of schools are also very important in supporting LBC. Concrete actions must be taken with support from government organizations, community and private enterprises, universities, international communities, and so forth. For example, to compensate for the lack of national fiscal capital investment in rural schools, local governments may try to change the investment system of funding rural education, possibly running a school with subsidized funding from the community and government. One option, public welfare organizations can organize well−educated caring volunteers to set up a Chinese

Pioneer Union, providing strong academic courses and offering supervised activities for LBC, even extending beyond the typical school hours. Additional options include offering parent hotlines or a website with family counseling services. In order to improve resource utilization, these services can be set up in targeted local areas or expanded nationwide and offered to individuals from all walks of life. Additionally, with available internet and technology, LBC and migrated parent(s) should be able to communicate more frequently, countering LBC's feelings of loneliness.

6 Conclusion

Parental migration in China's rural areas appears to be a common situation for many adolescents (Duan & Zhou, 2005; Zhou et al., 2005). Additionally, the negative impact of parent migration on LBC's psychological development and social adaptation is extensive. Understanding the common challenges facing LBC, educators and parents must strategically identify and implement interventions to address these needs, fostering better outcomes for LBC.

References

Aguilera–Guzman, R. M., de Snyder, V. N., Romero, M., & Medina–Mora, M. E. (2004). Paternal absence and international migration: Stressors and compensators associated with the mental health of Mexican teenagers of rural origin. *Adolescence, 39*(156), 711–723.

Amato, P., & Keith, B. (1991). Parental divorce and the well–being of children: A metaanalysis. *Psychological Bulletin, 110*(1), 26–46. doi: 10.1037/0033–2909.110.1.26.

Asher, S. R., & Wheeler, V. A. (1985). Children's loneliness: A comparison of rejected and neglected peer status. *Journal of Consulting and Clinical Psychology, 53*, 500–505. doi: 10.1037/0022–006X.53.4.500.

Attili, G., Vermigli, P., & Roazzi, A. (2010). Children's social competence, peer status, and the quality of mother–child and father–child relationships. *European Psychologist, 15*(1), 23–33. doi: 10.1027/1016–9040/a000002.

Bagner, D. M., Storch, E. A., & Roberti, J. W. (2004). A factor analytic study of the Loneliness and Social Dissatisfaction Scale in a sample of African American and Hispanic American children. *Child Psychiatry and Human Development, 34*, 237–250. doi: 10.1023/B:CHUD.0000014999.16111.2f.

Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high selfesteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest, 4*, 1–44. doi: 10.1111/1529–1006.01431.

Beck, K. H., Boyle, J. R., & Boekeloo, B. O. (2003). Parental monitoring and adolescent alcohol risk in a clinic population. *American Journal of Health Behavior, 27*(2), 108–115. doi: 10.5993/AJHB.27.2.2.

Berndt, T. J., & Keefe, K. (1995). Friends' influence on adolescents' adjustment to school. *Child Development, 66*(5), 1312–1329. doi: 10.1111/j.1467–8624.1995.tb00937.x.

Bryant, J. (2005). *Children of international migrants in Indonesia, Thailand and the Philippines: A review of evidence and polices (Innocenti Working Paper No. 2005–05)*. Florence: UNICEF Innocenti Research Centre.

Carandang, M. L. A., Sison, B. A., & Carandang, C. (2007). *Nawala ang ilaw ng tahanan: Case studies of families left behind by OFW Mothers*. Pasig City: Anvil Publishing.

Chaudoir, S. R., & Fisher, J. D. (2010). The disclosure processes model: Understanding disclosure decision making and postdisclosure outcomes among people

living with a concealable stigmatized identity. *Psychological Bulletin, 136*(2), 236－256. doi: 10.1037/ a0018193.

Chen, X., & Xie, Y. (2007). An investigation of rural left-behind children's problematic behaviors and the family factors. *Journal of Inner Mongolia Normal University (Philosophy and Social Science), 36*(6), 29－33.

Cronk, N. J., Slutske, W. S., Madden, P. A. F., Bucholz, K. K., & Heath, A. C. (2004). Risk for separation anxiety disorder among girls: Paternal absence, socioeconomic disadvantage, and genetic vulnerability. *Journal of Abnormal Psychology, 112*(2), 237－247. doi:10.1037/0021－843X.113.2.237.

Dawei, Y. (2000). Preliminary use of the children s depression inventory in China. *Chinese Mental Health Journal, 14*(4), 225－227.

DiClemente, R. J., Wingood, G. M., Crosby, R., Sionean, C., Cobb, B. K., & Harrington, K., et al. (2001). Parental monitoring: Association with adolescents' risk behaviors. *Pediatrics, 107*(6), 1363－1368. doi: 10.1542/peds.107.6.1363.

Dong, Q., & Lin, C. D. (2011). *The key index and assessment for mental development of 6－15 years old children and adolescents in China*. Beijing: Science Press.

Duan, C., Lu, L., Guo, J., & Wang, Z. P. (2013). Survival and development of left-behind children in rural China: Based on the analysis of sixth census data. *Population Journal, 35*(3), 37－49.

Duan, C. R., & Zhou, F. L. (2005). A study on children left behind. *Population Journal, 29*(1), 29－36.

Edillon, R. G. (2008). *The effects of parent´s migration on the rights of children left behind*. Quezon City: Asia Pacific Policy Center, UNICEF.

Fan, F., Su, L., Gill, M. K., & Birmaher, B. (2010). Emotional and behavioral problems of Chinese left-behind children: A preliminary study. *Social Psychiatry and Psychiatric Epidemiology, 45*(6), 655－664. doi: 10.1007/s00127－009－0107－4.

Fan, X. H., Fang, X. Y., Liu, Q. X., & Liu, Y. (2009). A social adaptation comparison of migrant children, rear children, and ordinary children. *Journal of Beijing Normal University (Social Sciences), 5*, 33－40.

Fang, X. Y., Li, X. M., & Dong, Q. (1996). The research of adolescent smoking and its relevant factors. *Chinese Mental Health Journal, 10*(2), 77－80.

Fang, X. Y., Zhen, Y., & Lin, D. H. (2001). Relationships between family factors and smoking behavior of junior middle school students. *Acta Psychologica Sinica, 33*(3), 244－250.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004, Spring). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59－109. doi: 10.3102/00346543074001059.

Gao, Y., Li, L. P., Kim, J. H., Congdon, N., Lau, J., & Griffiths, S. (2010). The impact of parental migration on health status and health behaviors among left behind adolescent school children in China. *BMC Public Health, 10*, 56－65. doi: 10.1186/1471－2458－10－56.

Greene, K., Krcmar, M., Walters, L. H., Rubin, D. L., & Hale, J. L. (2000). Targeting adolescent risk-taking behaviors: The contributions of egocentrism and sensationseeking. *Journal of Adolescence, 23*(4), 439－461. doi: 10.1006/jado.2000.0330.

Hartup, W. W. (1996). The company they keep: Friendships and their developmental significance. *Child Development, 67*(1), 1－13. doi: 10.1111/ j.1467－8624.1996.tb01714.x.

Huebner, E. S. (1991). Initial development of the student's life satisfaction scale. *School Psychology International, 12*, 231－240. doi: 10.1177/0143034391123010.

Huebner, E. S., Suldo, S. M., & Valois, R. F. (2003, March 12－13). Psychometric properties of two

brief measures of children's life satisfaction: The Students' Life Satisfaction Scale (SLSS) and the Brief Multidimensional Students' Life Satisfaction Scale (BMSLSS) Presented at the Indicators of Positive Development Conference, Washington, DC.Retrieved from http://www.childtrends.org/wp−content/uploads/2013/05/Child_Trends− 2003_03_12_PD_PDConfHSVP.pdf.

Hugo, G. (2002). Effects of international migration on the family in Indonesia. *Asian and Pacific Migration Journal, 11*(1), 13–46.

Jaggi, G., Rundle, M., Rosen, D., & Takahashi, Y. (1996). *China´s economic reforms: Chronology and statistics.* Working Paper 96–5. Washington, DC: The Peterson Institute for International Economics Retrieved from http://www.iie.com/publications/wp/96−5.pdf.

Jia, Z., & Tian, W. (2010). Loneliness of left−behind children: A cross−sectional survey in a sample of rural China. Child: Care, Health and Development, *36*(6), 812–817. doi: 10.1111/j.1365−2214. 2010.01110.x.

Joh, J. Y., Kim, S., Park, J. L., & Kim, Y. P. (2013). Relationship between family adaptability, cohesion and adolescent problem behaviors: Curvilinearity of circumplex model. *Korean Journal of Family Medicine, 34*(3), 169–177. doi: 10.4082/kjfm.2013.34.3.169.

Kim, K. J., Conger, R. D., Elder, G. H., & Lorenz, F. O. (2003). Reciprocal influences between stressful life events and adolescent internalizing and externalizing problem. *Child Development, 74*, 127–143. doi: 10.1111/1467−8624.00525.

Kovacs, M. (1992). Children's depression inventory manual. North Tonawanda, NY: MultiHealth Systems.Lin, D., Fan, X., Li, X., & Pan, J. (2010). The relationship between environmental and individual factors and adolescent drinking behavior. *Psychological Development and Education, 26*(3), 288–293.

Liu, B. Q., & Wang, W. (2010). Study on living stress events and psychological health of children remaining in rural areas. *China Journal of Health Psychology, 18*(2), 210–212.

Luo, J., Wang, W., & Gao, W. B. (2009). Review of the studies on rural left−behind children in China. *Advances in Psychological Science, 17*(5), 990–995.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. doi: 10.1037//1082−989X.1.1.30.

McLanahan, S. S., & Sandefur, G. (1994). *Growing up with a single parent: What hurts, what helps.* Cambridge, MA: Harvard University Press.

Melgar, G. A., & Borromeo, R. (2002). The plight of children of OFWs. In E. An onuevo, & A. An onuevo (Eds), *Coming home: Women, migration and reintegration* (pp. 106–114). Quezon City: Balikbayani Foundation and Atikha Overseas Workers & Communities Initiatives.

Morrison, W. M. (2014, October 9). China's economic rise: History, trends, challenges, and implications of the United States. *CRS report prepared for the members and committees of Congress.* Washington, DC: Congressional Research Service. Retrieved from http://fas.org/sgp/crs/row/RL33534.pdf.

Olson, D. H., Sprenkle, D. H., & Russell, C. S. (1979). Circumplex model of marital and family systems: Cohesion and adaptability dimensions, family types, and clinical applications. *Family Process, 18*(1), 3–28. doi: 10.1111/j.1545−5300.1979.00003.x.

Peterson, R. A., & Kim, Y. (2013). On the relationship between coefficient alpha and composite reliability. *Journal of Applied Psychology, 98*(1), 194–198. doi: 10.1037/a0030767.

Qu, Z. Y., & Zou, H. (2009). Juvenile delinquency: The role of self−control, family environment and parental

monitoring. *Psychological Science, 32*(2), 360‒363.

Reyes, M. M. (2008). *Migration and Filipino children left behind: A literature review*. Quezon City: Miriam College/UNICEF Retrieved from http://www.unicef.org/ philippines/ Synthesis_StudyJuly12008.pdf

Rosenberg, M. (1965). S*ociety and the adolescent self‒image*. Princeton, NJ: Princeton University Press.

Rosenberg, M. (1979). *Conceiving the self*. New York, NY: Basic Books.

Rosenberg, M., Schooler, C., Schoenbach, C., & Rosenberg, F. (1995). Global self‒esteem and specific self‒esteem: Different concepts, different outcomes. *American Sociological Review, 60,* 141‒156. doi: doi.org/10.2307/2096350.

Schmitt, D. P., & Allik, J. (2005). Simultaneous administration of the Rosenberg SelfEsteem Scale in 53 nations: Exploring the universal and culture‒specific features of global self‒esteem. *Journal of Personality and Social Psychology, 89*(4), 623‒642. doi:10.1037/0022‒3514.89.4.623.

Shaffer, D. R., & Kipp, K. (2010). *Developmental psychology: Childhood and adolescence* (8th ed.). Belmont, WA: Wadsworth.

Shen, J. L. (2009). *Looking in depth of the mental world of disadvantaged children*. Beijing: Beijng Normal University Press.

Siegel, A. W., & Scovill, L. C. (2000). Problem behavior: The double symptom of adolescence. *Development and Psychopathology, 12*, 763‒793. doi: 10.1017/ S0954579400004119.

Straus, M. A., & Kantor, G. K. (2005). Definition and measurement of neglectful behavior: Some principles and guidelines. *Child Abuse and Neglect, 29,* 19‒29. doi: 10.1016/j.chiabu.2004.08.005.

Su, S., Li, X., Lin, D., Xu, X., & Zhu, M. (2013). Psychological adjustment among leftbehind children in rural China: The role of parental migration and parent‒child communication. Child: Care, *Health and Development, 39*(2), 162‒170. doi:10.1111/j. 1365‒2214.2012.01400.x.

Sun, X. J., Zhou, Z. K., Wang, Y., & Fan, C. Y. (2010). Loneliness of children left in rural areas and its relation to peer relationship. *Psychological Science, 33*(2), 337‒340.

Tarroja, M. C. H., & Fernando, K. C. (2013). Providing psychological services for children of overseas Filipino workers (OFWs): A challenge for school psychologists in the Philippines. *School Psychology International, 34*(2), 202‒212. doi: 10.1177/0143034312453399.

Tian, L. M. (2006). Shortcoming and merits of Chinese version of Rosenberg (1965) selfesteem scale. *Psychological Exploration, 26*(2), 88‒91.

Turner, J. R., & Wheaton, B. (1995). Checklist measurement of stressful life events. In L. Gordon (Ed.), *Measuring stress: A guide for health and social scientists* (pp. 29‒58). New York, NY: Oxford University Press.

Unnever, J. D., Cullen, F. T., & Travis, P. C., (2003). Parental management, ADHD, and delinquent involvement: Reassessing Gottfredson and Hirschi's general theory. *Justice Quarterly, 20*(3), 471‒500. doi: 10.1080/07418820300095591.

Wang, X. L., Hu, X. Y., Shen, J. L. (2011). Affection of left children's friendship quality on loneliness and depression. *Chinese Journal of Clinical Psychology, 19*(2), 252‒254.

Wei, X., & Chen, X. (2010). Prosocial tendencies and its relationship to personalities and family functioning among the left‒home and family functioning among the left‒home kids in junior middle school. *Psychological Development and Education, 4*, 402‒408.

Weir, J. P. (2005). Quantifying test‒retest reliability using the intraclass correlation coefficient and the SEM. *Journal of Strength and Conditioning Research,*

19(1), 231‒240.

Wen, M., & Lin, D. (2012). Child development in rural China: Children left behind by their migrant parents and children of nonmigrant families. *Child Development, 83*(1), 120‒136. doi: 10.1111/j.1467‒8624.2011.01698.x.

Xiang, B. (2007). How far are the left‒behind left behind? A preliminary study in rural China. *Population Space and Place, 13*, 179‒191. doi: 10.1002/psp.437.

Yeoh, B. S. A., & Lam, T. (2007). The costs of (im)mobility: Children left behind and children who migrate with a parent. In *Perspectives on gender and migration* (pp. 120‒149). In the United Nations Economic and Social Commission for Asia and the Pacific Regional Seminar on Strengthening the Capacity of National Machineries. Bangkok, Thailand: United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP).

Yu, D. W., & Li, X. (2000). Preliminary use of the children's depression inventory in China. *Chinese Mental Health Journal, 14*(4), 225‒227.

Zhang, W. W. (2004). *Transforming China: Economic reform and its political implications.* New York, NY: St. Martin's Press.

Zhao, J. X., Liu, X., & Zhang, W. X. (2013). Peer rejection, peer acceptance and psychological adjustment of left‒behind children: The roles of parental cohesion and children's cultural beliefs about adversity. *Acta Psychologica Sinica, 45*(7), 797‒810. doi: 10.3724/SP.J.1041.2013.00797.

Zhao, J. X., & Shen, J. L. (2010). The effect of pre‒study or post‒study emotional arousal on implicit and explicit memory. *Chinese Journal of Special Education, 7*, 65‒70, 76.

Zhao, J. X., Shen, J. L., & Liu, X. (2008). Left‒at‒home‒adolescents' perception of social support networks and their relevance to individual self‒esteem and social initiative: Variable‒centered and person‒centered perspectives. *Psychological Science, 31*(4), 827‒831.

Zhou, C., Wang, D. B., & Hong, Q. (2010). Influencing factors and related measures of sex abuse cases among left‒behind children. *Medicine and Society, 23*(1), 54‒56.

Zhou, Z. K., Sun, X. J., Liu, Y., & Zhou, D. M. (2005). Rural left‒behind children's psychological development and education problems. *Academic Journal of Beijing Normal University, 1*, 71‒78.

Zhou, Z. K., Zhao, D. M., Chen, J., Jiang, J. C., & Hundley, R. (2003). Loneliness as a function of sociometric status and self‒perceived social competence in middle childhood. *Psychological Development and Education, 19*(4), 70‒74.

Zou, H. (1998). Developmental function and influence factors of peer relationship. *Psychological Development and Education, 12*(2), 40‒46.

Childhood adversity, recent life stressors and suicidal behavior in Chinese college students

Zhiqi You[1, 2], Mingxi Chen[1, 2], Sen Yang[1, 2], Zongkui Zhou[1, 2,*] Ping Qin[3, 4]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan, China)

([2] School of Psychology, Central China Normal University, Wuhan, China)

([3] National Center for Register–based Research, University of Aarhus, Aarhus, Denmark)

([4] National Center for Suicide Research and Prevention, Institute of Clinical Medicine, University of Oslo, Oslo, Norway)

Abstract   Background: Although the independent effects of childhood adversities and of recent negative events on suicidality have been well–documented, the combinative role of childhood and recent adversities on risk for suicidality is still underexplored, especially in the context of Chinese culture and in consideration of specific types of negative events. Method: 5989 students, randomly sampled from six universities in central China, completed the online survey for this study. Suicidal behavior, life adversity during childhood and stressful events in recent school life were assessed with designed questionnaires. Results: Students experiencing recent stressful life events more often reported an experience of life adversity during childhood. While recent stressful life events and childhood life adversity both were associated with an increased risk for suicidal behavior, the two exposures presented conjunctively and acted interactively to increase the risk. There was noticeable variation of effects associated with specific childhood life adversities, and sexual abuse, poor parental relationship, divorce of parents and loss of a parent were among the adversities associated with the highest increased risk. Recent conflicts with classmates, poor school performance and rupture of romantic relationships were the recent school life stressors associated with the highest increased risk. Conclusions: Childhood adversity and recent school life stressors had a combinative role in predicting suicidality of young people studying in Chinese colleges. Unhappy family life during childhood and recent interpersonal conflicts in school were the most important predictors of suicidality in this population.

Keywords suicidal behavior; childhood adversity; school life stressors; college students; risk factors

---

1 Introduction

Suicidal behavior in young people has become a significant public health problem and has been the focus of many studies world-wide. Consequently it is urgent to study the associated risk factors for a better understanding of this problem and to establish assessment and prevention systems. In China, suicide among young people has become a particularly serious challenge for public health. There were up to 2,000,000 Chinese attempting suicide every year, with yearly suicide casualties of 287, 000 in the period of 1995–1999. Although it is the leading cause of death for young people in China, suicide in this population has not been much investigated and relevant evidence of significant correlates is meager compared with that of western countries.

Suicidality, including suicidal ideation and suicide attempt, may be influenced by many factors, ranging from genetics, family functions, social-economic status, personality and psychiatric comorbidity. The experience of adverse events during childhood has long been known to be a significant risk factor for suicidal behaviors among young adults. A large body of research, either with clinical samples, community samples, or specific groups such as drug abusers and males with hyperactivity-inattention symptoms, has revealed that a higher number of adverse experiences in childhood were linked with higher risks for suicidal ideation or suicide attempt. Similar results were also found in one study focusing on college students. At the same time, recent negative life events confer another striking risk factor for suicidal behaviors among young people. Consistent evidence has shown that recent negative life events are associated with an increased risk for onset of suicidal ideation and suicide attempt, and that the number or the severity of recent life events could predict the severity of suicidality and its repetition.

Though childhood adversity and recent negative events are individually known as prominent risk factors for suicidal behavior, few studies have explored the combinative role of these two exposures on suicidality. According to the stressor-diathesis model of suicidal behaviors proposed by Dr van Heeringen, suicidal behavior should be not only influenced by the independent effects of diathesis and stress, but also further formed by their interaction. Moreover, certain diatheses (usually associated with early childhood experience and genetics) could increase the likelihood of people encountering stressful events later in life. It is therefore very likely that childhood adversities and recent life stressors have a multiplicative impact on suicidality.

Available studies on childhood adversities and suicidal behavior have delved into the influence of sexual abuse and neglect, leaving the influence of many common adversities in one's early life unascertained. To our awareness, in research on negative events prior to suicidal behaviors, no study has simultaneously considered types of childhood adversity and recent negative events as well as their joint effects on suicidality.

To amend the missing knowledge in the area, this study aims: a) to examine the independent association of childhood adversities and recent life stressors with suicidality in a large sample of Chinese college students; b) to explore the interactive effect of the two exposures; and c) to probe the specific impacts of different types of childhood adversities and recent school life stressors on suicidal behaviors in this population.

2 Method

2.1 Ethical approval

The study was approved by the Ethical Committee for Scientific Research at Central China Normal University. A signed consent form was collected from each student.

2.2 Participants of study

Of eight universities in Wuhan, a city in central China, six agreed to join in the survey. All were attached directly to the ministries of P. R. China. Stratified cluster sampling was used first to identify 10% of students at each university and then to randomly select participants from classes. Of a total 7022 students selected into the study, 6096 students participated in the survey (response rate: 86.8%). 5989 respondents completed all question items designed for the present study and therefore were considered as study population in the data analyses. The sample comprised 3156 (52.8%) male and 2768 (46.2%) female students. There were 56 respondents who did not report their gender. The age range of the sample was from 14 to 26 years old (M=19.94; SD=1.38).

2.3 Instrument for assessment

Childhood Adversities(CA). In this section, participants were asked to report the worst adverse experience in the first 10 years of their life. The Childhood Adversity questionnaire listed 6 adversities that children might experience: divorce of parents, poor parental relationship, loss of a parent, family financial problems, sexual abuse, and severe physical illness, together with the option of "None of the above". Participants were asked to review each item in this part and to report if they had ever undergone any of these adversities, and if yes, then to specify which adverse experience had the worst impact on them. The score range of CA was 0–1. Students reporting "None of the above" were assigned a score with "0" and classified as "without childhood adversity", while those reporting they had experienced one or more adversities were assigned a score of "1" and classified as cases "with childhood adversity".

School Life Stressors (SLS). To assess recent negative life events in college students, the SLS questionnaire was created to measure frustrations in different domains that college students might encounter.

Based upon results from the pilot investigation for this study, nine stressors were considered: 1) poor school performance, 2) rupture of romantic relationships, 3) difficulty of adapting college life, 4) financial problems, 5) employment stress, 6) conflicts with classmates, 7) failure in making close friends and 8) other frustrations. The participants were asked which frustrations they had encountered during the past one year period. Students who reported at least one of the listed stressors were assigned a score of "1" and regarded as "with school life stressor', while those reporting "None of the above" were assigned a score of "0" and classified as "without school life stressor".

Suicidality. Participants' suicidal ideation and suicide attempts were assessed by a questionnaire comprising 5 questions: 1) Did you ever seriously consider killing yourself in the past year; 2) Have you ever seriously considered killing yourself in your life; 3) Did you ever try or attempt to kill yourself within the past one year; 4) Have you ever tried or attempted to kill yourself in your life; and 5) Have you ever had a nonfatal suicide action. Participants answered the first 4 questions by marking 0 (never), 1 (sometimes) or 2 (very often). For the last question respondents had to answer only "Yes" or "No". Participants were asked to respond to all items. Students who answered "never" for both items about suicidal ideation (items 1and 2) were classified as "Without suicidal ideations", while the others were classified as "With suicidal ideations". Similarly, students who chose "no" or "never" as the answer to all three questions about suicide attempts (items 3–5) were classified as "Without suicidal attempts"; otherwise, they were regarded as "With suicidal attempts."

Demographic information. Personal demographic information was also collected during the survey. Respondents reported personal details including age, gender, specialty of study, and region of permanent family residence (urban or rural).

2.4 Procedure

An independent website was designed for this project with all relevant questionnaires and the site was accessible only through a unique password assigned to each selected student. The students were asked to complete the survey online. Before the survey, they were given a brief introduction to the study and were ensured of the confidentiality of personal data. Students who agreed to participate signed a consent form. The on-line survey started with an overall introduction about the research purposes, and then moved to specific instructions for each questionnaire. Several pilot studies were carried out to examine whether the questionnaires were suitable and understandable, and also to test the functionality of the website. There was no report of technical problems during the final online survey collecting data from the students.

2.5 Data analyses

Data were first generated into a spreadsheet file and then analyzed with SPSS17.0. Firstly, a t-test was used to detect if there was any difference in the number of school life stressors according to the presence of childhood adversity. Secondly, three logistic regression models were conducted to assess the effects of gender, region, childhood adversity and school life stressors on risks for suicidal ideation and for suicide attempts. Model 1 estimated the individual effect of each of these variables; Model 2 estimated the separate effects of CA and the SLS with the adjustment for gender and region; while in Model 3 the effects of gender, region, SLS and CA on suicidality were considered simultaneously in one analytic model. Lastly, separate regression analyses were conducted to examine specific CA and SLS events as risk factors for suicidal ideation and suicide attempt.

3 Results

Among the 5989 college students, 982 (16.40%)

presented a positive answer to suicidal ideation some time during their life course, while 115 (1.92%) reported the presence of a suicide attempt. The prevalence of both suicidal ideation and suicide attempt was significantly higher in female students than in male students. Table 1 shows the prevalence of suicidal behaviors by variables of interest in the present study.

Examination of the relationship between childhood adversities and school life stressors showed that in the group of students "with childhood adversities", the mean number of school life stressors was 2.51 ± 1.47, whereas the corresponding number was 1.86 ± 1.34 for the group "without childhood adversities". The t-test result indicated that compared to those without any adversity in childhood, individuals who had one or more adverse childhood experience experienced more recent negative events in school ($t=30.28$, $p<0.001$).

Table 2 shows the effect of study variables on risk for suicidal ideation, derived from three logistic regression models. Results from model 1 revealed that gender, recent school life stressors and childhood adversities all had a significant association with suicidal ideation, while region of family residence did not. Compared with model 1, the OR values associated with SLS and CA in Model 2 were slightly reduced after adjustments for the effects of gender and region. When all variables were simultaneously included in model 3, gender, region, SLS and CA all had a notable influence on suicidal ideation. Moreover, compared with the estimates from model 2, the OR associated with SLS decreased by 3.60% and the OR associated with CA reduced by 15.38%, suggesting that there was an interactive effect of SLS and CA on risk for suicidal ideation.

Table 2 also displayed the results of regression analyses focusing on suicide attempt. Again, the effect sizes associated with the variables under study, as well the pattern of the effects, were very similar to those found in

Table 1. Prevalence of suicidal ideation and suicide attempts in the study population

Variables under study		Number	Suicidal ideation		Suicide attempts	
		N	%	(n)	%	(n)
Gender	Male	3165	13.02	(412)	1.33	(42)
	Female	2768	20.38	(564)	2.60	(72)
Region	Urban	3942	16.39	(646)	2.23	(88)
	Rural	2044	16.39	(335)	1.32	(27)
School life Stressor	No	953	6.40	(61)	0.84	(8)
	Yes	5036	18.29	(921)	2.13	
Childhood adversity	No	3698	13.63	(504)	1.33	(49)
	Yes	2291	20.86	(478)	2.88	(66)
Total		5989	16.40	(982)	1.92	(115)

Note. In the study population, 56 students did not report gender information and 3 students did not report the region of permanent family residence.

Table 2. Effects of gender, region, school life stressors and childhood adversities on risk for suicidal ideation and for suicide attempt

Variables	Model I				Model II				Model III			
	Wald	p	OR	95%CI	Wald	p	OR	95%CI	Wald	p	OR	95%CI
Suicidal ideation												
Gender (0/1=Male/Female)	57.87	0.00	1.71	1.49–1.97					60.19	0.00	1.76	1.53–2.03
Region (0/1=Urban/Rural)	0.71	0.94	0.94	0.82–1.08					8.62	0.00	0.80	0.70–0.93
School life stressors (0/1=No/Yes)	180.24	0.00	1.37	1.31–1.44	184.99	0.00	1.39	1.32–1.45	141.72	0.00	1.34	1.28–1.40
Childhood adversities (0/1=No/Yes)	87.64	0.00	1.93	1.68–2.22	100.39	0.00	2.08	1.80–2.40	56.48	0.00	1.76	1.52–2.04
Suicide attempt												
Gender (0/1=Male/Female)	12.26	0.00	1.99	1.35–2.92					11.24	0.00	1.94	1.32–2.86
Region (0/1=Urban/Rural)	6.64	0.01	0.60	0.41–0.89					11.34	0.00	0.51	0.34–0.75
School life stressors (0/1=No/Yes)	28.84	0.00	1.36	1.22–1.52	31.04	0.00	1.38	1.23–1.54	21.18	0.00	1.32	1.17–1.48
Childhood adversities (0/1=No/Yes)	17.85	0.00	2.24	1.54–3.25	23.54	0.00	2.58	1.76–3.79	14.96	0.00	2.17	1.47–3.22

Notes: Model I: analyses without any adjustment; Model II: adjusted for gender and region; Model III: adjusted for all variables listed in the table.

the analyses on suicidal ideation. From model 2 to model 3, there was a 4.35% reduction of the OR value associated with SLS and a 15.89% reduction of the OR associated with CA, suggesting an interactive effect of the two exposures on risk for suicide attempt.

Analyses were then conducted to examine the associations between suicidality and each item in the CA and SLS questionnaires. Associations between each specific childhood adversity and suicidality are shown in Table 3. Family financial problems constituted the most common childhood negative event (21.8%) for the Chinese college students, followed by poor parental relationship (9.5%). Among the childhood adverse experiences, sexual abuse, divorce of parents, poor parental relationship and

loss of a parent were associated with significantly increased risks for suicidal ideation ($p<0.01$ for all) and suicide attempt ($p<0.01$ for all). Though financial problems and severe physical illness in childhood had a significant effect on suicidal ideation ($p<0.01$), they did not have a significant influence on suicidal attempt ($p=0.20$ and $p=0.26$, respectively). Among all childhood adversities under study, the three carrying the most risk for suicidal ideation were sexual abuse, poor parental relationship and loss of a parent; for suicide attempts, the top three risk factors were sexual abuse, loss of a parent and divorce of parents. It is worth mentioning that even with the limited number of college students who reported sexual abuse, this form of childhood adversity carried the highest risk for suicidality.

Table 3. Logistic regression results showing risk of suicidal behaviors associated with specific adverse experiences in childhood

		case	Suicidal ideation				Suicide attempts			
			Wald	*p*	OR	95%CI	Wald	*p*	OR	95%CI
STEP 1	Gender		53.25	0.00	1.70	1.48–1.96	1.24	0.27	1.19	0.88–1.60
	Region		2.10	0.15	0.90	0.78–1.04	4.26	0.04	0.72	0.53–0.98
STEP 2	Divorce of parents	140	13.92	0.00	2.16	1.44–3.28	18.71	0.00	4.06	2.15–7.66
	Poor parental relationship	566	95.58	0.00	2.83	2.30–3.49	47.37	0.00	3.86	2.63–5.68
	Loss of parent(s)	110	17.92	0.00	2.62	1.68–4.10	16.20	0.00	4.39	2.14–9.02
	Financial problems	1307	28.05	0.00	1.62	1.36–1.94	1.64	0.20	1.32	0.86–2.01
	Sexual abuse	19	9.83	0.00	4.36	1.74–10.98	11.04	0.00	8.45	2.40–29.78
	Severe physical illness	149	19.4	0.00	2.44	1.64–3.62	1.30	0.26	1.71	0.68–4.29

Note. ORs were adjusted for all childhood adversities as well as demographic variables listed in the table.

Table 4. Logistic regression results showing risk of suicidal behaviors associated with specific stressors in school life

		case	Suicidal ideation				Suicide attempts			
			Wald	*p*	OR	95%CI	Wald	*p*	OR	95%CI
STEP 1	Gender		56.72	0.00	1.74	1.51–2.01	3.35	0.06	1.32	0.98–1.79
	Region		2.07	0.15	0.09	0.80–1.04	6.15	0.01	0.68	0.05–0.92
STEP 2	Poor school performance	3380	28.15	0.00	1.50	1.29–1.74	4.40	0.04	1.41	1.02–1.93
	Rupture of love affairs	1269	9.30	0.00	1.30	1.10–1.53	9.00	0.00	1.65	1.19–2.29
	Difficulty in adapting college life	1545	21.91	0.00	1.45	1.24–1.69	8.89	0.00	1.62	1.18–2.22
	Financial problems	1229	12.52	0.00	1.36	1.15–1.61	4.54	0.03	1.45	1.03–2.04
	Employment stress	1176	0.10	0.75	0.97	0.81–1.16	0.61	0.43	1.15	0.81–1.64
	Conflicts with classmates	706	21.09	0.00	1.59	1.30–1.93	3.03	0.08	1.42	0.96–2.10
	Failure in making close friends	2011	14.68	0.00	1.34	1.15–1.56	1.75	0.18	1.24	0.90–1.70
	Other frustration	1516	26.40	0.00	1.49	1.28–1.74	0.34	0.56	1.10	0.79–1.53

Note. ORs were adjusted for all school life stressors as well as demographic variables listed in the table.

Table 4 shows the relationships between each specific stressor on the SLS and suicidal behaviors in the study population. Poor school performance was the most common stressor in college life, presenting in 56.44% of the study population. Poor school performance, rupture of a romantic relationship, failure to adapt, financial problems, conflicts with classmates, failure to make close friends, and other frustrations, all had a significant effect (*p*<0.01) on suicidal ideation. For suicide attempts, only poor school performance (*p*<0.05), rupture of romantic relationships (*p*<0.01), failure in adaptation (*p*<0.01) and financial problems (*p*<0.05) had a significant influence. Conflicts with classmates, poor school performance and other frustrations carried the highest risk for suicidal ideation, while rupture of romantic relationships, difficulty in adaptation, and financial problems were carried the highest risk for suicidal attempt. Furthermore, there was

no significant effect of stress from employment competition on suicidal behavior in the study population.

4 Discussion

4.1 Demographic structure of suicidality in Chinese college students

The present study demonstrates that both suicidal ideation and suicide attempts were significantly more common in female students as compared with their male counterparts. This observation is consistent with the conclusion of a previous study on Chinese college students , and also in high concordance with many reports internationally. As for the effect of region, there was no significant rural–urban difference for suicidal ideation, but the risk of suicide attempt was significantly higher among students coming from urban areas than those from rural

China. This finding is somewhat different from the previous view that the suicide rate in rural China is much higher than that in urban places. A possible reason could be that the sample population in the current study is very different from those in previous reports. College students coming from rural areas were now studying and hence living in a city of central China. They were pursuing an education and enjoying a much better living environment than what they had in the rural area. It is therefore understandable that these students are at a relatively lower risk for attempting suicide compared with counterpart students coming from the urban areas.

4.2 Childhood adversity, school life stressors and suicidality

This study shows that both adverse experiences in childhood and recent life stressors in school life are strongly associated with risk for suicidal behaviors in young people studying in university. These results are consistent with previous reports focusing on early life adversities, and also studies on recent stressful life events in relation to suicidality.

This study adds to the literature demonstrating that experience of childhood adversity (CA) could increase the likelihood of exposure to negative life events (SLS) during young adulthood and that there is an interactive effect between SLS and CA on suicidality. Such observations support the stress–diathesis model of suicidal behaviors, proposed by Dr van Heeringen. Apart from the respectively independent effect of school life stressors and childhood adversities on suicidality, the two types of exposure combined to exert an interactive role on suicidality. According to the stress–diathesis model, childhood adversities could act as a diathesis that could predispose individuals to more negative events in their future life; the occurrence of later stressful events could in turn further predispose these people to become more vulnerable to the effects of stressful events; and consequently, risk for suicidal behavior becomes higher and higher through

an interactive process. There are a number of published studies providing evidence of the proposed impact pathway. For example, childhood adversity was found to be linked with personality characteristics such as aggression; the aggressiveness made the subjects more maladapted and therefore induced more stressful events. Childhood adversity could increase the tendency to form specific cognitive characteristics, such as negative attribution style and hopelessness, which could in turn increase the vulnerability to new stressors.

The observed interaction between childhood adversity and recent school life stressors could also be interpreted from developmental perspectives. One possible explanation lies in the association between attachment and suicidality. Secure attachment has been acknowledged as a strong, if not the most important, protective factor for children's psycho–social development. Especially in the face of troubles or frustrations, secure attachment could be the child's major psychological resource. As well established, early adversities or traumas are detrimental for the formation of secure attachment.Therefore, children with severe childhood adversity might show less secure attachment during their upbringing and show greater vulnerability in response to stressful events compared to peer youth raised in families that experienced no obvious adversity. It is also more difficult for them to retain stable and durable relationship with others, including with family members and classmates–factors strongly associated with suicidality among young people. The current results are consistent with the attachment hypothesis, in that interpersonal stress and conflicts with classmates showed high prevalence among students with suicidal behavior.

Moreover, childhood adversity and recent life stressors may share certain characteristics as risk factors for suicidality, the most notable being difficult family relationships. Family circumstances are usually enduring and related both socially and psychologically to the

adjustment of individual family members. Children from such families may experience ongoing suffering and are more likely to experience stressors from childhood through adolescence to adulthood. This is supported by early research demonstrating strong associations between a problematic family background, early life adversity and recent stressful life events. It is also evident that children from family backgrounds characterized by low socio-economic status, poor communication or troublesome life are at an increased risk for suicidal behaviors.

4.3 Influence of specific life stressors or negative events

The present study indicates that different types of life stressors or negative events, either in childhood or in school life, had different effects on suicidal behavior in Chinese college students. Compared with other stressful events, poor school performance was the most common stressor that Chinese college students encountered within a one year period. Among all stressors under study, those that were significantly associated with suicidality (both suicidal ideation and suicide attempts) were poor school performance, rupture of romantic relationships, difficulty in adapting school life and financial problems; conflicts with classmates, failure to make close friends and other frustrations were linked with suicidal ideation but not suicide attempts. These findings are mostly in agreement with another study of college students, showing that suicidality was especially associated with interpersonal affairs, such as interpersonal conflicts and ruptured relationships as well as school performance .

It is worth noting that employment stress did not confer significant risk for suicidal behavior in the present study. In earlier research the association between unemployment and suicidality has been viewed as ambiguous and conclusions have varied under different research conditions. To our awareness, however, no study has focused on suicidal behavior in relation to stress from possible employment competition, let alone in a sample of Chinese college students. Yet we could not simply conclude from the present data that employment stress has no significant effect on suicidal behaviors in Chinese undergraduates, because the present sample mostly consisted of students who had not yet completed their third year at their universities. Due to their limited experiences in the labor market, it is likely that job competition is not yet an urgent consideration for the subjects of the present study. On the other hand, our data demonstrate that recent stress from financial problems is strongly associated with risk for both suicidal ideation and suicide attempt. This is in line with previous studies of young adults, and underlines the importance of financial support for college students and of students' proper management of their economic situation during school life.

Regarding childhood adversities, it is interesting to see that all parent-related stressors during childhood including divorce of parents, poor parental relationship and loss of a parent - were associated with a significantly increased risk for suicidal behavior. This finding is in line with the literature, and makes particular sense when thinking of the stress-diathesis model and attachment theory discussed in the previous section. Although very few students reported the experience of sexual abuse during childhood, this exposure denoted the highest risk for suicidal ideation and attempt. On the other hand, the experience of family financial problems during childhood was the most frequently reported childhood stressor in this population, although compared to other negative childhood experiences it had the lowest influence on suicidality. Of course, family financial problems during childhood do not necessarily mean low economic status, because the problems could be due to a short term crisis. Nevertheless, the results from this study indicate that compared with insufficient resources caused by financial problems, family stressors such as parental divorce or a poor parent-child

relationship play a more important role in the development of suicidality.

4.4 Limitations

There are several limitations in the present study. One limitation is that the study is a cross-sectional investigation, which does not allow detection of any causal relationship between exposure to adversity and suicidality. It would be ideal to test the findings from this study using a cohort design, following children from their early life. Another limitation concerns generalizability. The sample was obtained strictly by a stratified cluster sampling method in six of eight universities in a major city in Central China. Although we are confident that our sample represents well the undergraduate college students in the area of Central China, it is uncertain to us if it could represent undergraduates in other places of China such as the more developed coastal or the less developed western regions. It is therefore preferable that more studies should be carried out in larger and more diverse samples to test the present findings and evaluate their generalizability.

5 Conclusion

Based on a large sample of undergraduate students in China, the present study reports a 16.40% rate of suicidal ideation and a 1.92% rate of suicide attempts. The prevalence differed significantly by gender and region of family residence, with higher rates in females than males and lower rates among students from rural areas compared to urban areas. At the same time, students with suicidal behavior more often reported the experience of negative life stressors in both childhood and recent school life. While childhood adversity and recent school life stressors each denoted significant risk for suicidal behavior, the two types of exposure tended to co-occur and act interactively to increase the risk for suicidal behavior. With noticeable variation of effects associated with specific stressors,

family relationship problems during childhood and recent interpersonal conflicts in school are of great importance in predicting risk for suicidality. These findings should be taken into account when planning programs of mental health promotion, suicide prevention and suicide intervention in university settings in China, and perhaps internationally.

References

Ackerman, B. P., Brown, E. D., Izard, C. E. (2004). The relations between persistent overty and contextual risk and children's behavior in elementary school. *Dev Psychol, 40,* 367–377. doi: 10.1037/0012-1649.40.3.367.

Bedi, S., Nelson, E. C., Lynskey. M, T., McCutcheon, V. V., Heath AC, et al. (2011). Risk for suicidal thoughts and behavior after childhood sexual abuse in women and men. *Suicide Life Threat Behav, 41*, 406–415. doi: 10.1111/j.1943-278X.2011.00040.x.

Buchmann, A., Blomeyer, D., Laucht M. (2012). Suicidal behaviors among young adults: Risk factors during development from early childhood to adolescence. *Prax Kinderpsychol Kinderpsychiatr, 61,* 32–49.

Cupina, D. (2009). Life events, gender and suicidal behaviors in the acute community setting. *Australas Psychiatry, 17,* 233–236. doi: 10.1080/10398560802680746.

Frederick, J., Goddard, C. (2008). Living on an island: consequences of childhood abuse, attachment disruption and adversity in later life. *Child Fam Soc Work, 13*, 300–310. doi: 10.1111/j.1365-2206.2008.00554.x.

Galera, C., Bouvard, M., Encrenaz, G., Messiah, A. Fombonne E. (2008). Hyperactivity-inattention symptoms in childhood and suicidal behaviors in adolescence: The Youth Gazel Cohort. *Acta Psychiatr Scand, 118,* 480–489.doi:10.1111/j.1600-

0447.2008.01262.x.

Grover, K. E., Green, K. L., Pettit, J. W., Monteith, L. L., Garza, MJ, et al. (2009). Problem solving moderates the effects of life event stress and chronic stress on suicidal behaviors in adolescence. *J Clin Psychol, 65*, 1281 – 1290. doi: 10.1002/jclp.20632.

Hawton, K., Harriss, L., Hodder, K., Simkin, S., Gunnell, D. (2001). The influence of the economic and social environment on deliberate self–harm and suicide: an ecological and person–based study. *Psychol Med, 31*, 827 – 836. doi: 10.1017/S0033291701003993.

Hawton, K. (2000). Sex and suicide gender differences in suicidal behaviour. *Br J Psychiatry, 177*, 484 – 485. doi: 10.1192/bjp.177.6.484.

Hill, E. M., Young, J. P., Nord, J. L. (1994). Childhood adversity, attachment security, and adult relationships:A reliminary tudy. *Ethol Sociobiol, 15*, 323 – 338. doi:10.1016/0162–3095(94)90006–X.

Kleiman, E. M., Miller, A. B., Riskind, J. H. (2012). Enhancing attributional style as a protective factor in suicide. *J Affect Disord, 143*, 236 – 240. doi: 10.1016/j.jad.2012.05.014.

Kwok, SYCL. (2011). Perceived family functioning and suicidal ideation:Hopelessness as mediator or moderator. *Nurs Res, 60*, 422 – 429. doi: 10.1097/NNR.0b013e31823585d6.

Moscick, E. K. (1994). Gender differences in completed and attempted suicides.*Ann Epidemiol, 4,* 152 – 158. doi: 10.1016/1047–2797(94)90062–0.

Pedersen, N., Fiske, A. (2010). Genetic influences on suicide and nonfatal suicidal behavior: Twin study findings. *Eur Psychiatry, 25,* 264 – 267. doi: 10.1016/j.eurpsy.2009.12.008.

Phillips, M. R., Li, X., Zhang, Y. (2002). Suicide rates in China, 1995 – 99. *Lancet, 359,* 835 – 840. doi: 10.1016/S0140–6736(02)07954–0.

Pompili, M., Di., Cosimo, D., Innamorati, M., Lester, D., Tatarelli, R., et al. (2009). Psychiatric comorbidity in patients with chronic daily headache and migraine: A selective overview including personality traits and suicide risk. *J Headache Pain, 10*, 283 – 290. doi: 10.1007/s10194–009–0134–2.

Sarchiapone, M., Jaussent, I., Roy, A., Carli, V., Guillaume S, et al. (2009). Childhood trauma as a correlative factor of suicidal behavior – via aggression traits. Similar results in an Italian and in a French sample. *Eur Psychiatry, 24*, 57 – 62. doi: 10.1016/j.eurpsy.2008.07.005.

Sfoggia, A., Pacheco, M. A., Grassi–Oliveira, R. (2008). History of childhood abuse and neglect and suicidal behavior at hospital admission. *Crisis, 29*, 154 – 158. doi:10.1027/0227–5910.29.3.154.

Singh, S., Manjula, M., Philip, M. (2012). Suicidal risk and childhood adversity: A study of Indian college students. *Asian J Psychiatr, 5*, 154 – 159. doi: 10.1016/j.ajp.2012.02.024.

Skopp, N. A., Luxton, D. D, Bush N, Sirotin A. (2011). Childhood adversity and suicidal ideation in a clinical military sample: Military unit cohesion and intimate relationships as protective factors. *J Soc Clin Psychol, 30*, 361 – 377. doi:10.1521/jscp.2011.30.4.361.

Spokas, M., Wenze,l A., Stirman, S. W., Brown, G. K., Beck, A. T.(2009). Suicide risk factors and mediators between childhood sexual abuse and suicide ideation among male and female suicide attempters. *J Trauma Stress ,22*, 467 – 470. doi:10.1002/jts.20438.

Sundin, M., Spak, F., Spak, L., Sundh, V., Waern, M. (2011). Substance use/abuse and suicidal behavior in young adult women: A population–based study. *Subst Use Misuse, 46,* 1690 – 1699. doi: 10.3109/10826084.2011.605414.

Tang, J., Yu, Y., Wu, Y., Du, Y., Ma, Y., et al. (2011) . Association between nonsuicidal self–injuries and suicide attempts in Chinese adolescents and college

students: A crosssection study. *PloS One, 6*, e17977. doi: 10.1371/journal.pone.0017977.

Van, Heeringen. K. (2000). A stress diathesis model of suicidal behavior. *Crisis, 21,*192a. doi: 10.1201/b12215–7.

Van, Orden. K. A., Witte, T. K., Cukrowicz, K. C., Braithwaite, S. R., Selby, E. A., et al.(2010). The interpersonal theory of suicide. *Psychol Rev, 117,* 575–600. doi:10.1037/a0018697.

Vazquez, J. J., Panadero, S., Rincon, P. P. (2010). Stressful life events and suicidal ehaviour in countries with different development levels: Nicaragua, El alvador, Chile and Spain. *J Community Appl Soc Psychol, 20,* 288–298. doi:10.1002/casp.1036.

Violato, C., Arato, J. E. (2004). Childhood attachment and adolescent suicide: A stepwise discriminant analysis in a case–comparison study. *Individ Differ Res, 2,* 162–168.

Wang, Y., Sareen, J., Afifi, T. O., Bolton, S, L, Johnson, E. A., et al. (2012). Recent stressful life events and suicide attempt. *Psychiatr Ann, 42,* 101–108. doi:10.3928/00485713–20120217–07.

Yen, S., Shea, M. T., Sanislow, C. A., Skodol, A. E., Grilo, C. M., et al. (2009). Personality traits as prospective predictors of suicide attempts. *Acta Psychiatr Scand, 120,* 222–229. doi: 10.1111/j.1600–0447.2009.01366.x.

Zhang, J., Jia, S., Jiang, C., Sun, J. (2006). Characteristics of Chinese suicide attempters: An emergency room study. *Death Stud, 30,* 259–268. doi: 10.1080/07481180500493443.

Zhang, J., Ma, Z. (2012). Patterns of life events preceding the suicide in rural young Chinese: A case control study. *J Affect Disord, 140,* 161–167. doi: 10.1016/j.jad.2012.01.010.

Effects of life satisfaction and psychache on risk for suicidal behaviour: A cross-sectional study based on data from Chinese undergraduates

Zhiqi You[1,2], Juanjuan Song[1,2], Caizhi Wu[1,2], Ping Qin[3,4], Zongkui Zhou[1,2]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan, China)

([2] School of Psychology, Central China Normal University, Wuhan, China)

([3] National Center for Register-based Research, University of Aarhus, Aarhus, Denmark)

([4] National Center for Suicide Research and Prevention, Institute of Clinical Medicine, University of Oslo, Oslo, Norway)

Abstract Objectives: To examine predictive power of psychache and life satisfaction on risks for suicidal ideation and suicide attempt among young people. Design: A cross-sectional study. Setting: Data were collected from an online survey in Wuhan, China. Participants: 5988 university students from six universities were selected by a stratified cluster sampling method. Primary and secondary outcome measures: Suicidal ideation and suicide attempt at some point of the students' lifetime were the outcomes of interest. Results: Students with suicidal ideation or attempted suicide reported a lower level of life satisfaction and high degree of psychache than counterparts without suicidal ideation or attempt. Regression analyses indicated that life satisfaction and psychache were significantly associated with the risk of suicidal ideation and the risk of suicidal attempt. Though psychache showed a relatively stronger predictive power than life satisfaction, the effect of the two factors remained significant when they were individually adjusted for personal demographic characteristics. However, when the two factors were included in the model simultaneously to adjust for each other, psychache could fully explain the association between life satisfaction and suicidal attempt. Life satisfaction remained to contribute unique variance in the statistical prediction of suicidal ideation. Conclusions: Psychache and life satisfaction both have a significant predictive power on risk for suicidal behaviour, and life satisfaction could relieve the predictive power of psychache when suicidal behaviour is just starting. Shneidman's theory that psychache is the pre-eminent psychological cause of suicide is perhaps applicable only to a more serious form of suicidal behaviour.

Keywords suicidal ideation; suicide attempt; life satisfaction; psychache; Chinese undergraduates

1 Introduction

Suicidal behaviour in young people has become an increasing social and public health problem in contemporary China as well as in many other parts of the world. With around 287 000 casualties in the total population of China each year, suicide is the leading cause of death in young people aged 15–34 years and the mortality among Chinese aged 15–24 years stands in the second highest place among the 39 countries that provided the data on suicide to the WHO. Suicide in China is known for its high OR of the rates in rural versus urban areas, and in women versus men particularly in young age groups, but the problem in other forms as suicidal ideation and suicide attempt in Chinese youth is generally less known and warrants thorough investigation.

Suicidal behaviour is a complicated phenomenon. It is associated with social and environmental factors and particularly with a variety of traits and disadvantages on a personal level. A great number of studies have demonstrated significant relationships of suicide with personal sociodemographics, life stressors, biological features, psychological traits, psychiatric problems, coping skills, etc. Psychological pain is thought to be of great importance, playing an indispensable role in the process of becoming suicidal.

Psychache, according to Shneidman, is defined as an acute state of intense and intolerable psychological pain that encompasses shame, guilt, humiliation, loneliness, fear, angst, dread, anguish, etc. It is associated with many psychological suicidogenic factors such as depression and hopelessness, but is also conceptually distinct from these factors. Evidence has shown that psychache is a significant and unique predictor for various suicide criteria, including suicidal ideation and suicide attempt, even after being controlled for effects of depression and hopelessness.

According to Shneidman, psychache is fundamentally linked with suicide on its own, whereas other psychological factors and affective states (eg, depression, hopelessness, etc) are relevant to suicide only insofar as they lead to psychache. In other words, if psychache is controlled, the single effect of many factors would be largely attenuated or become insignificant. Thus, psychache is the switch deciding whether a suicide happens or not. In concrete terms, if psychache is too much to tolerate for a person, the person will think dying is the best solution. As a way of relief from the intolerable psychache the individual may choose a lethal method of suicide. In contrast, if psychache could be relieved or mollified, the choice of a lethal method of suicide would be reduced, thus extending their lives.

It is suggested that completion of suicide is a balanced consequence, depending on the presence of risk factors and the absence of protective factors against suicide. However, most of the available studies about the preeminence of psychache in suicidality have only involved negative psychological variables, such as hopelessness, depression and alexithymia. Little research attention has delved into positive factors, such as life satisfaction—a measure that is reported to be strongly and negatively correlated with suicidal ideation and suicide attempt and to have a long–term effect on the risk of suicidal behaviour. At the same time, suicide studies testing Shneidman's model have been predominantly from North America; to the best of our knowledge, there is a sparsity of studies that have tested Shneidman's theory with a large sample from another culture. To better understand the role of psychache in the mechanism of suicidality, we believe that studies that take into account effects of positive factors, such as life satisfaction, and that use data from diverse cultural backgrounds would provide interesting insights.

We therefore conducted the present study with the

primary purpose of testing Shneidman's theory of suicide by assessing the predictive power of life satisfaction and psychache on risk for suicidal behaviour in a large sample of Chinese college students. Our specific aims are: (1) to examine the prevalence of suicidal ideation and suicide attempt among Chinese college students, (2) to assess the predictive power of life satisfaction and psychache on risk for suicidal behaviour and (3) to explore the role of psychache in the mechanism of suicidality.

2 Method

2.1 Participants

Of eight universities that are attached directly to the ministries of the People's Republic of China and located in the city of Wuhan, six universities agreed to join the survey for this study. A stratified cluster sampling method was used to draw study subjects from all undergraduate students in these universities. A 10% sample of undergraduate students was assigned to each university, and then randomly selected in classes—the cluster unite that is organised according to specialty and school year with usually 30–80 students. In case a selected class has more than 100 students, 100 students were drawn randomly from this class. Otherwise, all students in the drawn classes were enrolled into the study. The rationale of restricting samples from a large class to up to 100 students is to reduce the likelihood of overweight of large classes and thus to ensure a better representativeness of the 10% sampled students in each university. Consequently, we drew a 10% sample students from all six universities, consisted of 93 cluster units (classes). Of 21 large classes (22.58%), 100 students were further sampled.

With this sampling procedure, a total of 7220 college students were sampled and 6096 students finally attended the questionnaire survey for the data collection, corresponding to a response rate of 84.44%. Most students who did not attend the survey were out of the university campus for their internship during the period when the survey was conducted.

The survey was conducted online. Each selected student was assigned with an encrypted code unique to their student identification to be used as a personal password for online access to the website designed for the survey. Students enrolled into the survey were informed about the purpose of the study, the confidentiality of personal information and the principle of voluntariness. Of 6096 students who attended the online survey, 5988 completed all question items designed for this study and were therefore included in the final dataset for analyses.

2.2 Measurements

For each participant, we collected data on personal general information and demographic status such as gender, age, place of family residence, school year and specialty of study alongside data on life satisfaction, psychache and suicidal behaviour as described below.

2.2.1 Life satisfaction

Life satisfaction was assessed with the Satisfaction with Life Scale, which is a self–report questionnaire comprising five items, that is, (1) 'In most ways my life is close to my ideal'; (2) 'My life condition is very good'; (3) 'I am satisfied with my life', (4) 'I have got the important things which I want in the life' and (5) 'If I had new life, I would enjoy it just as I do now'. All items are answered on a seven–point Likert scale (1, strongly disagree; 2, disagree; 3, incomplete disagree; 4, not sure; 5, incomplete agree; 6, agree and 7, strongly agree). The average score of the five items was calculated as the score of life satisfaction (scores range 1–7). The Satisfaction with Life Scale has been reported with high α reliability coefficients; 0.89 in one study and 0.86 in another. In the present study, the α reliability coefficient was 0.85.

2.2.2 Psychache

Psychache was measured with the Psychache

Scale, which is a 13—item (eg, My soul aches) self—report questionnaire designed to assess Shneidman's conceptualisation of psychache. All items are answered on a five—point Likert scale and the total scores range from 13 to 65 (high score indicates a high psychache level). In order to precisely capture the level of psychache, we constructed a continuous variable to categorise the score into five levels defined after taking into account the frequency distribution and score value as 1=13—15, 2=16—19, 3=20—22, 4=23—26 and 5=27—65. The α reliability coefficients of this scale were generally exceeding 0.90, and the scale could distinguish between suicide attempters and non—attempters. In the present study, the α reliability coefficient was 0.92.

2.2.3 Suicidal ideation and suicide attempt

Suicidal ideation was defined as thoughts or wishes to be dead or to kill oneself according to Schneidman and Silverman *et al*. It was assessed through the following two question items: (1) 'have you seriously considered about killing yourself in the past 1 year', and (2)'have you ever seriously considered about killing yourself in your life'. These two items were answered on a threepoint Likert rating (0, never; 1, sometimes and 2, very often). Participants who answered with '0' for both items were regarded as 'without suicidal ideation', whereas all others were regarded as 'with suicidal ideation'.

Suicide attempt was defined as a self—inflicted behaviour with a non—fatal outcome for which there is evidence of intent to die. It was assessed with three questions: (1) 'have you ever tried or attempted to kill yourself in the past year', (2) 'have you ever tried or attempted to kill yourself in your life' and (3) 'have you ever taken any non—fatal suicidal action'. The first two items were answered on a three—point Likert rating (0, never; 1, sometimes and 2, very often) and the third item had a binary answer of 'yes' or 'no'. Participants who answered with '0' for the first two items and 'no' for the

third item were regarded as 'without suicide attempt', and otherwise regarded as 'with suicide attempt'.

2.3 Statistical analysis

χ^2 Test was used to examine the distribution difference of suicidal behaviour (suicidal ideation and suicidal attempt) by demographic variables. T test was conducted to test the variation of scores on life satisfaction and psychache according to the presence of suicidal behaviour. Logistic regression was used to assess the predictive power of psychache and life satisfaction on risk of suicidal behaviour and also to evaluate the relative importance of these two factors via three models. Model I estimated the crude effect of each variable of interest; model II estimated the individual effect of life satisfaction and of psychache in the adjustment of demographic variables and model III adjusted the effect when all variables were included in the model.

3 Results

3.1 General description

The total sample of 5988 students comprised 3203 male and 2785 female students, with the age ranging from 14 to 26 years (mean=19.94; SD=1.38). Table 1 shows the demographic distribution of the study participants in details.

Of the total students, 16.4% (982) participants reported the presence of suicidal ideation and 1.9% (114) reported a history of suicide attempt at some point of their lifetime. The prevalence differed significantly with regard to most demographic variables as shown in table 2. Suicidal ideation was more prevalent in female than male students (20.4% vs 12.9%), in students coming from large cities as provincial capitals or direct—controlled municipalities (18.9%), and in students studying liberal arts (20.8%), in freshman students (17.8%), and consistent with this, in students in relatively younger age groups. These observed patterns remained very similar when looking at the self—reported presence of suicide attempt, although

the differences by most demographic factors did not reach a statistical significance. Still, it is evident that female students reported a significantly higher rate of suicide attempt (2.6%) than their male counterparts studying in the universities (2.6% vs 1.3%).

3.2 Distribution of life satisfaction and psychache according topresence of suicidal behaviour

Figure 1 shows the frequency distribution of students according to the level of life satisfaction and the raw score of psychache. For life satisfaction, the scores exhibited a normal distribution, ranging from 1 to 7. However, the scores of psychache exhibited a partial normal distribution with the majority having a low score and a few at the high end. According to Shneidman's theory, psychache is an intense and intolerable extreme state involving heavy psychological pain. It is therefore reasonable that most of the people experienced a mild psychological pain while intense psychological pain was present in the minority.

Table 3 presents the means and SDs of the scores on life satisfaction and psychache according to the presence of suicidal ideation and suicide attempt. The results from t tests indicate that students with suicidal ideation or suicide attempt reported a significantly lower score of life satisfaction and a significantly higher score of psychache. This, in other words, means that a high level of life satisfaction and low degree of psychache are proactive against suicidal behaviour within this population.

3.3 Effect of life satisfaction and psychache on risk for suicidal ideation and suicide attempt

Table 4 displays the results from modelling the data with logistic regression in order to assess the statistical predictive power of psychache and life satisfaction on risk for suicidal ideation and suicide attempt.

The results from model I and model II both indicated that life satisfaction and psychache had a significant power in predicting suicidal ideation and suicide attempt, and that a higher level of life satisfaction was associated with a reduced risk while a higher level of psychache was associated with an increased risk for suicidal ideation and for suicide attempt. In model III, psychache continued to

Table1 Demographic distribution of students in the study

Demographic variables	Number (%)		
	Total	Male	Female
Residence place of family			
Provincial capital or direct–controlled municipality	924 (15.4)	413 (44.7)	511 (55.3)
County–level city	1353 (22.6)	665 (49.2)	688 (50.8)
County	803 (13.4)	417 (51.9)	386 (48.1)
Township	861 (14.4)	465 (54.0)	396 (46.0)
Suburban countryside	930 (15.5)	510 (54.8)	420 (45.2)
Remote countryside	1114 (18.6)	733 (65.8)	381 (34.2)
Specialty of study			
Engineering	2036 (34.0)	1568 (77.0)	468 (23.0)
Science	2310 (38.6)	1107 (47.9)	1203 (52.1)
Liberal arts	1642 (27.4)	528 (32.2)	1114 (67.8)
Grade			
Freshman	2770 (46.3)	1351 (48.8)	1419 (51.2)
Sophomore	1748 (29.2)	1025 (58.6)	723 (41.4)
Junior	1134 (18.9)	589 (51.9)	545 (48.1)
Senior	336 (5.6)	238 (70.8)	98 (29.2)
Age,(years)			
14–1	828 (13.8)	352 (42.5)	476 (57.5)
19–21	4390 (73.3)	2349 (53.5)	2041 (46.5)
22–26	770 (12.9)	502 (65.2)	268 (34.8)

In the study population, three students did not report the region of permanent family residence.

have a significant effect on the risk for suicidal ideation and suicide attempt. The effect of life satisfaction, however, differed somehow in model III; its predictive power remained highly significant for the risk of suicidal ideation but attenuated into insignificant for suicide attempt. This means that when the effects of demographic variables and psychache were controlled, life satisfaction did not have a significant predictive effect on suicide attempt; it was relevant to suicide attempt only insofar as it led to psychache.

4 Discussion

In this study, we have investigated the role of life satisfaction and psychache on suicidal ideation and suicide attempt among university students in China using a large random sample from six universities. It is, to our awareness, the first study to examine Schneidman's theory on psychache and suicidology in a Chinese culture, and also the first to assess the role of psychache on suicidality in the context of the positive psychological factor of life satisfaction.

4.1 Differences in suicidal ideation and attempt by gender, grade and specialty

The present study demonstrates that suicidal ideation and suicide attempt were more prevalent in female students than in male students. This mostly aligns with previous studies from China and many other places reporting a higher prevalence of suicidal behaviours in young females than young male students. One explanation for the higher female versus male suicidal behaviour in our study population perhaps lies in the Chinese culture, especially the deep-rooted Confucianism produced in a patriarchy which tended to denigrate women. On the other

Table 2 Prevalence of suicidal ideation and suicide attempt by demographic variables

	Number	Suicidal deation	Test of ifference	Suicidal attempt	Test of difference
	N	N (%)		N	N (%)
Total	5988	982(16.4)		114(1.9)	
Gender			χ^2=60.63		χ^2=12.95
Males	3203	414(12.9)	df=1	42(1.3)	df=1
Females	2785	568(20.4)	p=0.0000	72(2.6)	p=0.0002
Residence place of family			χ^2=6.26		χ^2=8.92
Provincial capital or direct-controlled municipality	924	175(18.9)	df=5	26(2.8)	df=5
County-level city	1353	218(16.1)	p=0.2819	27(2.0)	p=0.1123
County	803	124(15.4)		19(2.4)	
Township	861	129(15.0)		15(1.7)	
Suburban countryside	930	151(16.2)		12(1.3)	
Remote countryside	1114	184(16.5)		15(1.3)	
Specialty of study			χ^2=33.16		χ^2=4.51
Engineering	2036	289(14.2)	df=2	30(1.5)	df=5.p=0.105
Science	2310	351(15.2)	p=0.0000	44(1.9)	
Liberal arts	1642	342(20.8)		40(2.4)	
School year			χ^2=21.22		χ^2=5.61
Freshman	2770	493(17.8)	df=3	58(2.1)	df=3
Sophomore	1748	270(15.4)	p=0.0001	31(1.8)	p=0.1322
Junior	1134	191(16.8)		24(2.1)	
Senior	336	28(8.3)		1(0.3)	
Age,(years)			χ^2=8.56		χ^2=3.79
14-18	828	150(18.1)	df=2	15(1.8)	df=2
19-21	4390	732(16.7)	p=0.0138	91(2.1)	p=0.1500
22-26	770	100(13.0)		8(1.0)	

In the study population, three students did not report the region of permanent family residence.
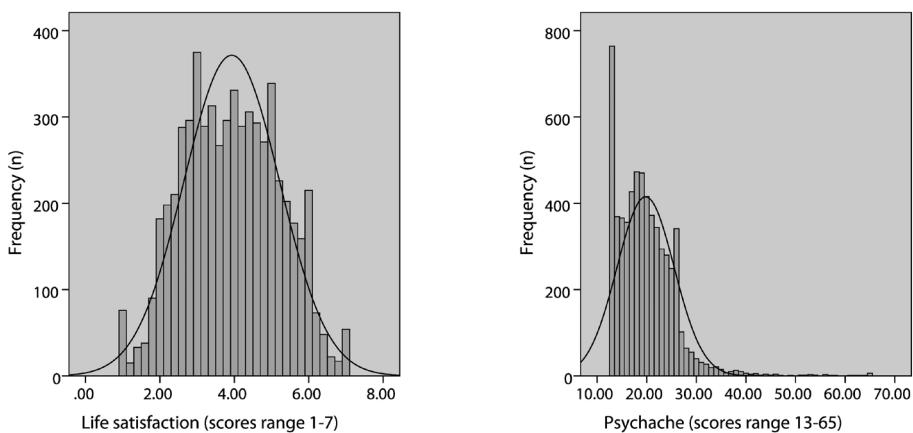
Fig. 1 Distribution of life satisfaction and psychache scores in the study subjects

hand, since the observed phenomenon exists worldwide regardless of culture, we believe that the gender-specific personality traits may, to a large extent, contribute to our observation.

Our finding that students studying liberal arts reported a significantly higher occurrence of suicidal ideation and also modestly higher suicide attempt than their peers studying Engineering and Science is interesting. Of course, the fact that relatively more female students choose to study liberal arts than other subjects may contribute to the result. It may also reflect a selection of personality traits that liberal arts students may react more sensitively and emotionally to events and stressors than their counterpart students studying other subjects.

In addition, we note that students in the early school years or at a younger age reported more suicidal ideation than their higher grade or elder peers. These results are somehow as expected and could well be explained by the hardship to adapt to the college life and limited experience in handling stressors at the beginning of college life. With the increase of age and experience, the students become better in coping with and handling stressors that they would not be able to handle at younger ages.

4.2 Psychache, life satisfaction and suicidal behaviour

The present study indicates that psychache and life satisfaction were both significant predictors for suicidal

ideation and suicide attempt in the study population. Psychache had stronger power in predicting suicidal ideation and suicide attempt than life satisfaction, as evidenced by its larger standardised regression coefficients. This result is in line with the finding from other studies testing Shneidman's theory with samples of general population and special groups as offenders, homeless people and patients with mental illness. The result supports Shneidman's view that psychache is a fundamentally important predictor for suicidal behaviour.

Moreover, our analyses further indicate that the mediating role of psychache on the occurrence of suicidal behaviour differed slightly between suicidal ideation and suicide attempt. In the prediction of suicide attempt, psychache accounted for a greater proportion of variance than did life satisfaction. The association between life satisfaction and suicidal attempt were completely residualised by psychache, as Shneidman had predicted. Life satisfaction was relevant to suicide attempt only when it was associated with psychache. However, for suicidal ideation, life satisfaction, independent of psychache, continued to have an additional significant contribution. In other words, psychache was not inevitable to suicidal ideation and played a partial role mediating the link between life satisfaction and suicidal ideation. This observation is in line with findings in a number of published studies. For instance, a large scale study of

Table 3 Scores of life satisfaction and psychache by presence of suicidal ideation and suicide attempt

Number of Participants	Mean (SD)	
	Life satisfaction	Psychache
Total	3.93（1.29）	1.56 (1.27)
Suicidal ideation		
Yes 982	3.50（1.22）	2.44 (1.21)
No 5006	4.02（1.28）	1.39 (1.20)
T Test	−11.97	24.95
Df	5986	5986
p Value	0.0000	0.0000
Suicide attempt		
Yes 114	3.48（1.44）	2.55 (1.36)
No 5874	3.94（1.28）	1.54 (1.26)
Test	−3.81	8.49
Df	5986	5986
p Value	0.0001	0.0000

Life satisfaction score ranges from 1 to 7; Psychache score ranges from 1 to 5.

college students showed that psychache did not fully mediate the association between suicidal ideation and hopelessness; the change in hopelessness during the follow-up period remained a significant predictor. One possible explanation to this result is that psychache may be associated with more serious forms of suicidal behaviour. Suicide is known as a continuous process with suicide attempt being a more severe form than suicidal ideation. Many factors influence suicidal attempt only insofar as they are related to psychache, so psychache is the inevitable channel to suicidal attempt. A study conducted by Mendonca and Holden on population-based psychiatric inpatients also demonstrated that psychache exhibited a stronger influence in persons who had actually formulated a plan for suicide than those who had general suicidal desires. That is to say that psychache represents a more distal vulnerability in a chain culminating in suicidality.

In addition, our results show that psychache is independently and positively associated with suicidal ideation and suicide attempt, and that the ORs associated with psychache were eliminated when adjusted for effects of life satisfaction and demographic variables simultaneously. These results suggest that life satisfaction may be opposed and relieve psychache, and thus verify Shneidman's theory that if the psychache

Table 4 Predictive power of life satisfaction and psychache on risk of suicidal ideation and of suicide attempt

Variables of study	Model I					Model II					Model III				
	B	Wald	p Value	OR	95%CI	B	Wald	p Value	OR	95%CI	B	Wald	p Value	OR	95%CI
Suicidal ideation															
Gender	−0.55	59.75	0.0000	0.58	0.51 to 0.67						−0.64	71.53	0.0000	0.53	0.45 to 0.61
Grade	−0.13	11.04	0.0009	0.88	0.81 to 0.95						0.04	0.69	0.4049	0.96	0.87 to 1.06
Ages	−0.19	7.48	0.0062	0.83	0.73 to 0.95						−0.07	0.78	0.3785	0.93	0.79 to 1.10
Life satisfaction	−0.32	127.81	0.0000	0.73	0.69 to 0.71	−0.36	151.82	0.0000	0.70	0.66 to 0.74	−0.15	19.89	0.0000	0.81	0.81 to 0.92
Psychache	−0.67	497.41	0.0000	1.96	1.85 to 2.08	−0.68	498.28	0.0000	1.98	1.86 to 2.10	0.63	379.54	0.0000	1.76	1.76 to 2.00
Suicide attempt															
Gender	−0.69	12.47	0.0004	0.50	0.34 to 0.74						−0.73	13.18	0.0003	0.48	0.33 to 0.72
Grade	−0.15	1.95	0.1626	0.86	0.70 to 1.06						−0.06	0.25	0.6164	0.94	0.73 to 1.20
Ages	−0.20	1.17	0.2802	0.82	0.57 to 1.17						−0.04	0.04	0.5810	0.96	0.63 to 1.46
Life satisfaction	−0.29	14.27	0.0002	0.75	0.65 to 0.87	−0.33	17.85	0.0000	0.72	0.62 to 0.84	−0.11	1.56	0.2123	0.90	0.76 to 1.06
Psychache	0.62	63.12	0.0000	1.86	1.60 to 2.17	0.62	62.90	0.0000	1.87	1.60 to 2.18	0.58	47.10	0.0000	1.79	1.52 to 2.11

Value of variables: Gender 0(female) −1 (male); Grade 1−4; Age group 1 (14−18 years old) − 3 (22−16 years old); Life satisfaction score 1−7; Psychache score 1−5.

Model I: crude analyses without any adjustment; model II: only adjusted for gender, grade and age; Model III: adjusted for all variables listed in the table.

surpasses the threshold of tolerance and is subjectively judged to be unbearable, intolerable or unacceptable, the individuals would then die from suicide in order to escape from the suffering. However, if psychache is relieved, for example, via effect of protective factors such as life satisfaction, then the originally unbearable psychological pain become acceptable and the individual would stay to live on.

From a clinical point of view, the phenomenology of suicide refers to the inner world of individuals and focuses on what the individual feels as well as understanding from the inside whenever a clinician encounters a patient. Shneidman considers psychache to be the main ingredient of suicide, and regards suicide not as a movement towards death but rather as a remedy to escape from intolerable emotion, unendurable or unacceptable anguish. Suicidal individuals experience dichotomous thinking, wishing for either some specific (almost magical) total solution for their psychache or cessation (suicide), and suicide is the result of an interior dialogue during which the mind scans its options. The present study indicates that life satisfaction may relieve the psychache and therefore reduces the risk for suicidal ideation and suicide attempt. Treatment for psychache, for example, using anodyne psychotherapy to mollify unbearable psychological pain, may well have an effect on reducing the risk for suicidal behaviour.

4.3 Limitations of the study

There are several limitations of the present study. First, like most research in this area, the present study is a cross-sectional investigation. This makes it impossible to document any causal relationship of life satisfaction and psychache with suicidal ideation or suicide attempt. While the present study is supportive of Shneidman's view concerning the pre-eminent role of psychache on suicidal behaviour, further studies with a cohort design are needed to verify the possible causal

pathways. Second, self-reported suicide attempts have limited validity or reliability due to recall bias; we are unable to verify if the reported suicide attempt truly happened in the students' real life. Another concern is related to the generalisability of the findings from the present study. The study students were sampled strictly by a stratified cluster sampling method and from six universities in a major city in Central China. Yet we are confident that these students could well represent all undergraduate college students in the area of Central China; we are uncertain whether they are representative of undergraduates studying in other places of China such as the more developed coasts and the less developed western China. This calls for more research to test the model using samples from other parts of China as well from areas with different cultures. Moreover, to our awareness, no protective variable such as life satisfaction has been encompassed in Shneidman's model so far. Some postulates, for example, that life satisfaction could relieve psychache and thus prevent suicide, need to be further verified by studies from various social settings.

4.4 Conclusions

The present study sought to test Shneidman's theory of suicide as psychache using a large sample of university students in China, and to extend existing research by including the protective variable of life satisfaction. The results indicate that psychache and life satisfaction both contribute to the risk for suicidal ideation and suicide attempt, and that psychache plays a mediating role on the link between life satisfaction and suicidal behaviour. More specifically, psychache could fully mediate the relationship of life satisfaction with suicidal attempt, but acts as a partial mediator linking life satisfaction with suicidal ideation, suggesting that Shneidman's theory is probably more applicable to severe forms of suicidality.

Referemces

Alexander, M. J., Haugland, G., Ashenden, P., et al. (2009). Coping with thoughts of suicide: techniques used by consumers of mental health services. *Psychiatr Serv, 60*, 1214 – 21.

Berlim, M. T., Mattevi, B. S., Pavanello, D. P., et al. (2003). Psychache and suicidality in adult mood disordered outpatients in Brazil. *Suicide Life Threat Behav, 33*, 242 – 8.

Blum, R., Sudhinaraset, M., Emerson, M. R. (2012). Youth at risk: suicidal thoughts and attempts in Vietnam, China, and Taiwan. *J Adolesc Health, 50*, 37 – 44.

Bray, I., Gunnell, D.(2006). Suicide rates, life satisfaction and happiness as markers for population mental health. *Soc Psychiatry Psychiatr Epidemiol, 41*, 333 – 7.

DeLisle, M. M., Holden, R. R. (2009). Differentiating between depression, hopelessness, and psychache in university undergraduates. *Meas Eval Couns Dev, 42*, 46 – 63.

DeLisle, M. M. (2007). *Understanding the relationship between depression, hopelessness, psychache and suicide risk*. Queen's University.

Desjarlais, R., Eisenberg, L., Good, B., et al. (1995). *World mental health: problems and priorities in low–income countries*. Oxford University Press.

Diener, E., Emmons, R. A., Larsen, R. J., et al. (1985). The satisfaction with life scale. *J Pers Assess, 49*, 71 – 5.

Diener, E., Fujita, F. (1995). Resources, personal strivings, and subjective well–being: a nomothetic and idiographic approach. *J Pers Soc Psychol, 68*, 926.

Diener, E., Suh, E. M., Lucas, R. E., et al. (1999). Subjective well–being: three decades of progress. *Psychol Bull, 125*, 276 – 302.

Dwivedi, Y., Pandey, G. N. (2011). Elucidating biological risk factors in suicide: role of protein kinase A. *Prog Neuropsychopharmacol Biol Psychiatry, 35*, 831 – 41.

Evans, E., Hawton, K., Rodham, K., et al. (2005). The prevalence of suicidal phenomena in adolescents: a systematic review of population–based studies. *Suicide Life Threat Behav, 35*, 239 – 50.

Fawcett, J. (2012). Suicide and life stress: past and present. *Psychiatr Ann, 42*, 82.

Garlow, S. J., Rosenberg, J., Moore, J. D., et al. (2008). Depression, desperation, and suicidal ideation in college students: results from the American Foundation for Suicide Prevention College Screening Project at Emory University. *Depress Anxiety, 25*, 482 – 8.

Heisel, M. J., Flett, G. L. (2004). Purpose in life, satisfaction with life, and suicide ideation in a clinical sample. *J Psychopathol Behav Assess, 26*, 127 – 35.

He, Z. X., Lester, D. (1997). The gender difference in Chinese suicide rates. *Arch Suicide Res, 3*, 81 – 9.

Holden, R. R., Mehta, K., Cunningham, E. J., et al. (2001). Development and preliminary validation of a scale of psychache. *Can J Behav Sci, 33*, 224 – 32.

Keefer, K. V., Holden, R. R., Gillis, K. (2009). *The mediational role of psychache in the relationship between alexithymia and suicidal ideation*. Montreal, Canada: Canadian Psychological Association Annual Convention.

Keenan, K., Hipwell, A., Feng, X., et al. (2008). Subthreshold symptoms of depression in preadolescent girls are stable and predictive of depressive disorders. *J Am Acad Child Adolesc Psychiatry, 47*, 1433 – 42.

Koivumaa–Honkanen, H., Honkanen, R., Viinamaeki, H., et al. (2001). Life satisfaction and suicide: a 20–year follow–up study. *Am J Psychiatry, 158*, 433 – 9.

Lau, R., Morse, C. A., Macfarlane, S. (2010). Psychological factors among elderly women with suicidal intentions or attempts to suicide: a controlled comparison. *J Women Aging, 22*, 3 – 14.

Law, B., Shek, D. (2013). Self–harm and suicide attempts

among young Chinese adolescents in Hong Kong: prevalence, correlates, and changes. *J Pediatr Adolesc Gynecol, 26*(Suppl 3), S26 – 32.

Lewinsohn, P. M., Rohde, P., Seeley, J. R. (1996). Adolescent suicidal ideation and attempts: prevalence, risk factors, and clinical implications. *Clin Psychol Sci Pract, 3*, 25 – 46.

Li, Y., Li, Y., Cao, J. (2012). Factors associated with suicidal behaviors in mainland China: a meta-analysis. *BMC Public Health, 12,* 524.

Mendonca, J. D., Holden, R. (1996). Are all suicidal ideas closely linked to hopelessness? *Acta Psychiatr Scand, 93*, 246 – 51.

Mills, J. F., Green, K., Reddon, J. R. (2005). An evaluation of the Psychache Scale on an offender population. *Suicide Life Threat Behav, 35*, 570 – 80.

Nock, M. K., Borges, G., Bromet, E. J., et al. (2008). Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *Br J Psychiatry, 192*, 98 – 105.

Patterson, A. A., Holden, R. R. (2012). Psychache and suicide ideation among men who are homeless: a test of Shneidman's model. *Suicide Life Threat Behav, 42*, 147 – 56.

Pelkonen, M., Karlsson, L., Marttunen, M. (2011). Adolescent suicide: epidemiology, psychological theories, risk factors, and prevention. *Curr Pediatr Rev, 7*, 52 – 67.

Pereira, E. J., Kroner, D. G., Holden, R. R., et al. (2010). Testing Shneidman's model of suicidality in incarcerated offenders and in undergraduates. *Pers Individual Dif, 49*, 912 – 17.

Phillips, M. R., Liu, H., Zhang, Y. (1999). Suicide and social change in China.*Cult Med Psychiatry, 23*, 25 – 50.

Phillips, M. R., Li, X., Zhang, Y. Suicide rates in China, 1995 – 99. Lancet 2002;359:835 – 40.

Pompili, M. (2010). Exploring the phenomenology of suicide. *Suicide Life Threat Behav, 40*, 234 – 44.

Qin, P., Mortensen, P. B. (2010). Specific characteristics of suicide in China. *Acta Psychiatr Scand, 103*, 117 – 21.

Reinecke, M. A., Simons, A. (2005). Vulnerability to depression among adolescents: implications for cognitive-behavioral treatment. *Cogn Behav Pract, 12*, 166 – 76.

Shneidman, E. S. (2005). Anodyne psychotherapy for suicide: a psychological view of suicide. *Clin Neuropsychiatry, 2*, 7 – 12.

Shneidman, E. S. Aphorisms of suicide and some implications for psychotherapy. Am J Psychother 1984;38:319 – 28.

Shneidman, E. S. (1999). Perturbation and lethality: a psychological approach to assessment and intervention. In: Jacobs DG.ed The Harvard Medical School guide to suicide and intervention. *San Francisco: Jossey-Bass,* 83 – 97.

Shneidman, E. S. (1993). *Suicide as psychache: a clinical approach to self-destructive behavior.* Northvale, NJ: Jason Aronson.

Shneidman, E. S. (1993). Suicide as psychache. *J Nerv Ment Dis, 181*, 145 – 7.

Shneidman, E. S. (1996). *The suicidal mind.* New York: Oxford University Press.

Silverman, M. M., Berman, A. L., Sanddal, N. D., et al. (2007). Rebuilding the tower of babel: a revised nomenclature for the study of suicide and suicidal behaviors part 1: background, rationale, and methodology. *Suicide Life Threat Behav, 37*, 248 – 63.

Silverman, M. M., Berman, A. L., Sanddal, N. D., et al. (2007). Rebuilding the tower of Babel: a revised nomenclature for the study of suicide and suicidal behaviors part 2: suicide-related ideations, communications,and behaviors. *Suicide Life Threat Behav, 37*, 264 – 77.

Troister, T., Holden, R. R. (2010). Comparing psychache, depression, and hopelessness in their associations with suicidality: a test of Shneidman's theory of suicide. *Pers Individual Dif, 49,* 689‒93.

Valois, R. F., Zullig, K. J., Huebner, E. S., et al. (2004). Life satisfaction and suicide among high school adolescents. *Soc Indic Res, 66,* 81‒105.

Yip, P. S., Liu, K. Y., Hu, J., et al. (2005). Suicide rates in China during a decade of rapid social changes. *Soc Psychiatry Psychiatr Epidemiol, 40,* 792‒8.

Zhang, J., Wieczorek, W., Conwell, Y., et al. (2010). Characteristics of young rural Chinese suicides: a psychological autopsy study. *Psychol Med, 40,* 581‒9.

Cyberbullying and its risk factors among Chinese high school students

Zongkui Zhou [1]，Hanying Tang [1]，Yuan Tian [1]，Hua Wei [1]，Fengjuan Zhang [1]，Chelsey M. Morrison[2]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan, China；School of Psychology, Central China Normal University, Wuhan, China)

([2] Wheaton College, USA)

Abstract Cyberbullying has become a common occurrence among adolescents worldwide; however, it has yet to receive adequate scholarly attention in China, especially in the mainland. The present study investigated the epidemiological characteristics and risk factors of cyberbullying, utilizing a sample of 1,438 high school students from central China. Findings revealed that cyberbullying among high school students in the heartland of central China is relatively common with 34.84% (N=501) of participants reported having bullied someone and 56.88% (N=818) reported having been bullied by online. Significant gender differences were found, suggesting that boys are more likely to be involved in cyberbullying both as perpetrators and victims. Students with lower academic achievement were more likely to be perpetrators online than were students with better academic achievement. Students who spend more time on online, have access to the internet in their bedrooms, have themselves experienced traditional bullying as victims, and are frequently involved in instant−messaging and other forms of online entertainment are more likely to experience cyberbullying. Increased parent and teacher supervision reduced students' involvement in cyberbullying. Implications for intervention are explored.

Keywords Chinese high school students, cyberbullying, intervention, risk factors

1 Introduction

With the popularization of the internet and development of information and communication technology (ICT), online communication has become a common mode of communication. With this change, social phenomena existing off−line have begun to appear online. Cyberbullying is an example of this trend, which has received increasing attention because of its potentially serious consequences and increasing prevalence. However, relatively little is known about cyberbullying (O'Keeffe,

Clarke–Pearson, & Council on Communication and Media, 2011) especially in non–Western settings. Although there remains confusion about a standard definition for cyberbullying, most researchers agree that cyberbullying is an intentional, repeated, and aggressive act or behavior carried out by a group or individual instrumentally employing information and communication technology (ICT) (von Marees & Petermann, 2012). Studies show that the prevalence of cyberbullying is high in China; for example, research suggests that 34.9% of Chinese Taiwanese adolescents have been cyberbullied, and 20.4% had cyberbullied others (Huang & Chou, 2010; see Chen & Cheng [2013] for traditional bullying prevalence in a Taiwanese sample). Adolescents who have been bullied online are more likely to experience psychological problems (e.g. anxiety, depression) and engage in problem behaviors (e.g. skipping school, drug and alcohol use) (Beran & Li, 2005; Fosse & Holen, 2006; Juvonen & Gross, 2008; Mitchell, Ybarra, & Finkelhor, 2007; Wolak, Mitchell, & Finkelhor, 2006; Ybarra, Diener–West, & Leaf, 2007; Ybarra & Mitchell, 2007). In extreme cases, cybervictimization has been linked with suicide (Hinduja & Patchin, 2010). These factors have caused cyberbullying to attract attention as an important public health problem.

The majority of previous studies have focused on prevalence and types of cyberbullying. Studies that have systematically examined the risk factors for cyberbullying are rare and recent (e.g. Huang & Chou, 2010; Mishna, Khoury–Kassabri, Gadalla, & Daciuk, 2012; von Marees & Petermann, 2012). To date, most studies investigating risk factors for cyberbullying have focused on demographic variables(Li, 2006; Ortega, Elipe, Mora–Merchán, Calmaestra, & Vega, 2009; Wade & Beran, 2011; Wang, Iannotti, & Nansel, 2009), internet usage (Aricak et al., 2008; Erdur–Baker, 2010; Mesch, 2009; Navarro, Serna, Martínez, & Ruiz–Oliva, 2012; Smith et al., 2008; Topccu, Erdur–Baker, & Capa–Aydin, 2008; Twyman,

Saylor, Taylor, & Comeaux, 2010), and experiences of traditional bullying (Hinduja & Patchin, 2010; Kowalski, Morgan, & Limber, 2012; Li, 2008; Tokunaga, 2010; Wang, Iannotti, & Luk, 2012). There have been many contradictory findings, suggesting that more information is needed to clarify the relationship between these factors and cyberbullying. For example, some studies determined that boys are more likely to be involved in cyberbullying as perpetrators, whereas girls are more likely to be cybervictims (Li, 2006; Ortega et al., 2009; Wade & Beran, 2011; Wang, Iannotti, & Nansal, 2009). However, other studies show that boys are more likely to be involved in cyberbullying as *both* bullies and victims (Aricak et al., 2008; Nansel et al., 2001). In addition, Li (2007) found that students with lower academic status were likely to become cyberbullies, whereas Ma (2001) argued that these students were likely to become victims. It is essential to have a thorough understanding of the risk factors to inform intervention and prevention strategies.

To date, very little is known about cyberbullying among adolescents from the Mainland of China, as the majority of existing studies were conducted in Chinese Taipei (the capital of Taiwan, as it is known in the West) (e.g. Hokoda, Lu, & Angeles, 2006; Huang & Chou, 2010; Wei, Jonson–Reid, & Tsao, 2007). The China Internet Network Information Center (CINIC, 2012) reported that the total number of Chinese *netizens* (or 'internet citizens') reached 537.6 million in June 2012, with adolescents (10– to 19–years–old) accounting for 25.4%. The popularity of internet use among Chinese adolescents makes cyberbullying a social phenomenon deserving of attention.

Culture is a strong predictor for both cyberbullying and cybervictimization (Li, 2007, 2008). Huang and Chou (2010) found that Chinese Taiwanese students usually took no action after being victimized online because of a cultural imperative to avoid conflict so as to maintain group harmony; further, that study found no relationship

between cyberbullying and academic achievement. Cultural differences between the West and general Chinese culture may account for some of the differences observed in the empirical literature (e.g. frequency, attitude, setting, motive). Therefore, conclusions about cyberbullying from studies utilizing samples influenced by Western cultures may not be generalizable to Chinese culture. These findings are a reminder that it is essential to examine cultural factors related to the development, presentation, and intervention of cyberbullying.

Due to the limited number of studies on cyberbullying in mainland China, the present study aimed to clarify epidemiological characteristics and risk factors among Chinese mainland high school students with an emphasis on cultural differences between China and Western countries. We also sought to explore effective prevention and intervention measures of cyberbullying.

2 Methods

2.1 Participants

Participants were 1,438 students (42.56% female, 57.44% male, mean age=15.91, SD=1.02) from central China. The grade level of participants ranged from grad 10 to grade 12 (43.91%, 10th grade; 39.81%, 11th grade; 16.28%, 12th grade).

2.2 Measures

Students anonymously completed a survey which included questions on: Demographics; internet usage; Cyberbullying Inventory (CBI); traditional bullying scale; motivation for cyberbullying; and parents' and teachers' supervision of internet usage. With regard to internet usage, participants were asked how long they had been internet users; how many hours they spend online per week; what device they use to access the internet; and where they access the internet. They were also asked how often they engaged in each of the following internet activities: Instant messaging; visiting social networking sites; emailing; phone messaging; online entertainment (e.g. listening to streamed music, watching video); searching for information; playing online games; and online shopping. Answers were provided on a five-point scale ranging from 1=Never to 5=Always (where 1 to 3 was coded as low frequency; 4 to 5, high frequency). Finally, students were asked to indicate whether or not they had a mobile-phone and whether or not it could access the internet.

Students' experiences of cyberbullying were assessed by a Chinese-language questionnaire based on the Cyberbullying Inventory (CBI) (Erdur & Kavsut, 2007) which included two forms: *Cyberbullying* (CB) and *cybervictimization* (CV). Both forms consisted of 18 items that described different forms of CB or CV. Two items were added to the cybervictimization form so that both forms would have an equal number of items. Items described experiences such as sending/ receiving hurtful emails and making/receiving threats in a chat room. In this study, internal consistency of the CB and CV scales was 0.88 and 0.90, respectively.

The traditional bullying and victimization scale was developed by Li, Zhang, and Yu (2012). This scale consists of six items including, 'people make-up cruel nicknames about me', 'people scold me', or 'people tease and mock me', 'students hit, kick, punch or threaten me', and 'students spread rumors about me and try to make others not like me'. The inverse of each of these questions was asked to assess bullying behavior (i.e. 'I/we make-up cruel nicknames for other people', etc.). Each item was assessed using a five-point likert scale. In this study, internal consistency of traditional bullying and victimization was 0.80 and 0.69, respectively.

Eight major motives for cyberbullying were investigated, which had been identified from prior studies. Motives included 'for fun', 'to vent', 'because I dislike someone', 'for revenge', 'out of boredom', 'because it

looks cool', 'to attract someone's attention', and 'for other benefits'. Only participants who had already endorsed the statement that they had bullied others online were asked to answer this portion of the questionnaire. Participants could choose more than one motive based on their own experiences. Victims' reactions to cyberbullying were also collected using multiplechoice questions with several options, such as 'ignore/don't care', 'talk about the experience with someone for help', and 'seek revenge on people who hurt me'.

In order to examine the relationship between academic achievement and cyberbullying, we collected information on participants' academic achievement. Due to the difficulty involved in obtaining official academic records for every participant, academic achievement data was obtained only through self−report (a method that has been used by other researchers of Chinese cyberbullying; e.g. Huang & Chou, 2010). Students reported their academic achievement on a three−point scale (1='above average', 2='average', 3='below average').

To assess parental supervision and restriction of internet usage, three questions were asked: 'Do your parents supervise your online activities?', 'Do your parents control your online activities?', and 'Do your parents control your use of your mobile−phone at home?'. The same three questions were asked in reference to teachers' supervision. Additionally, two additional multiple−choice questions were asked to identify parents' methods for controlling children's online activities (e.g. install software to prevent access to some websites, install software to monitor online activities, check history of visited sites, control/limit time online).

2.3 Procedure

Data were collected during Summer 2012 in classroom−settings by trained graduate students and teachers. Information about the purposes of the survey and confidentiality of responses was clearly explained and assent was obtained. Students were informed that there was no right or wrong answer. Extra support was provided for students who had difficulty completing the surveys.

3 Results

3.1 Prevalence rate of cyberbullying

In this sample, 34.84% (N=501) of participants reported that they had bullied someone online and 56.88% (N=818) reported that they had been bullied by someone online within the last semester (one semester continues for about five months in China); in addition, 26.84% (N=386) reported being *both* cyberbullies and cybervictims. A total of 37.34% (N=537) of participants reported that someone in their class had cyberbullied, and 40.33% (N=580) reported that someone in their class had been cybervictimized.

Our study mirrored previous studies that found varying prevalence rates among different types of cyberbullying. Specifically, 'to kick out someone from a chat room' (17.94%), 'to insult someone in a chat room' (12.80%), and 'to spread private information discussed by instant messaging tools (e.g. QQ/MSN)' (7.79%) were the three most frequently reported cyberbullying behaviors in our study. 'Someone stole my passwords of my ICM so that I cannot use them anymore' (24.13%), 'someone kicked me out from a chat room' (22.81%), and 'someone stole my network account and accessed my personal information' (15.09%) were the three most frequently reported forms of cybervictimization.

3.2 Gender differences

Our study found that boys (39.59%) were more likely to report that they were involved in cyberbullying as perpetrators than girls (28.43%), χ^2 (1, N=1438)=19.28, p<0.0001; boys (59.81%) were also more likely to be cybervictims than girls (52.94%), χ^2 (1, N=1438)=6.76, p<0.01. Regarding the relationship between gender and

Table 1. Comparisons of the gender differences on different forms of cyberbullying

Behavior	As a perpetrator			As a victim		
	Boys N=826	Girls N=612	χ^2 df=1	Boys N=826	Girls N=612	χ^2 df=1
Send hurtful e-mails	3.1%	1.0%	7.59**	12.7%	9.6%	3.28
Hiding the name via sending SMS	2.2%	0.8%	4.14	6.8%	5.1%	1.82
Take embarrassing photos	9.1%	4.2%	12.57***	14.6%	8.0%	14.88***
Threaten in chat room	3.5%	1.0%	9.48**	6.3%	4.2%	2.87
Spread rumors	4.2%	1.8%	6.76**	13.3%	9.6%	4.58*
Spread information on SMS	8.0%	7.5%	0.11	10.5%	10.3%	0.02
Kick out from chat room	18.9%	16.7%	1.18	25.5%	19.1%	8.25**
Humiliate by using fake photos	6.7%	1.6%	20.56***	6.9%	3.1%	10.12**
Send hurtful SMS	5.2%	2.6%	6.00*	6.4%	4.7%	1.84
Insult in chat room	17.1%	7.0%	31.78***	16.5%	6.7%	31.06***
Get information without permission	2.4%	0.7%	6.69**	13.6%	12.4%	0.40
Block access by stealing password	2.4%	0.3%	10.24**	11.4%	6.9%	8.38**
Reach messages by stealing passwords	1.5%	0.3%	4.62*	10.0%	6.9%	4.50*
Block from using SMS	2.3%	0.3%	9.51**	26.0%	21.6%	3.82
Use other's username without permission	1.6%	0.2%	7.25**	18.8%	10.1%	20.45***
Violate privacy via webcam	1.1%	0.3%	2.70	3.4%	2.1%	2.03
Humiliate by create Web pages	1.1%	0.2%	4.37	2.7%	0.5%	9.72*
Harm someone known from Internet	5.3%	2.0%	10.64**	7.7%	6.7%	0.57

Note: *** p<0.001, ** p<0.01, * p<0.05

types of cyberbullying or cybervictimization, more boys than girls reported cyberbullying someone using each type of bullying behavior measured with the exception of five types (e.g. 'To spread the information without permission', 'To kick out someone from a chat room', and 'To harm someone you know from the internet') (See Table 1). However, male-cybervictims reported experiencing only about half of all 18 types; these results suggest that the gender differences are weaker among cybervictims. In addition, we also found that' to kick someone from a chat room' was the most frequent bullying behavior perpetrated by both boys and girls online, and 'someone stole my passwords of my ICM so that I cannot use them anymore' was the most frequent type of victimization experienced by both boys and girls online.

3.3 Academic achievement differences

The relationship between academic achievement difference and participants' experience of cyberbullying was examined using a one-way ANOVA analysis. As presented in Table 2, there were significant differences found between the three academic achievement groups (F=3.89, p<0.05). Multiple comparisons indicate that compared to students who perform better academically, the students with lower academic achievement were more likely to be online perpetrators. However, there were no differences found between the three academic-achievement groups and experiences of cybervictimization. Therefore, lower academic achievement maybe a potential risk factor for those engaging in cyberbullying as perpetrators.

3.4 Internet usage and cyberbullying

Descriptive data of internet usage was collected. The mean age of when the participants originally gained access to the internet was 5.60 years (SD=2.58). Students reported

Table 2. Cyberbullying experience scores of the three academic-achievement groups

	Above average N=756		Average N=383		Below average N=253		
Cyberbullying experience	Mean	SD	Mean	SD	Mean	SD	F
Cybervictims	1.12	0.23	1.16	0.35	1.14	0.25	2.66
Cyberbullies	1.03	0.12	1.07	0.20	1.07	0.23	3.89*

Note: * p<0.05

spending time accessing the online world between one and two days each week (1.87 days; SD=1.77), with 2.13 (SD=2.06) hours per day, on average. This means that on average, students spend about 4.82 hours per week on the internet. Correlation analysis showed that both cyberbullying and cybervictimization were significantly related to total time spent online each week (r=0.22. p<0.01; r=0.19, p<0.01). Time spent online appears to be an important factor for predicting cyberbullying.

For technology use, 71.42% of participants (N=1027) reported having a mobile−phone, and among them 76.73% (N=788) reported that they could access the internet by phone. More than four−in−five of the youth (84.77%) reported having access to the internet by computer, 58.97% used a mobilephone, and 16.41% accessed via tablet/pad computer. Additionally, 83.80% of participants reported having access to the internet at home; 17.45% reported accessing the internet at a commercial setting (e.g. 'coffee−shop'), while 13.70% reported access at school and 9.39% at the home of friends. (As participants could choose more than one option, the totals reported here do not equal 100%).

When asked how often they were involved in any of several major online activities using a five−point scale (from1=never to 5=always), results revealed that participants often spent their time on searching for information (M=3.84), communicating via instant messages (M=3.79), and online entertainment (M=3.61). They were less frequently involved in social networking sites (M=3.13), phone messaging (M=2.93), and online game playing (M=2.55). They were rarely involved in sending emails (M=1.95) or shopping online (M=1.74). Further analyses were run in order to gain a deeper understanding of the relationship between the three online activities most frequently endorsed by participants and their experience of cyberbullying. Students were classified into two groups: Students who endorsed 'often' and 'always' on these items

were classified as Group 1 (often involved in), and all others were classified as Group 2(rarely involved in). The results of t−test showed that students who extensively used instant messaging were more likely to be experienced both as cyberbullies and cybervictims (t=3.46, p<0.001; t=3.73, p<0.0001) and that students who were often involved in online entertainment were more likely to experience cybervictimization (t=2.21, p<0.05).

3.5 Parents and cyberbullying

With regard to parental supervision, 45.83% of the participants reported having their internet usage monitored by parents. The extent of parental restriction was measured on a five−point scale, showing that only 8.41% of parents fail to restrict their child's internet usage, and 73.50% of parents restrict children's internet usage at least a moderate level. T−tests examined the differences in parental restriction between perpetrators and non−perpetrators, as well as victims and non−victims. Non−perpetrators reported that their parents were more restrictive of their internet usage than were perpetrators' parents (t=2.65, p<0.05). No significant differences in parental restriction were found between victims and non−victims (t=0.71, p>0.05).

With regard to specific restriction strategies, 71.77% of youth reported that their parents implemented rules to limit the amount of time they are allowed to spend online, 7.79% reported their parents had installed filtering software to block specific web sites, 6.61% reported that their parents check their browsing history, and 2.36% reported that their parents had installed monitoring software to record online activities.

Participants were asked about the location where they accessed the internet at home: 544 participants (37.83%) reported having access to the internet in their own bedroom and 42.91% of youth reported only having internet access in common areas of their house. Chi square analyses showed that both cyberbullies x^2 (1, N=544)=21.44;

$p<0.0001$) and cybervictims (x^2 (1, N=544)=36.03, $p<0.0001$) were more likely to have access to the internet in a private space at home than were non−cyberbullies and noncybervictims.

3.6 Educators and cyberbullying

This survey also showed that 70.45% of participants reported being monitored by teachers when using the internet at school—suggesting that Chinese teachers have a cautious attitude about students' online activities. The results of t−tests indicate that both perpetrators and victims reported less restriction by teachers while engaging in online activities at school than non−perpetrators and non−victims (t=−2.31, $p<0.05$; t−=2.06, $p<0.05$).

There are limited opportunities for high school students in China to access the internet using a computer while at school; however, as 58.97% of the participants reported having internet access using their mobile−phones, it can be speculated that students use their phones to access the internet while at school. Therefore, we further examined the relationship between teachers' restriction on students' phone usage and students' experiences of cyberbullying. Results indicated that perpetrators reported their mobile−

phone use as being less restricted by teachers than non− perpetrators (t=−2.03, $p<0.05$). However, there was no significant difference in teachers' restrictions of phone use between victims and non−victims.

3.7 Traditional bullying and cyberbullying

A significant positive correlation was found between traditional bullying and cyberbullying (see Table 3). To further examine the relationship between traditional and cyberbullying forms, we analysed the overlap between traditional and cyberbullying using the method utilized by Kowalski, Morgan, and Limber (2012). Participants were classified into four groups: Victims only; bullies only; bully/victims; and neither bullies nor victims. Separate categories were created for cyberbullying and traditional bullying. Table 4 presents the overlap between involvement in traditional bullying and involvement in cyberbullying. Participants who were cyberbullies/victims were likely to be involved in traditional bullying as both perpetrators (53.11%) and as victims (68.13%). Similarly, individuals who were bully/victims in traditional bullying were likely to be both bullies (50.49%) and victims (76.83%) online. Individuals who were cybervictims were

Table 3. Correlation matrix between traditional and cyberbullying

	Traditional victimization	Traditional perpetration	Cyber victimization	Cyber perpetration
Traditional victimization	1			
Traditional perpetration	0.48**	1		
Cyber−victimization	0.35**	0.35**	1	
Cyber−perpetration	0.29**	0.44**	0.58**	1
Mean	1.22	1.13	1.13	1.06
SD	0.36	0.31	0.27	0.17

Note: ** $p<0.01$

Table 4. Overlap between traditional and cyberbullying

Cyberbullying status	Traditional victims	Traditional bullies	All participants
Victim	233(53.94%)	146(33.80%)	432(30.04%)
Bully	47(40.87%)	28(24.35%)	115(8.00%)
Bully/Victim	263(68.13%)	205(53.11%)	386(26.84%)
Neither	160(32.32%)	95(19.19%)	495(34.42%)
Traditional status	Cyber victims	Cyber bullies	All participants
Victim	179(61.72%)	101(34.83%)	290(20.17%)
Bully	34(56.67%)	24(40.00%)	60(4.17%)
Bully/Victim	315(76.83%)	207(50.49%)	410(28.51%)
Neither	258(41.41%)	150(24.08%)	623(43.32%)

also likely to be victims of traditional bullying (53.94%). Both bullies (56.67%) and victims (61.72%) of traditional bullying were themselves likely to be victims online. This provides supporting evidence for the predictive effects of traditional bullying on cyberbullying.

3.8 Motives for cyberbullying

We also explored motives for cyberbullying. Based on the previous literature, several types of motives were investigated (Kowalski, Limber, & Agatston, 2012). A frequency distribution showed the following motives as most commonly reported by cyberbullies (N=501): 'I dislike someone' (29.14%); 'for fun' (23.95%); 'out of boredom' (19.16%); 'to vent' (15.77%); 'to get revenge' (7.58%);'to conform/fit in' (5.59%), 'to attract his/her attention' (3.19%); 'it looks cool' (1.60%); and 'to get some other benefit' (1.60%).

3.9 Students´ reactions to cybervictimization

Students may react to cybervictimization in different ways. This survey showed that the most frequent reaction of participants who had experienced cybervictimization was to 'ignore/not react' (45.84%) followed by 'talking about the experience with someone for help' (35.57%). Other outcomes less frequently cited were 'delete the materials which may hurt me' (32.27%), 'change my online account' (24.69%), and'seek revenge on people who hurt me online' (11.86%). Participants who reported'talking about the experience with someone for help' most frequently wanted to talk with classmates/friends (65.64%), followed by parents (28.87%) and siblings (27.84%). Only 2.75% reported they would talk with their teachers for help.

4 Discussion and implications

There are few studies examining the prevalence of cyberbullying in China, especially in the mainland. Li (2008) utilized a study of 197 Chinese students and found that 33% were cybervictims and 7% were cyberbullies.

Another study from Chinese Taipei showed that 34.9% of 545 participants had been cyberbullied, 20.4% had cyberbullied others, and 63.4% reported having witnessed or being aware of cyberbullying (Huang & Chou, 2010). As the majority of the students who participated in these studies were from middle schools, the prevalence rate of cyberbullying among Chinese high school students remains unknown. The results of the present study provide preliminary data indicating that high school students in mainland China are frequently involved in cyberbullying. Studies from Western countries have examined the prevalence rate of cyberbullying among high school students with varying results (von Marees & Petermann, 2012). Based on his review of 14 studies conducted on cyberbullying in Australia, the USA, the UK, and Canada, Kraft (2006) summarized that reported levels of cybervictimization varied between 10% and 42%, and that rates of cyberbullying varied from 6% to 33% (with 11.5% recently confirmed in Australia by Sakellariou, Carroll, & Houghton, 2012).

Many researchers have argued that these discrepancies may be due to differing definitions of cyberbullying utilized by assessment instruments, the age−range of participants, and the timeframe of participants' response (e.g. Kowalski, Limber, & Agatston, 2012). Some researchers have posited that involvement in cyberbullying as perpetrators or victims increases from the age 10− to 16−years−old (von Marees & Petermann, 2012). In the present study, the mean age of participants is 15.93 (SD=1.02), which is very close to 16−years−old. In accordance with the age trend mentioned above, our participants may be more likely to be involved in cyberbullying. Nevertheless, the prevalence rates of cyberbullying revealed in our research are higher than the upper limits summarized by Kraft (2006). Therefore, this suggests that cyberbullying among high school students in China is relatively common and should attract the attention of parents, educators, and public society.

Consistent with the findings of previous studies from Western countries, the most common venues for victimization by cyberbully perpetrators were chat rooms (18.0%) and instant messaging (8.5%) (Hinduja & Patchin, 2008; Kowalski & Limber, 2007; Patchin & Hinduja, 2006). Within these venues, 23.9% of the respondents reported having their ICM passwords stolen and 22.8% reported having been kicked out from chat rooms. This study also found that instant messaging (e.g. QQ/MSN) was one of the most common activities that students participated in online. Among those who were bullied online (*N*=818), 41.08% claimed that they were aware of the bully's identity. More than one-half (55.06%) of these students reported that the perpetrators were their classmates. This finding may provide support for existing studies that have investigated the anonymity of cyberbullying (Huang & Chou, 2010; Juvonen & Gross, 2008). The proportion of classmates being perpetrators should attract educators' attention in framing prevention and intervention activities.

4.1 Risk factors for cyberbullying among high school students

With regard to gender differences, we found that boys were significantly more likely to be cyberbullies or cybervictims. Although these results are consistent with findings from the traditional bullying literature, they are not aligned with results from early research on cyberbullying. Previous studies have found that girls were more likely to be cybervictims (Smith et al., 2008; Wang et al., 2009) or have found no significant gender differences between cybervictims (Hinduja & Patchin, 2008; Slonje & Smith, 2008; Williams & Guerra, 2007). Our results are supported by the work of Olweus (2003). In a sample of Chinese adolescents from Taipei, Huang and Chou's (2010) findings mirror our own, with male students reporting greater levels of both perpetration and victimization experiences than females. This suggests that these findings

may be due to cultural differences. In traditional Chinese culture, girls are raised to be gentle, polite and kind, while boys are encouraged to be active, brave, and independent. Boys are told that it is not brave to be aggressive towards or even bully girls. This may lead to fewer girls being involved in bullying, whether offline or online.

Our results confirmed that the frequency of internet access may be another risk factor for cyberbullying. Our results are consistent with finings from previous that youth who use the internet more frequently and spend more time online per day may be more likely to become involved in cyberbullying as perpetrators or victims (Mishna et al., 2012; Navarro et al., 2012; Wolak, Mitchell, & Finkelhor, 2007). Chi square analyses showed that having internet access by phone or accessing the internet at a commercial location increased student's risk of involvement in cyberbullying. Descriptive results showed that almost three-quarters of the youth had a phone, and nearly one-fifth of the youth had access to the internet commercially. This implies that youth are more likely to become involved in cyberbullying in unsupervised spaces. As more youth have access to the internet using mobile-phones, additional attention is warranted regarding youths' phone usage. Our results indicate that some types of online activities increase the odds of involvement in cyberbullying. Specifically, the more often students are involved in instant messaging, online entertainment, and information searches, the more likely they are to become involved in cyberbullying. Instant messaging may be an activity in which Chinese high school students are the most likely to experience cyberbullying (Huang & Chou, 2010). In the USA, socialnetwork sites and chat rooms have served as fertile ground for cyberbullying (Mesch, 2009). Although social networking sites are not yet as popular in China as in Western countries, nearly half of Chinese youth (42.98%) report frequently visiting social networking sites. This number is expected to increase which may

expose youth to additional risks.

Our study revealed a relationship between traditional bullying and cyberbullying which mirrors the findings of previous studies (Hinduja & Patchin, 2010; Kowaski, Morgan, & Limber, 2012; Li, 2007). Students who report involvement in traditional bullying were found to be at a greater risk for involvement with cyberbullying. We speculate three possible reasons for this result. First, previous studies have suggested traditional bullying and victimization are related to personal traits. Specifically, traditional bullies are typically emotionally impulsive, irritable, and lacking in self−control; in contrast, traditional victims are typically introverted, sensitive, and easily restrained (Zhang, Gu, & Ju, 2001). We speculate that students with these personal traits are also likely to experience bullying online. Second, because bullies and victims often know each other in real life, the social interaction that usually occurs at school may be extended online, along with their status of bully or victim (Kowaski, Morgan, & Limber, 2012). Third, because of the power imbalance between perpetrators and victims, victims may be unable to fight back or transform his/her role in traditional bullying. However, the anonymity of the internet makes this transformation easier and may give traditional victims the courage to counterattack. Therefore, the victims of traditional bullying may become cyberbullies (40.85% in our study); consequently, traditional bullies transition into cybervictims (61.80% in our study). This finding sheds light on the similarities between traditional and cyberbullying while highlighting one of the key differences. Therefore, we contend that there should be a different approach to prevention and intervention strategies for traditional bullying and cyberbullying.

The present study revealed that a relatively high proportion of Chinese parents (73.50%) place at least a moderate level of restriction on their children's internet usage. This suggests that Chinese parents are broadly aware of the risk involved in their children's online activities. Consistent with previous studies (Mesch, 2009; Navarro et al., 2012), our study found that parental restriction reduces the risk of children's involvement in cyberbullying. However, parental restriction was not found to be significantly related to children's experience of cybervictimization. This suggests that although parental restriction of children's internet usage may effectively reduce the risk of children perpetrating online, it may not protect children from being bullied online. This may be due to the unique feature of cyberbullying—that is, unlike traditional bullying, cyberbullying can occur anytime, anywhere (Li, 2008; Tokunaga, 2010). Thus, even though someone may rarely access the internet, he/she may still be vulnerable to online victimization.

The most important agenda for high school students in China is to prepare for their college entrance examinations. In order to achieve a high score in this critical watershed, they must spend almost all of their spare time studying, leaving little time for leisure activities. It is typical for Chinese high school students to spend more than ten hours daily in study. Going online while at school is virtually impossible; our survey revealed that 70.45% of participants reported that their online activities were closely monitored by teachers at school. It is logical to conclude that teacher−supervision and restrictions on students' online behaviors can reduce cyberbullying. While restricting students' computer use at school may be relatively easy, restricting their access to the internet using portable devices may is more complicated. We conclude that teachers' interventions for cyberbullying should include targeted measures to guide students' usage of mobile devices.

4.2 Implications for students

As adolescents are the direct participants of cyberbullying, prevention, and intervention efforts should help them better understand cyberbullying. Our survey

indicated that more than two−thirds (69.14%) participants believe that the harm brought to others through cybervictimization is only 'moderate' or 'minimal'. Further, the two most common motives for cyberbullying were 'for fun' and 'out of boredom', indicating that adolescents know little about the seriousness of their online behaviors. Therefore, it is not surprising that the prevalence of perpetrating reported by the participants was relatively high (34.84%). We speculate that if adolescents better understood the potential consequences of their online behaviors, the prevalence of perpetration would decrease.

4.3 Implications for parents

Our results indicate that while three−quarters of parents limit students' time online, relatively few use technological strategies such as installing online−filters or monitoring−software. Though this strategy did have some positive effects preventing cyberbullying among Chinese high school students, it was rudimentary and limited by comparison to Western countries. For example, Mesch (2009) investigated 935 US youth and found that 56% reported that their parents had installed filtering software and had rules on the type of information they were allowed to share over the internet. In addition to direct restriction on internet usage, other studies describe group interventions which involve educating students and creating rules about what personal information is appropriate to share (Navarro et al., 2012). In our study, 83.80% of participants reported having internet access at home. Further, youth who had internet access in the private space of their bedrooms were more likely to be both cyberbullies and cybervictims. In addition to controlling children's time online, Chinese parents may more effective if they intervened by placing the computer in relatively public spaces.

Chinese parents need to better understand cyberbullying; as they gain understanding they can help their children become more aware of the possible negative consequences of online activities. Our study indicated that one−in−three cyberbullied students were willing to talk about their experiences with parents—suggesting that parents could be important supports for children who experience cybervictimization; parents should know how to recognize the signs of cyberbullying and hat to do when they suspect cyberbullying.

4.4 Implications for educators

Supervision and restriction on students' online behaviors in school can effectively reduce cyberbullying (Cassidy, Brown, & Jackson, 2012); however, as mobile devices with internet access become increasingly available, supervision and restriction on students' online access may not be sufficient. The internet is an important part of modern everyday lives; even without computers, youth can go online by using portable devices. The most effective intervention for cyberbullying may be proper guidance. Specifically, schools should re−evaluate their methods of supervision and equip educators to better understand cyberbullying. Although only 4.90% of the cybervictims reported they would talk to their teachers about their experiences, this does not excuse educators from the role they should play to intervene in cyberbullying.

References

Aricak, T., Siyahhan, S., Uzunhasanoglu, A., Saribeyoglu, S., Ciplak, S., Yilmaz, N., & Memmedov, C. (2008). Cyberbullying among Turkish adolescents. *Cyberpsychology and Behavior, 11*(3), 253‒261. doi:10.1089/cpb.2007.0016.

Beran, T., & Li, Q. (2005). Cyber−harassment: A study of a new method for an old behaviour. *Journal of Educational Computing Research, 32*(3), 265‒277.

Cassidy, W., Brown, K., & Jackson, M. (2012). 'Under the radar': Educators and cyberbullying in schools. *School Psychology International, 33*(5), 520‒532.

doi:10.1177/0143034312445245.

Chen, M. L., & Cheng, Y (2013). Prevalence of school bullying among secondary students in Taiwan: Measurements with and without a specific definition of bullying. *School Psychology International*. doi:10.1177/0143034313479694 (in press).

China Internet Network Information Center (CINIC) (2012). 30th China Internet Development Statistics Report. Retrieved August 25, 2012, from http://www.cnnic.cn/ hlwfzyj/hlwxzbg/hlwtjbg/201207/P020120723477451202474.pdf.

Erdur–Baker, O ., & Kavsut, F. (2007). Cyber bullying: a new face of peer bullying. *Eurasian Journal of Educational Research, 27*: 31–42.

Erdur–Baker, O. (2010). Cyberbullying and its correlation to traditional bullying, gender and frequent and risky usage of internetmediated communication tools. *New Media Society, 12*(1), 109–125. doi:10.1177/1461444809341260.

Fosse, G. K., & Holen, A. (2006). Childhood maltreatment in adult female psychiatric outpatients with eating disorders. *Eating Behaviours, 7*(4), 404–409. doi:10.1016/ j.eatbeh.2005.12.006.

Hinduja, S., & Patchin, J. W. (2008). Cyberbullying: An exploratory analysis of factors related to offending and victimization. *Deviant Behavior, 29*(2), 129–156. doi:10.1080/01639620701457816.

Hinduja, S., & Patchin, J. W. (2010). Bullying, cyberbullying, and suicide. *Archives of Suicide Research, 14*(3), 206–221. doi:10.1080/13811118.20 1.494133.

Hokoda, A., Lu, H.–H. A., & Angeles, M. (2006). School bullying in Taiwanese adolescents. *Journal of Emotional Abuse, 6*(4), 70–90. doi:10.1300/ J135v06n04_04.

Huang, Y. Y., & Chou, C. (2010). An analysis of multiple factors of cyberbullying among junior high school students in Taiwan. *Computers in Human Behavior, 26*(6), 1581–1590. doi:10.1016/j.chb.2010.06.005.

Juvonen, J., & Gross, E. F. (2008). Extending the school grounds? Bullying experiences in cyberspace. J*ournal of School Health, 78*(9), 496–505. doi:10.1111/ j.1746–1561.2008.00335.x.

Kowalski, R. M., & Limber, S. P. (2007). Electronic bullying among middle school students. *Journal of Adolescent Health, 41*(6), 22–30. doi:10.1016/ j.jadohealth.2007.08.017.

Kowalski, R. M., Limber, S. P., & Agatston, P. W. (2012). Cyberbullying: *Bullying in the digital age*. Chicester: Wiley–Blackwell.

Kowalski, R. M., Morgan, C. A., & Limber, S. P. (2012). Traditional bullying as a potential warning sign of cyberbullying. *School Psychology International*, *33*(5), 505–519. doi:10.1177/0143034312445244.

Kraft, E. (2006). Cyberbullying: A worldwide trend of misusing technology to harass others. *WIT Transactions on Information and Communication Technologies*, *36*, 155–166. doi:10.2495/IS060161.

Li, H. L., Zhang, W. X., & Yu, F. J. (2012). The relationship between victimization and depression of adolescents. *Psychological Development and Education, 1*, 77–82. doi:CNKI:SUN:XL FZ.0.2012–01–011.

Li, Q. (2006). Cyberbullying in schools: A research of gender differences. *School Psychology International, 27*(2), 157–170. doi:10.1177/0143034306064547.

Li, Q. (2007). New bottle but old wine: A research of cyberbullying in schools. *Computers in Human Behavior, 23*(4), 1777–1791. doi:10.1016/ j.chb.2005.10.005.

Li, Q. (2008). A cross–cultural comparison of adolescents' experience related to cyberbullying. *Educational Research, 50*(3), 223–234. doi:10.1080/00131880802309333.

Ma, X. (2001). Bullying and being bullied: To what extent are bullies also victims? *American Educational Research Journal, 38*(2), 351 – 370. doi:10.3102/00028312038002351.

Mesch, G. S. (2009). Parental mediation, online activities, and cyberbullying. *CyberPsychology and Behavior, 12*(4), 387 – 393. doi:10.1089/cpb.2009.0068.

Mishna, F., Khoury–Kassabri, M., Gadalla, T., & Daciuk, J. (2012). Risk factors for involvement in cyber bullying: Victims, bullies and bully–victim. *Children and Youth Services Review, 34*(1), 63 – 70. doi:10.1016/j.childyouth.2011.08.032.

Mitchell, K. J., Ybarra, M., & Finkelhor, D. (2007). The relative importance of online victimization in understanding depression, delinquency, and substance use. *Child Maltreatment, 12*(4), 314 – 324. doi:10.1177/1077559507305996.

Nansel, T., Overpeck, M., Pilla, R., Ruan, W., Simons–Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment. *The Journal of the American Medical Association, 285*(16), 2094 – 2100. doi:10.1001/jama.285.16.2094.

Navarro, R., Serna, C., Martnez, V., & Ruiz–Oliva, R. (2012). The role of Internet use and parental mediation on cyberbullying victimization among Spanish children from rural public schools. *European Journal of Psychology of Education,* 1 – 21. doi:10.1007/s10212–012–0137–2.

O'Keeffe, G. S., Clarke–Pearson, K., & Council on Communication and Media. (2011). Clinical report—The impact of social media on children, adolescents, and families. *Pediatrics, 127*(4), 800 – 805. doi:10.1542/peds.2011–0054.

Olweus, D. (2003). A proEle of bullying at school. *Educational Leadership, 60*(6), 12 – 17.

Ortega, R., Elipe, P., Mora–Merchan, J. A., Calmaestra, J., & Vega, E. (2009). The emotional impact on victims of traditional bullying and cyberbullying. *Zeitschrift fu´r Psychologie/Journal of Psychology, 217*(4), 197 – 204. doi:10.1027/0044–3409.217.4.197.

Patchin, J. W., & Hinduja, S. (2006). Bullies move beyond the school yard: *A preliminary look at cyberbullying. Youth Violence and Juvenile Justice, 4*(2), 148 – 169, 376 – 385. doi:10.1177/1541204006286288.

Sakellariou, T., Carroll, A., & Houghton, S. (2012). Rates of cyber victimization and bullying among male Australian primary and high school students. *School Psychology International, 33*, 533 – 549. doi:10.1177/0143034311430374.

Slonje,R.,&Smith,P.K.(2008).Cyberbullying:Another maintypeofbullying?*Scandinavian Journal of Psychology, 49*(2), 147 – 154. doi:10.1111/j.1467–9450.2007.00611.x.

Smith, P. K., Mahdavi, J., Carvalho, M., Fisher, S., Russell, S., & Tippett, N. (2008). Cyberbullying: Its nature and impact in secondary school pupils. *Journal of Child Psychology and Psychiatry, 49*(4), 376 – 385. doi:10.1111/j.1469–7610.2007.01846.x.

Tokunaga, R. S. (2010). Following you home from school: A critical review and synthesis of research on cyberbullying victimization. *Computers in Human Behavior, 26*(3), 277 – 287. doi:10.1016/j.chb.2009.11.014.

Topcu, C., Erdur–Baker, O., & Capa–Aydin, Y. (2008). Examination of cyberbullying experiences among Turkish students from different school types. *CyberPsychology and Behavior, 11*(6), 643 – 648. doi:10.1089/cpb.2007.0161.

Twyman, K., Saylor, C., Taylor, L. A., & Comeaux, C. (2010). Comparing children and adolescents engaged in cyberbullying to matched peers. Cyberpsychology, *Behavior, and Social Networking, 13*(2), 195 – 199. doi:10.1089/cyber.2009.0137.

Von Marees, & Petermann. (2012). Cyberbullying: An increasing challenge for schools. *School Psychology International, 33*(5), 467–476. doi:10.1177/0143034312445241.

Wade, A., & Beran, T. (2011). Cyberbullying: The new era of bullying. *Canadian Journal of School Psychology, 26*(1), 44–61. doi:10.1177/0829573510396318.

Wang, J., Iannotti, R. J., & Luk, J. W. (2012). Patterns of adolescent bullying behaviors: Physical, verbal, exclusion, rumor, and cyber. *Journal of School Psychology, 50*(4), 521–534. doi:10.1016/j.jsp.2012.03.004.

Wang, J., Iannotti, R. J., & Nansel, T. R. (2009). School bullying among adolescents in the United States: Physical, verbal, relational, and cyber. *Journal of Adolescent Health, 45*(4), 368–375. doi:10.1016/j.jadohealth.2009.03.021.

Wei, H.-S., Jonson-Reid, M., & Tsao, H. L. (2007). Bullying and victimization among Taiwanese 7th graders: A multi-method assessment. *School Psychology International, 28*(4), 479–500. doi:10.1177/0143034307084137.

Williams, K. R., & Guerra, N. G. (2007). Prevalence and predictors of Internet bullying. *The Journal of Adolescent Health, 41*, S14–S21. doi:10.1016/j.jadohealth.2007.08.018.

Wolak, J., Mitchell, K., & Finkelhor, D. (2006). Online victimization of youth: 5 years later. *National Center for Missing and Exploited Children, 7*, 06–025.

Wolak, J., Mitchell, K. J., & Finkelhor, D. (2007). Does online harassment constitute bullying? An exploration of online harassment by known peers and online-only contacts. *Journal of Adolescent Health, 41*(6), S51–S58. doi:10.1016/j.jadohealth.2007.08.019.

Ybarra, M. L., Diener-West, M., & Leaf, P. J. (2007). Examining the overlap in Internet harassment and school bullying: Implications for school intervention. *Journal of Adolescent Health, 41*(6), 42–50. doi:10.1016/j.jadohealth.2007.09.004.

Ybarra, M. L., & Mitchell, K. J. (2007). Prevalence and frequency of Internet harassment instigation: Implications for adolescent health. *Journal of Adolescent Health, 41*(2), 189–195. doi:10.1016/j.jadohealth.2007.03.005.

Zhang, W. X., Gu, C. H., & Ju, Y. C. (2001). Review of the relationship between personality and bullying in children. *Advances in Psychological Science, 9*(3), 215–220.

Friendship quality, social preference, proximity prestige, and self-perceived social competence: Interactive influences on children's loneliness

Fengjuan Zhang[1,2,3], Zhiqi You[1,2,4,5], Cuiying Fan[1,2], Chuang Gao[1,2] , Robert Cohen[6], Yeh Hsueh[7], Zongkui Zhou[1,2,*]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan 430079, China)

([2] School of Psychology, Central China Normal University, Wuhan 430079, China)

([3] Institute of Complexity Science and Big Data Technology, Guangxi University, Nanning 530004, China)

([4] Department of Sociology, Huazhong Agricultural University, Wuhan 430070, China)

([5] Research Center For Rural Social Construction and Management, Wuhan 430070, China)

([6] Department of Psychology, University of Memphis, Memphis, TN 38152, USA)

([7] College of Education, Health and Human Sciences, University of Memphis, Memphis, TN 38152, USA)

Abstract The purpose of this study was to test an integrative model in which peer relations at different levels of social complexity (friendship quality, social preference, and proximity prestige) are associated with children's loneliness, with children's self－perceived social competence acting as a mediator of these associations. A middle childhood sample of 509 Chinese children (233 girls and 276 boys; 3rd to 6th grade) completed a battery of sociometric and self－report questionnaires. Bootstrap analysis showed that self－perceived social competence mediated the relations between each peer variable and loneliness. In the integrative model tested with SEM, the mediating effect of self－perceived social competence in the relation between friendship quality and loneliness and between social preference and loneliness remained significant. However, self－perceived social competence no longer mediated the association between proximity prestige and loneliness, when considering the simultaneous influences of the three peer variables (friendship quality, social preference, and proximity prestige). The whole model accounted for 56% of the variance in loneliness. These findings suggest that self－perceived social competence played an important role in children's loneliness, that the quality and the quantity of direct peer relations (friendship quality, social preference, and part of proximity prestige) were associated with loneliness, and that indirect friends had a relatively lower but significant influence on children's loneliness. The results are discussed in terms of their implications for preventing children's loneliness.

Keywords friendship quality; social preference; proximity prestige; self－perceived social competence; loneliness

*Corresponding author at: Key Laboratory of Adolescent Cyberpsychology and Behavior(CCNU), Ministry of Education, Wuhan 430079, China. School of Psychology, Central China Normal University, Wuhan 430079, China. Tel./fax: +86 2767868632. E－mail address: zhouzk@mail.ccnu. edu.cn (Z. Zhou).

1 Introduction

As a social species, human beings innately have the need for social connection. An absolute or relative lack of social connection can result in the painful emotion of loneliness (Peplau & Perlman, 1982; Weiss, 1973, p. 15). Loneliness is a common emotional experience among children. As many as 80% of children report having experienced loneliness at school (Berguno, Leroux, McAinsh, & Shaikh, 2004). One study of third-grade children found that 23% had a moderate level of loneliness, a feeling that steadily increased during the next two years (Jobe-Shields, Cohen, & Parra, 2011). Substantial evidences show that the feeling of loneliness in childhood not only is associated with children's current life quality but also predicts future maladjustment (Masi, Chen, Hawkley, & Cacioppo, 2011; Rotenberg, 1999; van Dulmen & Goossens, 2013). Lonely children are more likely to experience low self-esteem (Sletta, Valås, Skaalvik, & Sbstad, 1996), increased levels of social anxiety and social avoidance (Vanhalst, Goossens, Luyckx, Scholte, & Engels, 2012), poorer academic performance (Benner, 2011), higher risk of dropping out of school and of delinquency, and more mental and physical health problems (Harris, Qualter, & Robinson, 2013; Heinrich & Gullone, 2006). Loneliness in childhood is a valuable predictor of depressive symptoms in adolescence (Qualter, Brown, Munn, & Rotenberg, 2010).

Two theories have systematically illuminated the nature of loneliness and the factors that influence it: the social needs theory and the cognitive processes approach (Terrell-Deutsch, 1999). The social needs theory claims that loneliness is a response to unmet needs for social connection or to unsatisfactory interpersonal relationships. Consistent with this theory, the majority of research has focused on peer relations to understand children's loneliness, and during the past three decades, peer relations have been found to be a critical factor in children's loneliness (Asher & Paquette, 2003). In contrast to the social needs theory, cognitive processes theory suggests that loneliness is not a result of unmet inherent social needs but of dissatisfaction with one's perceived social relationships. In other words, it is the cognitive awareness of a deficiency in either the quality or the quantity of one's social relationships that leads to the discomfort of loneliness (Peplau & Perlman, 1982; Terrell-Deutsch, 1999). However, there is little research examining the effects of individuals' internal cognitive representations of social relations (e.g., perceived social competence) in the link between peer relations and children's feelings of loneliness.

1.1 Peer relations and loneliness

Experiences with peers constitute an important developmental context for children. Children's peer experiences can be divided into several levels of analysis—individual characteristics, social interactions, dyadic relationships, and group membership and composition. The latter three levels of peer system reflect social participation at different interwoven orders of social complexity (Hinde, 1987; Rubin, Bukowski, & Parker, 2006). Previous research has extensively explored the relations between loneliness and multiple types of peer relations, including peer relations at the dyadic and group levels (Margalit, 2010). Acceptance by the peer group and friendship have been found to be important factors for understanding children's experiences of feeling lonely (e.g., Asher & Paquette, 2003; Margalit, 2010).

The preponderance of research on children's loneliness has focused on the possible influence of acceptance versus rejection by peers (Asher & Paquette, 2003). The group's acceptance of a child refers to the degree to which the child is liked or disliked by group members, and group acceptance is an indicator of the child's social status

in the group (Ladd, 1999). Peer acceptance is typically assessed using peer nominations, a sociometric method in which children identify (or "nominate") group members they like most and like least. Acceptance is measured by the number of "like most" nominations a child receives, standardized according to class size, and rejection is the standardized number of "like least" nominations received. These two scores can be combined to create a social preference score (the difference between the acceptance and rejection scores), reflecting the child's relative standing in terms of acceptance by the peer group (Hymel, Vaillancourt, McDougall, & Renshaw, 2002). Previous studies have shown that children who have higher levels of peer acceptance or are more socially preferred are less likely to report feeling lonely (e.g., Mouratidis & Sideridis, 2009; Shin, 2007; Yu, Zhang, & Yan, 2005; Zhou, Sun, Zhao, & Hsueh, 2005).

Friendship lies at the dyadic level of peer experience. A "friend" is defined as a person you know well and like, usually not a member of your family (Hornby, 2010). Friendship is the relationship between friends. The definition of friend suggests that friendships may vary in their degree of mutual knowledge and affection, characteristics which constitute friendship quality. Researchers have consistently found a high negative correlation between friendship quality and children's loneliness (e.g., Hoza, Bukowski, & Beery, 2000; Nangle, Erdley, Newman, Mason, & Carpenter, 2003; Parker & Asher, 1993; Sun, Zhou, Fan, & Ke, 2009). Specifically, studies have indicated that companionship and support from friends were conducive to lessening or eliminating children's loneliness (Parker & Asher, 1993; Uruk & Demir, 2003). Moreover, longitudinal research has shown that friendship quality in middle childhood significantly and negatively predicted loneliness two years later (Zhou, Zhao, Sun, & Ding, 2006).

Nonetheless, relatively few studies have examined the link between the quantity of friends children have and children's loneliness, and the findings of these studies were inconsistent. Ladd, Kochenderfer, and Coleman (1997) found that although the quantity of friends did not predict children's loneliness as they made the transition to kindergarten, it positively predicted other aspects of children's adjustment (such as academic readiness and school involvement). Parker and colleagues (Parker & Asher, 1993; Parker & Seal, 1996) have shown that in middle childhood having one friend is predictive of lower loneliness, but they did not explore the effect of having more than one friend. Recent research in middle childhood samples has suggested that the number of friends is important in predicting loneliness. With a sample of third– through sixth–grade students, Nangle et al. (2003) demonstrated that the number of both best friends and good friends was important in predicting children's positive adjustment, including a lower level of loneliness. The number of early mutual friends at the ages of 9–10 has been shown to significantly and negatively predict loneliness two years later (Zhou et al., 2006).

Although research consistently suggests that both friendship processes and acceptance by peers can predict children's feelings of loneliness, these variables as typically measured (e.g., by friendship nominations and social preference scores) still leave much of the variance in loneliness unexplained (Asher & Paquette, 2003). One possible explanation is that not all friends are equally important in protecting a child from being lonely. Indeed, a child's friendships and social status in the peer group are embedded in a larger social network of peer relations and experiences (Gifford–Smith & Brownell, 2003). The peer group is a network in which most children connect with several other children by friendship. However, each of these friends has a distinct influence on the child, and one reason for this variation is that a child's friends may vary in social status. For example, friends with many connections

in the peer group might provide more information or resources to the child than more isolated friends could. Thus one way to assess a child's connectedness in the peer group is to take into account not just the number of friends, but also the number of "friends of friends." Even if a child has few direct friends, a high number of indirect friends might reduce the risk of loneliness. A useful index of the quantity of indirect friends is called proximity prestige, a measure that developed from social network analysis.

In recent years, social network analysis, a method used to study the structure and characteristics of social networks (Scott, 2000), has been identified as an appropriate method for studying the peer context of child and adolescent social behaviors and emotions, such as aggression (Faris & Ennett, 2010), substance use (Ennett et al., 2006; Mercken, Steglich, Sinclair, Holliday, & Moore, 2012; Mundt, 2011), prosocial behaviors (Gest, Graham–Bermann, & Hartup, 2001), depression (Okamoto et al., 2011), and loneliness (Chamberlain, Kasari, & Rotheram–Fuller, 2007). This method generates a visual depiction of the complex inter–relations in a social network. Each actor (e.g., a child) is represented by one node, and the relationships (e.g., friendships) between any two actors are represented by lines. Some actors are linked directly, and some are linked indirectly through relationships with other actors. In the visual representation of these relationships, the distance between two actors is one unit if the actors are linked directly, two units if they are linked indirectly through one other actor, and so on. The shorter the distance is between actors, the closer the relationships are between them. An actor's influence domain is the number or percentage of all actors linked to him or her directly and indirectly in a closed group. However, the intensity of the actor's influence varies depending on the closeness of the relationship. The farther away that neighbors are from an actor, the smaller the actor's influence on them.

Using this type of information about a social network,

Lin (1976) first proposed a measure of prestige called proximity prestige. According to Wassermann and Faust (1994), the proximity prestige of an actor is the proportion of all other actors directly or indirectly connected to the actor in a closed group (influence domain) divided by the average distance these other actors are from the actor (the mean of influence intensity). This means that when relationships are defined by friend nominations, then the relationships are weighted differently depending on whether the link is direct (one child is nominated by another) or indirect (one child is nominated by another, who in turn is nominated by someone else). In a friend nominations network, the influence domain is the quantity of direct and indirect friends. The influence intensity represents the closeness of these relationships. Therefore, proximity prestige is a composite representation of the quantity and the closeness of children's direct and indirect friends. Meanwhile, this index, which is calculated based on friend nominations by others, also reflects children's objective friendship status in the peer group.

In social network analysis, one widely used index is centrality, which is the proportion of nominations one actor receives from other actors in a group. The actors who receive many positive nominations are considered to be prominent and to occupy a central position in the peer group. However, the centrality index makes no distinction in terms of the popularity of the nominators. For example, two actors who receive the same number of nominations, but who differ in how active they are in the larger network, would have the same centrality index score. Proximity prestige extends the concept of centrality by taking the popularity of the nominators into consideration.

It is important to distinguish proximity prestige from other concepts of peer relationships (e.g., peer acceptance, social preference), especially because the number of direct and indirect friends in the social network may highly coincide with the number of "like most" nominations, the

basis of peer acceptance and social preference scores. But naming a specific child as a friend involves affirmation of a special dyadic relationship, whereas peer acceptance and social preference are unilateral constructs representing the general affective inclination of the group toward an individual. Even though these three constructs (peer acceptance, social preference, and proximity prestige) all lie at the group level of peer experience, proximity prestige is based on the social network of friendships and thus represents greater social complexity than the other two indexes.

Friendship quality, social preference (a measure of acceptance by the group) and proximity prestige lie at the dyadic and group level of peer relations, and they reflect different facets of children's experience with peers. Therefore, the first goal of the present study was to test whether variables representing peer relationships at different levels of social complexity were associated with children's loneliness, and it was hypothesized that the proximity prestige index would provide extra explanatory power above that of friendship quality and social preference in predicting children's loneliness.

1.2 Self-perceived social competence and loneliness

Measures of peer relations are often objective measures of social status, which may not correspond with children's subjective experience of loneliness. According to cognitive processes theory, it is a child's cognitive appraisal of peer relations that gives rise to the feeling of loneliness (Terrell-Deutsch, 1999). Self-perceived social competence is the internal cognitive estimation of one's own social competence. Children who perceive higher social competence are more likely to be satisfied with their peer relations, and less likely to feel lonely. Sun et al. (2009) discovered that self-perceived social competence was more predictive of loneliness than was objective social status. In addition, children's social success may lead to

positive social evaluations of their own competencies (Cole, 1991); in turn, positive self-perception may further lead to less loneliness. It is supposed that self-perception of social competence is a mediator of the link between peer relations and loneliness. Studies with third- to sixth-grade children have shown that self-perceived social competence mediates the association between direct peer relations (peer acceptance and friendship quality) and loneliness (Sun et al., 2009; Zhou et al., 2005). However, these studies did not take indirect peer relations into consideration. Therefore, the second goal of the present study was to test whether proximity prestige (a reflection of both direct and indirect friendships and objective friendship status) in the peer network is associated with children's loneliness through the mediator of self-perceived social competence.

1.3 Peer relations, self-perception, and loneliness

Previous studies have extensively explored different levels of peer relations as predictors of children's loneliness; however, the studies have tended to concentrate on each predictor independently rather than integratively. Based on an extensive critical review, Gifford-Smith and Brownell (2003) emphasized that children's peer relations are multifaceted, intersecting and overlapping systems, and it is necessary and valuable to attend to multiple aspects of the peer experiences simultaneously from an integrative perspective. Hinde (1987) also emphasized that events and processes at each level of peer experience are constrained and influenced by events and processes at other levels (Rubin et al., 2006). Additionally, researchers have pointed out that peer relations constitute just one facet of the contexts in which peer interactions occur and develop. These contexts determine the particular skills, perceptions, and cognitions that will be most effective for children's successful functioning in the peer group (Brownell & Gifford-Smith, 2003; Sheridan, Buhs, & Warnes, 2003). Thus, the third goal of the present research

was to test a conceptual model in which peer experience at multiple levels of social complexity (namely friendship quality, social preference, proximity prestige, and self–perceived social competence) jointly predicts children's loneliness; in this model, children's self–perceived social competence is also tested as a mediator, representing a possible underlying mechanism in the association between peer relations and children's loneliness.

It was noted that the level of social interaction was not represented in the conceptual model. Social interaction encompassed different kinds of behaviors and behavioral tendencies (Rubin et al., 2006). The theoretical and substantial evidences showed that social behaviors were the antecedents or predictors of peer relations in childhood (Chen, Rubin, & Sun, 1992; Ladd, Buhs, & Troop, 2002). According to the concept and theory of loneliness, it was highly emphasized that loneliness was the response to the lack of social connection (peer relations). The present study mainly focused on the link between peer relations and the feeling of lonely in the theoretical framework of loneliness. Thus, the antecedents of peer relations (e.g., the factors of social interaction) were not included in the analyses.

Both the social needs theory and the cognitive processes approach of loneliness indicate that the satisfaction or perceived satisfaction of interpersonal relationships could lower or eliminate feeling lonely (Terrell–Deutsch, 1999). The interpersonal relationships mentioned by the theories are the positive relations what individuals want to obtain (contrary to negative relations, such as peer victimization). Most of the research on children's loneliness has focused on the great contributions of two kinds of positive peer relations, acceptance by peer group and the friendship (Asher & Paquette, 2003; Margalit, 2010). The research has consistently found that children who are better accepted by peers or have a high–quality friendship with a best friend report experiencing less loneliness than other children. On the contrary, as

one facet of friendship, the quantity of friends children has been occasionally examined its relation with loneliness, and the results are not consistent. Meanwhile, the research has not taken indirect friendships and the statuses of friends into consideration. Therefore, the index of proximity prestige is introduced to represent the number of indirect friends and the statuses of friends. The positive peer relations in the peer group can be represented by the total of proximity prestige (composite representation of the quantity and the closeness of children's direct and indirect friends), friendship quality (the relationship with best friend) and social preference (the affective inclination of the group toward an individual). The present study aimed to explore the simultaneous influence of the three kinds of peer relationships (friendship quality, social preference, and proximity prestige) on children's loneliness, especially the unique contribution of proximity prestige on loneliness.

1.4 Culture and loneliness

Most research on children's loneliness has been conducted in Western cultures, expecially North American, which are typified by an individualistic orientation. Individualist cultures value individualization, autonomy, and privacy. Asian countries including China are collectivistic cultures that prioritize relational bonds and group cohesion (Oyserman, Coon, & Kemmelmeier, 2002). The Western cultural orientation of individualism has been found to be associated with higher loneliness among adults, even in Western societies (Rokach & Bauer, 2004; Seepersad, Choi, & Shin, 2008). Multinational studies of the loneliness of 3rd to 6th grade children conducted in Canada, Brazil, China, and Southern Italy revealed no mean differences in self–reported loneliness across these four samples. However, the overall patterns of relations between social behaviors (such as aggression and shyness–sensitivity) and loneliness differed across samples. For example, shyness is viewed as maladaptive in Western cultures (Fox, Henderson, Marshall, Nichols,

& Ghera, 2005), and indeed, results of the multinational study showed that shyness was positively associated with loneliness in Canadian, Brazilian and Italian children, but not associated with loneliness in Chinese children. These results suggested that the nature of children's loneliness may be affected by the broad socio-cultural context (Chen, He, De Oliveira, et al., 2004).

In contemporary China, one major mission of schooling is to help children develop collectivistic ideologies (Chen, 2000). Children are encouraged to participate in group activities, develop positive attitudes toward the group, and learn social skills that facilitate group functioning. The emphasis on social connections results in an interdependent sense of self (Markus & Kitayama, 1991). In this context, social alienation and loneliness are more likely to be viewed as maladaptive or problematic (Chen, 2000). Chinese children's self-perceived social competence is also likely to be influenced by cultural context. Self-evaluation has been encouraged traditionally in Chinese culture (e.g., Luo, 1996). In Chinese schools, children are regularly required to engage in self-evaluation, in the belief that children who are conscious of their strengths and weaknesses are likely to perform and regulate their behavior better according to social norms and values. A particular emphasis is placed on self-awareness of the negative aspects of one's social and moral character (Chen, He, & Li, 2004). Therefore, self-perceived social competence would be expected to have close links with peer relations and feelings of loneliness in Chinese children.

1.5 Study hypotheses

Researchers consider peer relations to be a multidimensional social network structure (Gifford-Smith & Brownell, 2003; Margalit, 2010), and many studies have explored the effects of one or two facets of peer relations on children's loneliness. However, little research has examined multiple facets of peer relations in relation to

children's loneliness, and the underlying mechanisms of these associations. The goal of the current study was to explore multiple indexes of peer relations as predictors of loneliness in middle childhood, and to test children's self-perceived social competence as a mediator of these links in the context of Chinese culture. Based on the empirical research and associated conceptual models to date, a new conceptual model was developed and tested in the current study (see Figure 1). It was hypothesized that: (a) friendship quality, social preference, proximity prestige, and self-perceived social competence would be negatively correlated with children's loneliness; (b) proximity prestige would explain unique variance in loneliness scores; (c) self-perceived social competence would mediate the association between peer relations and loneliness in both separate and integrated analyses.

2 Method

2.1 Participants

Data for the present research were collected from 509 third- to sixth-grade students of eight classes in a public ordinary elementary school in Wuhan, a big city in central China, at the end of the fall semester, December, 2008. The school had six classes at each grade level, and two classes of each grade were randomly selected for participation. The sample was composed of 233 girls and 276 boys (roughly equal numbers of girls and boys in each class). Almost all children fell in the age range of 8 to 12, with 125 third-graders, 127 fourth-graders, 129 fifth-graders, and 128 sixth-graders (62 to 66 students in each class). Four children were absent from school on the day of the survey, so the overall participation rate was 99.2%. The participation rates for six classes were 100%, and that for the other two classes were 95% and 98%. By teachers' report most of the students (>95%) were Han, the majority ethnic group in China, with a small number of students being from
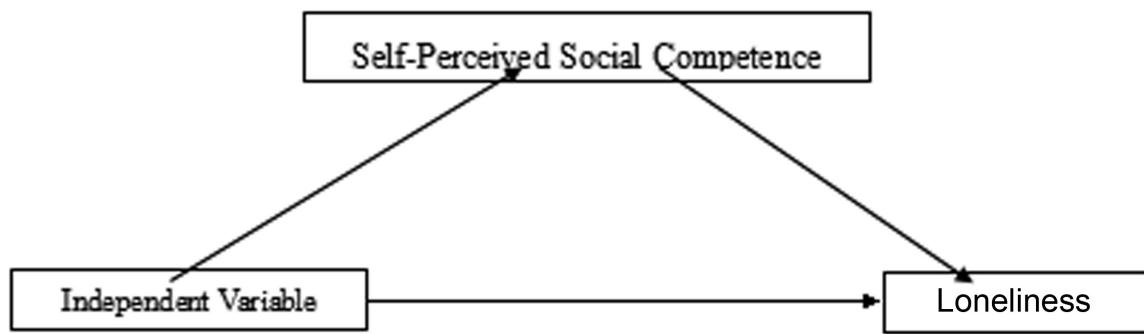
Fig. 1. The mediation model of peer variables, self−perceived social competence and loneliness. The independent variables were friendship quality, social preference, and proximity prestige.

minority groups. The students were primarily of middle socioeconomic status (> 80%), the majority of parents received college or university education (> 85%), and most (>95%) of the children were the only child in the family, an effect of China's 'one−child−per−family' policy implemented in the late 1970s. The sample was representative of ordinary elementary school children in urban China.

2.2 Measures

All the measures used in this study were originally English versions, which have been widely used in Western countries. Previous researchers translated these measures into Chinese language to apply to Chinese children, and found that the Chinese versions achieved the standards of psychometrics. The measures were commonly used in the research of Chinese children. The information about reliability and validity of each measure in prior research of Chinese children are reported in the descriptions of each measure.

2.2.1 Peer nominations for "like most" and "like least"

Researchers provided each child a printed name list of all children in the class, and children were asked to circle the names of classmates whom they "liked most" and "liked least" with no limit on the nomination number. Compared with the limited nomination method, the unlimited−choice procedure could provide a greater range, more normal−like distribution of values on the nomination items, and make the nominations more reliable

(Marks, Babcock, Cillessen, & Crick, 2012; Terry, 2000). Marks et al. (2012) demonstrated that the Cronbach's α s of unlimited liked−most nomination was approximately .75 with a 95% participation rate. The raw number of the children rated on the liked−most nomination was 9.73 (*SD*=6.92) on average, and 6.90 (*SD*=8.33) on the liked−least nomination. Then, the numbers of positive (like most) and negative (like least) nominations were calculated for each student and standardized within classrooms. This sociometric method has been widely used to assess children's peer acceptance and rejection in a variety of cultures including Chinese (e.g., Chen, Zappulla, Coco, et al., 2004; Zhou et al., 2006). Following Coie, Dodge, and Coppotelli's (1982) procedure, an index of social preference was formed to indicate the child's relative likeability in the group. This score was generated from the difference between the positive nomination score and the negative nomination score; higher scores indicate that compared to others in the class, the child is relatively more liked by peers. This index has been extensively used to reflect the peer relations of third− to sixth−grade children in China (e.g., Chen, Zappulla, Coco, et al., 2004; Zhou et al., 2005). One research with a sample of 3rd to 5th grade Chinese children showed that the test‒retest correlation of social preference was .65 over a one−year interval (Zhou, Sun, Xiang, & Liu, 2007). Social preference was moderately and positively associated with sociability (r= .35) and self−perceived social competence (r=0.23), and

negatively associated with aggression (r=−.46) in Chinese children (Chen, Zappulla,Coco, et al., 2004).

2.2.2 Friendship nominations

Participants were provided with a name list of all students in the class, and were asked to circle the names of classmates who were their friends, with no restriction on the number of nominations that could be made. Each child received 9.74 friendship nominations on average (*SD*=6.13). This procedure has been widely used across cultures, including Chinese culture, to identify children's friends (e.g., Chen et al., 1992; Parker & Asher, 1993; Zhou et al., 2006). Marks et al. (2012) demonstrated that unlimited friendship nominations were more reliable than 'top 3' friend nominations (Cronbach's α s were .60 and .38, respectively) when the participation rate was 95%. A previous study found that the quantity of good friends was positively correlated with children's popularity (r=0.65, .72), and negatively correlated with children's loneliness (r =−.39, −54) (Nangle et al., 2003).

In the current study, the key information was the number of "friend" nominations that each child received (rather than gave). A direct nomination occurred when a classmate nominated the child as a friend. An *indirect* nomination occurred when the classmate who had nominated the child had in turn been nominated as a friend by someone else in the class.

In mathematical terms, these connections can be described according to distance. The path distance of a direct nomination is 1; when the nomination is one step removed (that is, there is an indirect nomination), the path distance increases by 1. For example, child A names child B as a friend (a direct nomination with a path distance of 1), and child B names child C as a friend (another direct nomination, with a path distance of 1), but child A does not name child C as a friend. Child A and child C are indirect friends, and the path distance between them is 2. A student's proximity prestige is the ratio of the proportion

of all other classmates who directly or indirectly nominated the student to the mean path distance of these nominations. The formulation is as following (Wassermann & Faust, 1994):

$$P_P(n_i) = \frac{I_i/(g-1)}{\sum d(n_j, n_i)/I_i}$$

$P_{p(ni)}$ represents the prestige of actor I; I_i represents the total number of actors who directly or indirectly nominate actor i; g represents the total number of actors in the group; $d_{(nj, ni)}$ represents the distance between the nominator j and the actor i. If no one nominates actor i, $P_{p(ni)}= 0$; if all actors in the group directly nominate actor i, $P_{p(ni)}= 1$.

2.2.3 Loneliness

Children's loneliness was assessed by the modified version of the Loneliness and Social Dissatisfaction Questionnaire (Asher & Wheeler, 1985). This self-report measure provides a five-point scale of 1 (*not at all true*) to 5 (*always true*) to rate 24 items, 16 of which assess feelings of loneliness and social dissatisfaction at school (e.g., "I have nobody to talk to in class;" "I am lonely at school;" "I do not have any friends in class"), and eight of which are filler items about personal interests (e.g., "I like to read"). The average score of the 16 items was used to represent loneliness, with higher the scores representing higher self-reported loneliness. Zhou et al. (2006) revised this measure for use with Chinese children, and found the 16 items of loneliness loaded on one factor which did not related to the 8 filler items, the Cronbach's alpha coefficient was .92. In a sample of Chinese children from grade 3 to grade 6, Chen, He, De Oliveira, et al. (2004) reported an alpha reliability coefficient of .95, and loneliness was moderately and negatively associated with social preference (r =−.31) and sociability (r =−.27). The test-retest correlation was .42 over a one-year period in another sample of Chinese children from grade 3 to grade 5 (Zhou et al., 2007). The internal consistency (Cronbach's

alpha) was .92 in the present study.

2.2.4 *Self−perceived social competence*

Harter's (1982) Perceived Competence Scale for Children (PCSC) consists of four subscales representing children's self−perceived competence in four domains: social, cognitive, sport skill and global competence. Stigler, Smith, and Mao (1985) revised this scale for use with Chinese children and showed that the four−factor structure identified in U.S. samples of children in grades 3−9 (Harter, 1982) was replicated in a sample of Chinese fifth−grade children. For the social subscale (self−perceived social competence), in prior research on samples of Chinese children, children's self−perceived social competence was moderately and positively correlated with their sociability (.29) and social preference (.23) (Chen, He, & Li, 2004) and negatively related to victimization (−.19) (Zhou et al., 2006). The Cronbach's alpha coefficient was reported to be .67 and .64 (Chen, He, & Li, 2004; Stigler et al., 1985). Test−retest correlation was .48 over a one−year period in a sample of Chinese children in grades 3−5 (Zhou et al., 2007). The self−perceived social competence subscale was used in the present study. This subscale contains six items, with each item depicting two kinds of kids in two statements (e.g., "Some kids find it hard to make friends;" "Other kids find it's pretty easy to make friends"). Children were asked first to decide whether they were more like the child depicted on the left or the one on the right, and then to decide whether that part of the statement was "really true for me" or only "sort of true for me." This procedure of narrowing the choice gradually is helpful to allow even young children to produce meaningful ratings on a 4−point scale (Stigler et al., 1985). The average score of the 6 items was calculated as the subscale score, and the higher the score, the more positive the children perceived their social competence to be. In the present sample, Cronbach's alpha for the subscale was .56.

2.2.5 *Friendship quality*

Friendship quality was measured using an abbreviated version of the Friendship Quality Questionnaire (FQQ) (Parker & Asher, 1993). The abbreviated version of the FQQ consisted of the three items that showed the highest factor loadings on each dimension of the FQQ: Validation and Caring (e.g., "He or she tells me I am good at things"), Companionship and Recreation (e.g., "We always pick each other as partners for things"), Conflict and Betrayal (e.g., "We argue a lot"), Help and Guidance (e.g., "We help each other with schoolwork a lot"), Conflict Resolution (e.g., "We make up easily when we have a fight"), and Intimate Exchange (e.g., "We always tell each other our problems") as reported by Parker and Asher (1993), for a total of 18 items (Nangle et al., 2003). Zhou et al. (2006) revised the abbreviated version of the FQQ for use with Chinese children, and confirmed this measure's six−factor structure in the sample of Chinese children in grades 3−6, with an alpha reliability coefficient of .83; scores on this measure were moderately and positively correlated with peer acceptance (.26, .27) and number of mutual friends (.29, .36). Test−retest reliability was .40 over a one−year period in a sample of Chinese children in grades 3−5 (Zhou et al., 2007). In the present study, the children were asked to nominate their very best friend and assess their relationship using this questionnaire. Children responded to these items on a 5−point scale ranging from 0 (*not at all true*) to 4 (*really true*). After reversing the item scores on the dimension of Conflict and Betrayal, the scores of the 18 items were averaged, with higher scores indicating better friendship quality. In the present sample the Cronbach's alpha for the scale was .85.

2.3 Procedure

The survey was conducted through a passive parental consent process, and was overseen by the office of the Research Provost at the researchers' institution. The children were also told that they could drop out if they did not want to participate before the survey. All measures were collected in two group sessions in each classroom, and

each session lasted approximately 40 minutes. The order of presentation of the surveys in the group sessions was counterbalanced by classroom. Surveys were administered by two trained graduate students in developmental psychology. These graduate students had received one-hour-training which including how to conduct the survey to children, how to cope with the problems they might meet, the points of caution and the special requirements for researchers. These researchers told the purpose of the investigation, assured the children that their answers were confidential, gave instructions on each measure and provided individual assistance when children had difficulties completing the survey. Researchers checked the questionnaires when they were turned in and asked children to complete any missing items. By using a Filemaker 6.0 template, all data were entered and cross-checked by two groups of trained graduate students in developmental psychology, those who had got half-an-hour-training about the form of the Filemaker template, especially how to enter the data of nominations.

2.4 Method of statistical analysis

First, descriptive statistics and correlations among all variables were generated. T-tests and F-tests were then performed to examine gender and grade differences on the study variables. Then, bias-corrected bootstrap analyses were employed to test whether self-perceived social competence mediated the effect of each peer variable (friendship quality, social preference, proximity prestige) on children's loneliness, using Mplus 7.0 statistical software (Muthén & Muthén, 1998-2010). Bootstrap analysis is a nonparametric approach that directly tests the significance of mediation effects, makes no assumptions about the distribution of variables or the sampling distribution of the statistic, and can be applied to smaller samples with greater confidence than is possible with other methods (Fairchild & McQuillin, 2010). The present study used a 99% confidence interval and 1000 bootstrap

samples to evaluate the magnitude and significance of indirect effects. An effect is considered to be statistically significant if the confidence interval does not include zero.

Finally, structural equation model (SEM) analyses were conducted to test an integrative mediation model in which peer relationship variables (friendship quality, social preference, and proximity prestige) predicted children's loneliness both directly and indirectly through children's self-perceived social competence. Path analyses were conducted with Mplus 7.0 using maximum likelihood estimation of model parameters, and bias-corrected bootstrap analyses were also used to test the significance of the mediation effects (Muthén & Muthén, 1998-2010). A just-identified model was first analyzed, including all direct and indirect pathways from friendship quality, social preference and proximity prestige to self-perceived social competence and loneliness; this model assumed that there were correlations among friendship quality, social preference and proximity prestige. Then, based on the modification indices, a trimmed model was analyzed excluding the non-significant (p > .01) pathways from the initial saturated model. Lastly the Monte Carlo simulation was used to verify whether modification model via modification indices would generalize. The Monte Carlo simulation is increasingly popular in evaluating statistical estimators for structural equation models (see ntroduction in Paxton, Curran, Bollen, Kirby, & Chen, 2001). The present study specified and analyzed the hypothesized integrative model and the modified model using the same simulated data set (10000 replications for N=509), and compared the modelfit statistics and indices of the two models from the Monte Carlo simulation.

2.4.1 Missing data

Social preference and proximity prestige for the children in two classes were calculated with 95% and 98% participation rates. According to Cillessen and Marks (2011), at least 60－70% were necessary to obtain reliable

peer nomination data for social acceptance or preference. Thus, the class participation rates in the present research were sufficient to achieve reliable peer nomination data. Missing information on other variables was as following: loneliness (8 children, 1.6%), friendship quality (5 children, < 1%), and self−perceived social competence (26 children, 5.1%). In order to account for the missing data, all the analyses were completed using maximum likelihood estimation (ML) which is considered a state−of−the−art missing data technique and produces less biased estimates than traditional techniques (listwise deletion and single imputation methods; see Baraldi & Enders, 2010).

3 Results

3.1 Descriptive statistics

Descriptive statistics for the observed variables are presented by gender and grade in Table 1. Compared to boys, girls reported lower loneliness, higher perceived social competence and friendship quality; girls also had higher social preference scores and higher proximity prestige. The only significant grade difference was in proximity prestige. The proximity prestige of sixth graders was higher than that for the other grades, and the proximity prestige of fourth and fifth graders was higher than that of third graders, but there was no significant difference between fourth and fifth graders' proximity prestige. Fig.2 shows students' proximity prestige within one class. In this figure, students (called actors in the language of social

network analysis) are situated within the network as a function of direct and indirect friendship nominations. The students who received more friendship nominations (direct and indirect) had higher proximity prestige (e.g., v2, v3). In addition, some students received few nominations, but the nominators themselves had many nominations, so that the students indirectly had comparatively high proximity prestige. For example, v30 and v35 each had one direct nomination, but their proximity prestige differed because the proximity prestige of their nominators differed: The child who nominated v30 (i.e., v16) received more nominations than the child who nominated v35 (i.e., v44).

3.2 Correlations among peer relations, self−perceived social competence, and loneliness

Table 2 shows that, as expected, loneliness was negatively and significantly associated with self−perceived social competence (the hypothesized mediator) and with friendship quality, social preference, and proximity prestige (the hypothesized predictors of children's self−reported loneliness). Self−perceived social competence was positively and significantly correlated with the three indicators of peer relations (friendship quality, social preference, and proximity prestige). This pattern of correlations indicates that it would be appropriate to test the proposed mediation model.

3.3 Mediating effect of self−perceived social competence

Table 3 presents the results of the bootstrap analyses for each of the separate mediation models. The 99% bias−

Table 1 Means and standard deviations of the research variables by gender and grade

Variable	Gender				T	Grade								F
	Boys n = 276		Girls n = 233			Third n = 125		Fourth n = 127		Fifth n = 129		Sixth n = 128		
	M	(SD)	M	(SD)		M	(SD)	M	(SD)	M	(SD)	M	(SD)	
Loneliness	1.88	(0.78)	1.56	(0.60)	5.17***	1.76	(0.70)	1.70	(0.73)	1.73	(0.76)	1.74	(0.70)	0.16
Self−perceived Social Competence	2.78	(0.54)	2.93	(0.50)	−3.37***	2.85	(0.53)	2.85	(0.57)	2.86	(0.52)	2.84	(0.49)	0.05
Friendship quality	2.89	(0.75)	3.11	(0.73)	−3.34***	2.95	(0.63)	2.86	(0.89)	3.07	(0.75)	3.06	(0.68)	2.24
Social preference	−0.53	(1.78)	0.63	(1.40)	−8.20***	0.00[a]	(1.65)	0.00[a]	(1.76)	0.00[a]	(1.72)	0.00[a]	(1.75)	0.00
Proximity prestige	0.41	(0.12)	0.44	(0.08)	−3.90***	0.35	(0.11)	0.43	(0.10)	0.42	(0.08)	0.49	(0.09)	45.07***

Note: N = 509. [a] Values are Z scores. *** p < .001.
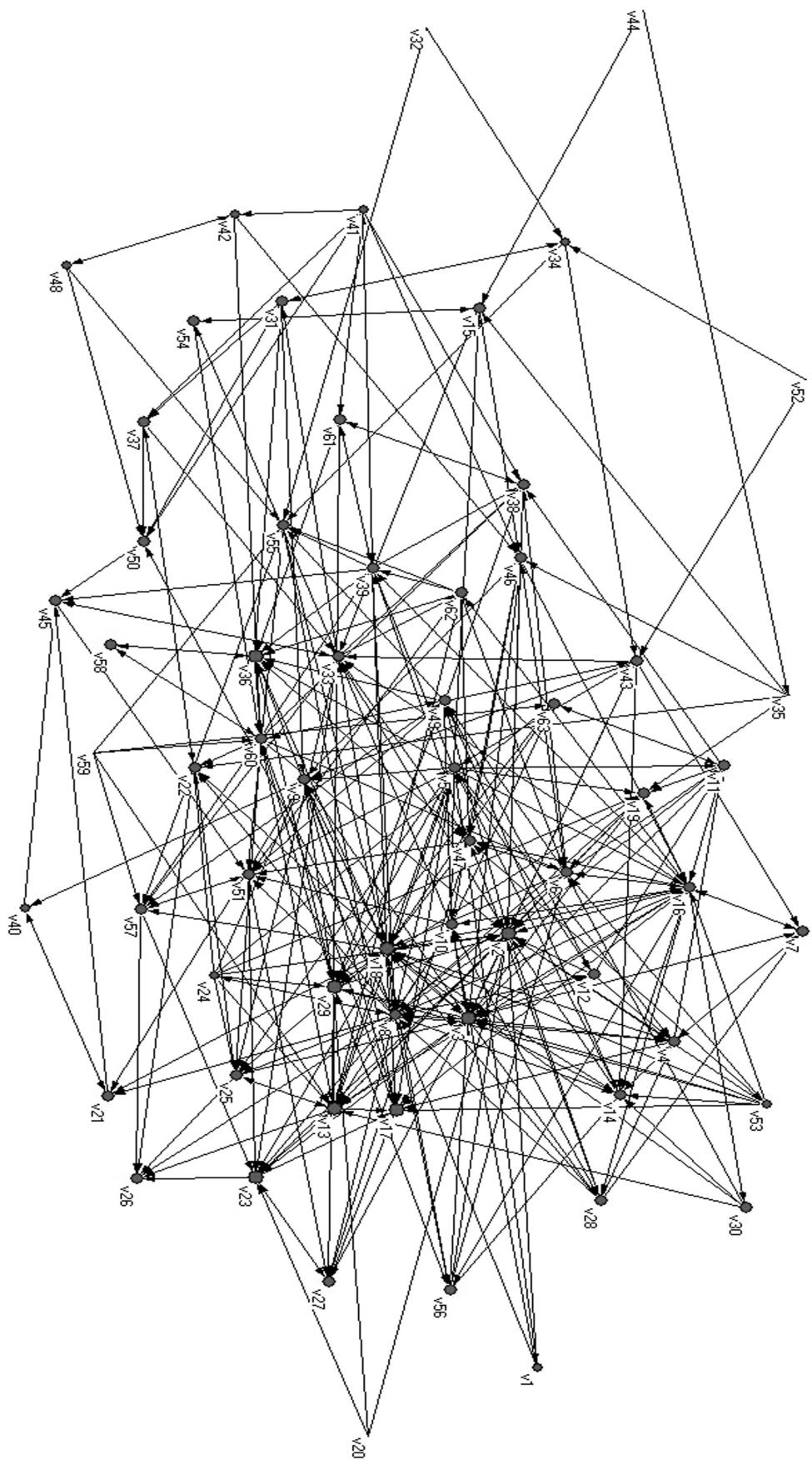
Fig. 2. The sociogram of one class (class size = 63), depicting each child's proximity prestige. Each node denotes one child in the class; the line denotes one child (sender) who nominates another child (receiver) as a friend. The size of the nodes denotes the children's proximity prestige, and the bigger the node is, the more prestigious the child is. The sociogram was drawn with Pajek.

Table 2. Correlations among the research variables

Variables	1	2	3	4	5
1. Loneliness					
2. Self-perceived Social Competence	−.69**				
3. Friendship quality	−.41**	.34**			
4. Social preference	−.44**	.29**	.31**		
5. Proximity prestige	−.37**	.23**	.26**	.60**	
Mean	1.73	2.85	2.99	0.00ᵃ	0.42
Standard deviation	0.72	0.53	0.75	1.72	0.10

Note. N = 509. ᵃValues are Z scores. ** $p < .01$.

corrected bootstrap confidence interval excluded zero for each of the three peer variables. In regard to model fit, the three models are all just-identified which have $df=0$, $CFI=1$, $TLI=1.000$, $RMSEA=0$, $SRMR=0$. The results suggested that self-perceived social competence was the mediator of the relations between friendship quality and loneliness, social preference and loneliness, and proximity prestige and loneliness, respectively.

3.4 Path model linking peer relations with loneliness

The hypothesized integrative model was a just-identified model, the information for model fit as following: $df=0$, $CFI=1$, $TLI=1.000$, $RMSEA=0$, $SRMR=0$. There was no significant pathway from proximity prestige to self-perceived social competence (.06, $p=0.24>.01$). The 99% bias-corrected bootstrap confidence intervals (99% BC CI) for the indirect effect of friendship quality or social preference did not contain zero (−.240, −.083; −.084, −.009, respectively). However, there was zero in the 99% BC CI for the indirect effect of proximity prestige (−.949, .376). The indirect effect of proximity prestige was not significant. The insignificant pathway was then excluded and the modified model was set. The result showed that the modified model was acceptable and a good fit to the data: $\chi^2 (1)=1.39$ ($df=0.24$), $CFI =1.00$, $TLI=0.99$ and RMSEA $=0.03$ (90% confidence interval values ranging from .00 to.13). Monte Carlo simulations also gave support for the modified model over the hypothesized integrative model. The result for the hypothesized integrative model from the Monte Carlo simulation showed that: for the information criteria (AIC, BIC, $ABIC$), the values observed in the Monte Carlo replications were close to the theoretical values; for χ^2, $RMSEA$, and $SRMR$ the observed values were not close to the theoretical values. Meanwhile, for the pathway from proximity to self-perceived social competence, there was only 77.8% of 10000 replications for which the parameter estimate was significantly different from zero at $\alpha = .05$, and it was smaller than the cut-off point value of .80 often used (Cohen, 1988). In regard to the modified model, the observed values were all close to the theoretical values for the model fit statistics and indices from the Monte Carlo simulation, and the lowest percentage of 10000 replications for which the parameter estimate was significantly different from zero at $\alpha =0.05$ was 92.4% (larger than the cut-off point value of .80). The Monte Carlo simulation for the modified model in present study provided good results, and demonstrated that the modifications made via modification indices would generalize. Fig.3 presents the final integrative model of the SEM analyses.

The hypotheses were partially supported. In the modified model, there was a direct link from proximity prestige to loneliness, and a lack of a direct link between proximity prestige and self-perceived social competence. This means that self-perceived social competence did not act as a mediator between proximity prestige and loneliness when the influences of social preference and friendship quality were taken into account. There were direct links between friendship quality and loneliness, and between social preference and loneliness. Meanwhile,
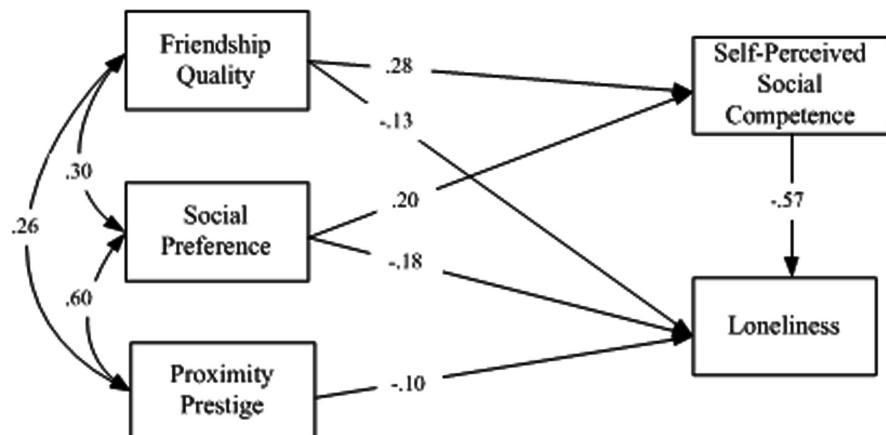
Fig. 3 Path model showing standardized path coefficients for associations among peer variables, self–perceived social competence and children s loneliness. Data were also analyzed using MLR in Mplus to correct for non–normality and the results did not change. *N*=509.

the effects of friendship quality on loneliness and of social preference on loneliness were transmitted through the intermediary predictor variable of self–perceived social competence, with the mediated (indirect) pathways having effects of −.16 and −.12 (p<.01), respectively. The 99% BC CI for the mediation (indirect) effect with friendship quality was (−.225, −.088), and that for social preference was (−0.183, −.049). These mediating effects accounted for 55% of the total effects of friendship quality on loneliness (standardized path coefficient of −.29), and 38% of the total effects of social preference on loneliness (−.29). The peer relationship variables (friendship quality, social preference, and proximity prestige) and self–perceived social competence together accounted for 56% of the variance in loneliness.

4 Discussion

According to the cognitive processes theory, loneliness could be influenced by self–cognition of social relationships, and peer relations provide the basis for self–cognition. Differing from past studies on loneliness in middle childhood, the present study assessed peer relations at multiple levels of social complexity, and introduced a measure of prestige in the social network. It was hypothesized that self–perceived social competence (self–cognition) would mediate the associations between peer relations (friendship quality, peer acceptance and proximity prestige) and children's loneliness, and mixed support for this hypothesis was found. The three

Table 3. The Confidence Mediated Effects of Self–Perceived Social Competence and the Corresponding Bias–corrected Bootstrap Confidence Intervals for Peer Variables as Predictors of Loneliness

Independent Variable	Estimate	SE	99% BC CI a	R^2_{med} [b]
Friendship quality	−.21**	.03	(−.28, −.14)	.13
Social preference	−.17**	.03	(−.24, −.10)	.13
Proximity prestige	−.13**	.04	(−.23, −.04)	.09

Note. Data set contains 4 cases with missing on all three variables (friendship quality, self–perceived social competence and loneliness). These cases were not included in the analysis of the mediation model with friendship quality (N = 505). All cases were included in the analyses of the other two mediation models (N = 509).

[a] BC CI represents bias–corrected confidence intervals. [b]R^2_{med} represents one effect size measure for mediated effect. R^2_{med} is computed by using squared bivariate correlations and the overall model R^2 from a model where Y is predicted from both X and M, as following (Fairchild & McQuillin, 2010): . $R^2_{med} = r^2_{Mr} - (R^2_{r,Mx} - r^2_{xr}) . r^2_{Mr}$ is the squared correlation between the outcome (Y) and the mediator (M), $R^2_{Y,Mx}$ is the overall model R^2 from the regression equation where Y is predicted from X and M. r^2_{xr} is the squared correlation between the outcome and the independent.

** *p* < .01 (significant mediation effect).

measures of peer relations were each significantly and negatively associated with loneliness, and each made a unique contribution to the variance in loneliness scores. Specifically, friendship quality made the highest contribution, in line with previous studies (Nangle et al., 2003; Sun et al., 2009), suggesting that a high quality friendship may be an important protective factor for children's loneliness. Tests of an integrative model showed that when the relations among all variables were taken into consideration, friendship quality and social preference were both associated with loneliness through the mediating effect of children's self-perceived social competence. However, self-perceived social competence did not mediate the relation between proximity prestige and loneliness.

4.1 Proximity prestige and children's loneliness

Previous studies have extensively explored the possible influence of friendships and acceptance by the peer group on children's loneliness. The present study also examined the association between loneliness and a more comprehensive index of peer relationships at the group level, namely proximity prestige, which reflects the quantity and the closeness of children's friendships in the peer social network. Proximity prestige was significantly correlated with loneliness, and in the integrative model it made a unique contribution to the variance in loneliness scores. But it was also highly associated with the other measures of peer relations, namely social preference (r =0.60) and friendship quality (r=0.25). Previous research also has found moderate correlations between network centrality (proximity prestige being one extension of network centrality) and peer acceptance (r=0.49; see Gest et al., 2001). What is unique about the proximity prestige measure is that it takes into account not only children's direct relations with each other (as do the measures of friendship quality and social preference), but also the indirect relations among children. Therefore, the significant

path coefficient from proximity prestige to loneliness in the integrative model suggests that indirect friends were conducive to a reduction in the experience of loneliness. Although a single indirect nomination contributes little to a child's score for proximity prestige, many "distant nominations" may contribute as much as one "close nomination" (de Nooy, Mrvar, & Batagelj, 2011). Being nominated by one friend with many friends has a direct benefit to the child but also increases the child's number of indirect friends, thus promotes the child's prestige (see Fig.2).

4.2 The mediating rffect of delf-perceived social competence

In the present study, children's subjective cognition about social competence, namely self-perceived social competence, made a greater contribution to loneliness scores than did other aspects of peer relations. This finding is consistent with the results of previous studies (Cheng & Furnham, 2002; Zhou et al., 2005) and suggests that subjective cognition about one's own competence plays a crucial role in the experience of loneliness. This result provides substantial support for cognitive processes theory as it relates to children's loneliness.

Previous studies also have found that cognitive factors mediated the relation between peer relationships and emotion adaption (Hymel et al., 2002). For example, Boivin, Hymel and Bukowski (1995) found that the association between negative peer relations and depression was mediated by the experience of loneliness and cognition about the social environment. The present study tested the mediation effect of self-perceived social competence in the relations between peer variables (friendship quality, social preference, and proximity prestige) and loneliness. The three peer variables lie at the dyadic and group levels of peer relations. When considered individually, each variable not only had a direct effect on the experience of loneliness, but also imposed an indirect effect through the

individual's cognition. Using an integrative perspective, the simultaneous influence of the peer variables and self-perceived social competence on loneliness was explored. When all variables were included in the model, friendship quality and social preference still had indirect effects on loneliness, in line with previous studies (Sun et al., 2009; Zhou et al., 2005). However, the indirect effect of proximity prestige was no longer apparent. The results suggest that there is a mediational pathway between some aspects of peer relations and children's loneliness. Specifically, it is possible that the poorer friendship quality or the lower social preference that children have, the more negatively they perceive their social competence, resulting in greater loneliness. These results are consistent with the cognitive processes perspective of loneliness, which holds that loneliness is the distressing feeling accompanying the perception that one's social needs are not being met by the quantity or especially the quality of one's social relationships, and that the experience of loneliness depends on the perception of deficiencies in one's own social relations (Hawkley & Cacioppo, 2010; Peplau & Perlman, 1982; Terrell-Deutsch, 1999).

There are two possible reasons that proximity prestige did not have an indirect influence on loneliness through the mediator of self-perceived social competence in the integrative model. First, the prestige index takes into account a different type of information of which the child may not be aware, namely the social connectedness of the child's friends and the friends' friends. It is likely difficult for children to assess the extent of indirect relationships in the social network and the impact of remote relationships on their own social standing. This would be even more difficult over time, as children's social circles may have changed, for instance, their social connections become more complex (a pattern seen in this study, in which proximity prestige scores increased from grades 3 to 6), and some children may transfer to or from the classes.

Second, prestige is measured based on friendships identified by others; as the reflection of a more objective measure of friendship status, children may be unaware of their prestige in the peer group. The results of the current study suggest that, although the indirect and the objective relationships in the peer network are out of children's perception, they can influence how lonely a child feels, independent of the effects of other aspects of peer relations in which the child is more directly involved.

It should be noted that the present study examined perceived social competence only as the mediator between peer relationships and loneliness based on cognitive processes theory of loneliness. Another possibility is that perceived social competence is itself an important predictor of loneliness, and that the link between perceived social competence and loneliness may be mediated by variables of peer experience. For example, perhaps as potential indicative of actual social competence, self-perceived social competence could influence the extent of children's success in peer relations, which could in turn negatively predict loneliness. One longitudinal study with a sample of Chinese children found that self-perception of social competence had positive and unique contribution to the prediction of social preference two years later (Chen, He, & Li, 2004). Social preference may mediate the association between self-perceived social competence and loneliness. Furthermore, longitudinal research is needed to clarify the relations among children's perceived competence, peer relationships and loneliness.

4.3 The integrative perspective on the research of peer relations

Considering children's peer relations are multifaceted, intersecting and overlapping systems, Gifford-Smith and Brownell (2003) suggested that researchers could be better informed of children's psychosocial development by attending to multiple aspects of the peer experiences simultaneously from an integative perspective. Based on

this perspective, the present study conceived and tested one conceptaul model of the links among peer experiences and children's emotional adjustment (loneliness). In this model, three kinds of peer relationships at the dyadic and group levels of peer experiences (friendship quality, social preference, and proximity presitge) were presumed to have simultaneous impacts on children loneliness through the mediator of individuals' cognition (self-perceived social competence). The pathways of the integrative model were compared with those of single peer relationship model. It was found that the indirect effect of proximity prestige on loneliness which was significant in the single peer relationship model was no longer significant in the integrative model. This result implied that the conclusions drawn at single levels of analysis may be not adaptive to the situations with multiple levels. Rubin et al. (2006) noted about children's peer experiences that: the analyses in each level of peer relations—interactions, relationships, groups—are scientifically legitimate and raise interesting questions. However, it has not always clearly demonstrated the important ways in which processes at one level are influenced by those at other levels. Sometimes there were limits for the conclusions drawn at the single levels of analysis. The integrative model of the present study took into consideration the correlations among the three different kinds of peer relationships, which are closer to children's real peer relations.

4.4 Strengths, limitations, and future research

The current study contributes to the literature on the relations between peer experiences and children's loneliness in four distinct ways. First, a social network index was introduced to assess the effect of indirect friends, a topic to which relatively less attention has been afforded in previous research. Proximity prestige reflects the quantity and the closeness of direct and indirect friendships (de Nooy et al., 2005), and it also demonstrates the objective status of a child's friends. Second, based

on the cognitive processes theory of loneliness, the present research explored multidimensional aspects of peer relations (friendship quality, social preference, and proximity prestige) in relation to children's loneliness, and an individual cognitive factor (self-perceived social competence) as a possible underlying mechanism responsible for these effects. The mechanism revealed by this research might be suggestive for formulating measures to prevent children from loneliness. Third, under the framework of an integrative and comprehensive perspective of peer experiences, the current study tested the simultaneous influence of the multiple aspects of peer relations on children's loneliness with individual's cognition as a mediator, and compared results to the pathways of the single peer relationship model. The pathways of the integrative model were different from that of the single peer relationship model. The results indicated that it was valuable and necessary to explore the integrative effects of the multiple aspects of peer experiences on children's social development and adjustment. Fourth, much of the previous work on the effects of peer experience on children's loneliness has been conducted within Western cultures, and the current study adds perspectives based on a representative sample from China, a country typified by collectivist culture. The meaning and influence of social interactions for Chinese children may be related to their cognitions of peer relations and loneliness in ways that would be less common in the West. Cross-culture research is a promising avenue for future work.

Several issues should be noted. First, the present study was conducted in a sample of Chinese children. Chinese culture places great value on relational bonds and group cohesion, and the self-evaluation that has been traditionally encouraged places more stress on social connections than the self-evaluation seen in Western cultures (Heine, 2001; Markus & Kitayama, 1991). The

neuroimaging research also provided evidence that the "Western self" was different from the Chinese self at a neural level (Zhu, Zhang, Fan, & Han, 2007). In addition, the children of this study were from an urban area of China. Thus, more investigation is needed to determine whether the results of the present study are generalizable to rural areas of China and to other countries and cultures, where most peer research has been conducted.

Second, due to the cross-sectional nature of the data, causal direction cannot be determined in the present study. The results can provide a foundation for future studies but cannot explain the causal mechanisms amongst peer relationships, children's cognition, and feeling of loneliness. Future research with longitudinal data and integrative perspectives is needed to study the dynamics of the relations among peer relationships, individual cognition and loneliness.

Third, the results indicated that there were significant gender differences on all five study variables. Boys felt lonelier than girls, had less positive self-perceived social competence, and had poorer peer relationships (lower friendship quality, social preference and proximity prestige). The analyses in this study focused on the general patterns of relations among the variables, without taking gender differences into consideration. Theory and research suggest that, compared to boys, girls are more involved in close relationships, receive greater emotional provisions in their friendships, more likely to adopt connection-oriented goals and pay greater concerns to social evaluation (for a review, see Rose & Rudolph, 2006). There may be gender differences in the patterns of relations among peer relationships, self-cognition (e.g., self-perceived social competence), and children's loneliness. Research on this topic would be benefit to provide process-oriented interpretation of gender differences in loneliness.

Another limitation with the analyses was that the present study did not take the nested nature of the data into consideration. The children were nested in eight classes and four grades, and there were significant grade differences in proximity prestige. Children's friendship networks may become more complex and closer with their increasing grades. Future studies could apply the analysis of multi-level models to provide more information about the effects of the class or grade.

The internal consistency of the Harter's Perceived Social Competence subscale was low (.56) in this study. Hoyle and Kenny (1999) showed that low reliability in the measure of the mediator leaded to underestimation of true mediation effect when analyzing observed variables. The simulation studies found that the bias associated with the mediation effect was approximately equal to the product of the reliabilities of two composite measures (the measures of the independent variable and the mediator; see Ledgerwood & Shrout, 2011). For instance, when the reliabilities of the two measures both were moderate (.70), the estimated effect was approximately half (.49) the size of the effect used to generate the data. It was supposed that the independent variable was measured without error, and the reliability of the measure of the mediator was .56, the estimated effect was approximately half (.56) the size of the real effect. The simulation studies also confirmed that the analysis of observed variables generally produces acceptable Type I error rates when there is in fact no real indirect effect (i.e., when pathway from independent variable to mediator and/or pathway from mediator to dependent variable is set to 0 in the simulation; see Ledgerwood & Shrout, 2011). Therefore, although the mediator was measured with low reliability in the present study, the results discovered in this study were credible. In addition, the reliability of the Harter's Perceived Social Competence subscale varied across studies in different cultural contexts. Other research in China with the same age group also found low reliability, with Cronbach's alphas of .67 and .64 (Chen, He, & Li, 2004; Stigler et

al., 1985). This subscale has also been found to have low reliability in Brazil (.50) and Italy (.57) samples (Chen, Zappulla,Coco, et al., 2004). However, the subscale has shown good reliability in North American samples (e.g., .80 with Canadian children and .75 to .84 with U.S. children; see Chen, Zappulla, Coco, et al., 2004; Harter, 1982). This range of values suggests that the internal consistency of the subscale was influenced by translation or cultural differences. Therefore, more efforts are needed to explore influences on the reliability of the Harter's Perceived Social Competence subscale, especially those due to cultural factors.

Finally, the test−retest correlations were moderate for the measures of loneliness (.42), self−perceived social competence (.48), and friendship quality (.40) over a one−year interval. The temporal stability of the measures decreases along with the interval increases. During the one−year interval, the classroom rosters may be changed across the school years. Meanwhile, the children in middle childhood may experience dramatically changes in peer relations (such as, the expansion of peer circle, increase in the number of friends) and cognitive abilities. These factors are all likely to lower the temporal stability of the measures. Future research would be informative to provide the test−retest correlations of the measures over shorter time intervals.

4.5 Implications for practice

Results from the present study may have important implications for enriching practical work in this domain. Hawkley and Cacioppo (2010) summarized interventions for loneliness, and concluded that there were four main types of interventions: (1) enhancing social skills, (2) providing social support, (3) increasing opportunities for social interaction, and (4) addressing maladaptive social cognition. A meta−analysis of loneliness reduction interventions revealed that among the randomized studies, interventions that addressed maladaptive social cognition

had a larger mean effect size compared to interventions that addressed social support, social skills, and opportunities for social intervention (Masi et al., 2011). The current study also found that a cognitive factor, self−perceived social competence, made a significant contribution to the variance in children's loneliness, and direct peer relations (friendship quality and social preference) were related to loneliness partially through self−perceived social competence.

As indicated by the current findings and the results of other research, it may be effective to prevent or diminish children's loneliness by changing maladaptive social cognitions such as sensitivity to social threats, preferential attention to negative social information, and negative social expectations. Meanwhile, in light of the cognitive level of children, it might be better to offer lonely children more opportunities for social interaction along with attempts to change their social cognition. A high quality friendship with a best friend, having friends who themselves have many friends, and acceptance by the peer group all appear beneficial for preventing the feelings of loneliness. Therefore, a three−tiered model is proposed to ameliorate social/emotional/behavioral concerns in order to prevent children's loneliness, and the improvements can be implemented in the classroom. Firstly, the teachers need to construct friendly atmosphere of peer relations, and encourage students to help each other. Secondly, based the effects of peer relations found in the present study, some peer−mediated interventions can be recommended. (a) In light of the finding that indirect friends might also make contributions to buffering children from feelings of loneliness, the teachers could guide highly prestigious children to befriend lonely classmates. This measure is beneficial to increase the friends of the lonely children, and also a way to directly and indirectly provide lonely children with more opportunities for peer interaction and potentially improved peer relations. (b) More class

activities involving all classmates can be organized to increase peer interaction and peer acceptance for lonely children. (c) Some social skills can be taught to children for making and boosting their friendships, such as learning how to join in an activity or game, communicating (e.g., talking and listening), cooperating (e.g., taking turns and sharing materials), validating or supporting (e.g., giving attention or help), resolving conflicts (Oden & Asher, 1977). Finally, individual help should be provided to lonely children for changing their maladaptive social cognitions. For instance, cognitive behavioral therapy (CBT), which targets maladaptive social cognitions is an effective approach for reducing loneliness. The lonely children can be taught to identify automatic negative thoughts and look for disconfirming evidence, to decrease biased cognitions, and/or to reframe perceptions of loneliness and personal control (McWhirter, 1990).

4.6 Conclusions

The present study provided valuable insights on the role of peer relations in understanding children's loneliness. Unlike previous research on loneliness, this study assessed peer relations at multiple levels of social complexity. A key finding was that children's self-perceived social competence predicted children's loneliness both directly and as a mediator of associations between other peer variables and loneliness. A unique contribution of this literature was the use of a proximity prestige index, which documented that "friends of friends" may act as a buffer against loneliness. Together, the results suggest possible pathways by which peer relations, assessed both subjectively and objectively, might contribute to children's feelings of loneliness, thus laying the groundwork for identifying and intervening with lonely children.

References

Asher, S. R., & Paquette, J. A. (2003). Loneliness and peer relations in childhood. *Current Directions in Psychological Science, 12*, 75–78. http://dx.doi.org/10.1111/1467–8721.01233

Asher, S. R., & Wheeler, V. A. (1985). Children's loneliness: A comparison of rejected and neglected peer status. *Journal of Consulting and Clinical Psychology, 53*, 500–505. http://dx.doi.org/10.1037/0022–006X.53.4.500

Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology, 48*, 5–37. http://dx.doi.org/10.1016/j.jsp.2009.10.001

Benner, A. D. (2011). Latino adolescents' loneliness, academic performance, and the buffering nature of friendships. *Journal of Youth and Adolescence, 40*, 556–567. http://dx.doi.org/10.1007/s10964–010–9561–2

Berguno, G., Leroux, P., McAinsh, K., & Shaikh, S. (2004). Children's experience of loneliness at school and its relation to bullying and the quality of teacher interventions. *The Qualitative Report, 9*, 483–499.

Boivin, M., Hymel, S., & Bukowski, W. M. (1995). The roles of social withdrawal, peer rejection, and victimization by peers in predicting loneliness and depressed mood in childhood. *Development and Psychopathology, 7*, 765–785. http://dx.doi.org/10.1017/S0954579400006830

Brownell, C. A., & Gifford-Smith, M. E. (2003). Context and development in children's school-based peer relations: Implications for research and practice. *Journal of School Psychology, 41*, 305–310. http://dx.doi.org/10.1016/s0022–4405(03)00052–9

Cillessen, A. H. N., & Marks, P. E. L. (2011). Conceptualizing and measuring popularity. In A. H. N. Cillessen, D. Schwartz, & L. Mayeux (Eds.), *Popularity in the peer system* (pp. 25–56). New York: Guilford Press.

Chamberlain, B., Kasari, C., & Rotheram–Fuller, E. (2007). Involvement or isolation? The social networks of children with autism in regular classrooms. *Journal of Autism and Developmental Disorders, 37*, 230–242. http://dx.doi.org/10.1007/s10803–006–0164–4

Chen, X. (2000). Growing up in a collectivistic culture: Socialization and socio–emotional development in Chinese children. *International perspectives on human development*. In A. L. Comunian & U. P. Gielen (Eds.), International perspectives on human development (pp. 331– 353). Lengerich, Germany: Pabst Science Publishers.

Chen, X., He, Y., & Li, D. (2004). Self–perceptions of social competence and self–worth in Chinese children: Relations with social and school performance. *Social Development, *, 570–589. http://dx.doi.org/10.1111/j.1467–9507.2004.00284.x

Chen, X., He, Y., De Oliveira, A. M., Coco, A. L., Zappulla, C., Kaspar, V., DeSouza, A. (2004). Loneliness and social adaptation in Brazilian, Canadian, Chinese and Italian children: A multi–national comparative study. *Journal of Child Psychology and Psychiatry, 45*, 1373–1384. http://dx.doi.org/10.1111/j.1469–7610.2004.00329.x

Chen, X., Rubin, K. H., & Sun, Y. (1992). Social reputation and peer relationships in Chinese and Canadian children: A cross–cultural study. *Child Development, 63*, 1336–1343. http://dx.doi.org/10.1111/j.1467–8624.1992.tb01698.x

Chen, X., Zappulla, C., Coco, A.L., Schneider, B., Kaspar, V., De Oliveira, A. M.,Bergeron, N. (2004). Self–perceptions of competence in Brazilian, Canadian, Chinese and Italian children: Relations with social and school adjustment. *International Journal of Behavioral Development, 28,* 129–138. http://dx.doi.org/10.1080/01650250344000334

Cheng, H., & Furnham, A. (2002). Personality, peer relations, and self–confidence as predictors of happiness and loneliness. *Journal of Adolescence, 25*, 327–339. http://dx.doi.org/ 10.1006/jado.2002.0475

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New Jersey: Erlbaum.

Coie, J. D., Dodge, K. A., & Coppotelli, H. (1982). Dimensions and types of social status: A cross–age perspective. *Developmental Psychology, 18*, 557–570. http://dx.doi.org/10.1037/0012–1649.18.4.557

Cole, D. A. (1991). Change in self–perceived competence as a function of peer and teacher evaluation. *Developmental Psychology, 27*, 682–688. http://dx.doi.org/10.1037/0012–1649.27.4.682

de Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with Pajek* (2nd ed.). New York: Cambridge University Press.

Ennett, S. T., Bauman, K. E., Hussong, A., Faris, R., Foshee, V. A., Cai, L., & DuRant, R. H. (2006). The peer context of adolescent substance use: Findings from social network analysis. *Journal of Research on Adolescence, 16*, 159–186. http://dx.doi.org/10.1111/j.1532–7795.2006.00127.x

Fairchild, A. J., & McQuillin, S. D. (2010). Evaluating mediation and moderation effects in school psychology: A presentation of methods and review of current practice. *Journal of School Psychology, 48*, 53–84. http://dx.doi.org/10.1016/j.jsp.2009.09.001

Faris, R., & Ennett, S. (2012). Adolescent aggression: The role of peer group status motives, peer aggression, and group characteristics. *Social Networks, 34*, 371–378. http://dx.doi.org/10.1016/j.socnet.2010.06.003

Fox, N. A., Henderson, H. A., Marshall, P. J., Nichols, K. E., & Ghera, M. M. (2005). Behavioral inhibition: Linking biology and behavior within a developmental framework. *Annual Review of Psychology, 56*, 235–262. http://dx.doi.org/10.1146/annurev.psych.55.090902.141532

Gest, S. D., Graham–Bermann, S. A., & Hartup, W. W. (2001). Peer experience: Common and unique features of number of friendships, social network centrality, and sociometric status. *Social Development, 10*, 23–40. http://dx.doi.org/10.1111/1467–9507.00146

Gifford–Smith, M. E., & Brownell, C. A. (2003). Childhood peer relationships: Social acceptance, friendships, and peer networks. *Journal of School Psychology, 41*, 235–284. http://dx.doi.org/10.1016/S0022–4405(03)00048–7

Harris, R. A., Qualter, P., & Robinson, S. J. (2013). Loneliness trajectories from middle childhood to pre–adolescence: Impact on perceived health and sleep disturbance. *Journal of Adolescence, 36*, 1295–1304. http://dx.doi.org/10.1016/j.adolescence.2012.12.009

Harter, S. (1982). The perceived competence scale for children. *Child Development, 53*, 87–97. http://dx.doi.org/10.1111/j.1467–8624.1982.tb01295.x

Hawkley, L. C., & Cacioppo, J. T. (2010). Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Annals of Behavioral Medicine, 40,* 218–227. http://dx.doi.org/10.1007/s12160–010–9210–8

Heine, S. J. (2001). Self as cultural product: An examination of East Asian and North American selves. *Journal of Personality, 69*, 881–905. http://dx.doi.org/10.1111/1467–6494.696168

Heinrich, L. M., & Gullone, E. (2006). The clinical significance of loneliness: a literature review. *Clinical Psychology Review, 26*, 695–718. http://dx.doi.org/10.1016/j.cpr.2006.04.002

Hinde, R. A. (1987). Individuals, relationships and culture: *Links between ethology and the social sciences*. New York: Cambridge University Press.

Hornby, A. S. (2010). *Oxford advanced learner´s dictionary* (8th ed.). Oxford: Oxford University Press.

Hoyle, R. H., & Kenny, D. A. (1999). Sample size, reliability, and tests of statistical mediation. In R. H. Hoyle (Ed.), *Statistical strategies for small sample research* (pp. 195–222). Thousand Oaks, CA: Sage.

Hoza, B., Bukowski, W. M., & Beery, S. (2000). Assessing peer network and dyadic loneliness. *Journal of Clinical Child Psychology, 29*, 119–128. http://dx.doi.org/10.1207/S15374424jccp2901_12

Hymel, S., Vaillancourt, T., McDougall, P., & Renshaw, P. D. (2002). Peer acceptance and rejection in childhood. In P. K. Smith & C. H. Hart (Eds.), *Blackwell handbook of childhood social development* (pp. 265–284). Malden: Blackwell.

Jobe–Shields, L., Cohen, R., & Parra, G. R. (2011). Patterns of change in children's loneliness: Trajectories from third through fifth grades. *Merrill–Palmer Quarterly, 57*, 25–47.

Ladd, G. W. (1999). Peer relationships and social competence during early and middle childhood. *Annual Review of Psychology, 50*, 333–359. doi: 10.1146/annurev.psych.50.1.333

Ladd, G. W., Buhs, E. S., & Troop, W. (2002). Children's interpersonal skills and relationships in school settings: Adaptive significance and implications for school–based prevention and intervention programs. In K. Smith & C. H. Hart (Eds.), *Blackwell handbook of childhood social development* (pp. 394–415). Malden: Blackwell.

Ladd, G. W., Kochenderfer, B. J., & Coleman, C. C. (1997). Classroom peer acceptance, friendship, and victimization: Destinct relation systems that contribute uniquely to children's school adjustment? *Child Development, 68*, 1181–1197. http://dx.doi.org/10.2307/1132300

Ledgerwood, A., & Shrout, P. E. (2011). The trade–off between accuracy and precision in latent variable models of mediation processes. *Journal of personality and social psychology, 101*, 1174–1188. http://dx.doi.

org/10.1037/a0024776

Lin, N. (1976). *Foundations of Social Research*. New York: McGraw-Hill.

Luo, G. (1996). Chinese traditional social and moral ideas and rules. Beijing, China: *The University of Chinese People Press*.

Margalit, M. (2010). *Lonely children and adolescents*. New York: Springer.

Marks, P. E. L., Babcock, B., Cillessen, A. H. N., & Crick, N. R. (2012). The effects of participation rate on the internal reliability of peer nomination measures. *Social Development, 22*, 609–622. http://dx.doi.org/10.1111/j.1467–9507.2012.00661.x

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224–253. http://dx.doi.org/10.1037/0033–295X.98.2.224

Masi, C. M., Chen, H. Y., Hawkley, L. C., & Cacioppo, J. T. (2011). A meta-analysis of interventions to reduce loneliness. *Personality and Social Psychology Review, 15*, 219–266. doi: http://dx.doi.org/10.1177/1088868310377394

McWhirter, B. T. (1990). Loneliness: A review of current literature, with implications for counseling and research. *Journal of Counseling & Development, 68*, 417–422. http://dx.doi.org/10.1002/j.1556–6676.1990.tb02521.x

Mercken, L., Steglich, C., Sinclair, P., Holliday, J., & Moore, L. (2012). A longitudinal social network analysis of peer influence, peer selection, and smoking behavior among adolescents in British schools. *Health Psychology, 31*, 450–459. http://dx.doi.org/10.1037/a0026876

Mouratidis, A. A., & Sideridis, G. D. (2009). On social achievement goals: Their relations with peer acceptance, classroom belongingness, and perceptions of loneliness. *The Journal of Experimental Education, 77*, 285–308. http://dx.doi.org/10.3200/JEXE.77.3.285–308

Mundt, M. P. (2011). The impact of peer social networks on adolescent alcohol use initiation. *Academic Pediatrics, 11*, 414–421. http://dx.doi.org/10.1016/j.acap.2011.05.005

Muthén, L. K. , & Muthén, B. O. (1998–2012). *Mplus user´s guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.

Nangle, D. W., Erdley, C. A., Newman, J. E., Mason, C. A., & Carpenter, E. M. (2003). Popularity, friendship quantity, and friendship quality: Interactive influences on children's loneliness and depression. *Journal of Clinical Child and Adolescent Psychology, 32*, 546–555. http://dx.doi.org/10.1207/S15374424JCCP3204_7

Oden, S., & Asher, S. R. (1977). Coaching children in social skills for friendship making. *Child Development, 48*, 495–506. http://dx.doi.org/10.2307/1128645

Okamoto, J., Johnson, C. A., Leventhal, A., Milam, J., Pentz, M. A., Schwartz, D., & Valente, T. W. (2011). Social network status and depression among adolescents: An examination of social network influences and depressive symptoms in a Chinese sample. *Research in Human Development, 8,* 67–88. http://dx.doi.org/10.1080/15427609.2011.549711

Oyserman, D., Coon, H. M., & Kemmelmeier, M. (2002). Rethinking individualism and collectivism: Evaluation of theoretical assumptions and meta-analyses. *Psychological Bulletin, 128*, 3–72. http://dx.doi.org/10.1037/0033–2909.128.1.3

Parker, J. G., & Asher, S. R. (1993). Friendship and friendship quality in middle childhood: Links with peer group acceptance and feelings of loneliness and social dissatisfaction. *Developmental Psychology, 29*, 611–621. http://dx.doi.org/10.1037/0012–1649.29.4.611

Parker, J. G., & Seal, J. (1996). Forming, losing, renewing, and replacing friendships: Applying temporal

parameters to the assessment of children's friendship experiences. *Child Development, 67*, 2248–2268. http://dx.doi.org/10.1111/j.1467–8624.1996.tb01855.x

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling, 8*, 287–312. http://dx.doi.org/10.1207/S15328007SEM0802_7

Peplau, L. A., & Perlman, D. (1982). Perspectives on loneliness. In L. A. Peplau & D. Perlman (Eds.), Loneliness: *A sourcebook of current theory, research and therapy* (pp. 1–18). New York: Wiley.

Qualter, P., Brown, S. L., Munn, P., & Rotenberg, K. J. (2010). Childhood loneliness as a predictor of adolescent depressive symptoms: An 8–year longitudinal study. *European Child & Adolescent Psychiatry, 19,* 493–501. http://dx.doi.org/10.1007/s00787–009–0059–y

Rokach, A., & Bauer, N.. (2004). Age, culture, and loneliness among Czechs and Canadians. *Current Psychology, 23*, 3–23. http://dx.doi.org/10.1007/s12144–004–1005–2

Rose, A. J., & Rudolph, K. D. (2006). A review of sex differences in peer relationship processes: Potential trade–offs for the emotional and behavioral development of girls and boys. *Psychological Bulletin, 132,* 98–131. http://dx.doi.org/10.1037/0033–2909.132.1.98

Rotenberg, K. J. (1999). Childhood and adolescent loneliness: An introduction. In K. J. Rotenberg & H. Shelley (Eds.), *Loneliness in childhood and adolescence* (pp. 3–8). Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511551888.001

Rubin, K. H., Bukowski, W. M., & Parker, J. G. (2006). Peer interactions, relationships, and groups. In N. Eisenberg (Ed.), Handbook of child psychology, *Vol.3: Social, emotional, and personality development* (6th ed., pp. 571–645). Hoboken, NJ: Wiley. http://dx.doi.

org/10.1002/9780470147658.chpsy0310

Scott, J. (2000). Social network analysis: *A handbook.* London: Sage Publications Ltd.

Seepersad, S., Choi, M. K., & Shin, N. (2008). How does culture influence the degree of romantic loneliness and closeness? The Journal of Psychology: *Interdisciplinary and Applied, 142,* 209–220. http://dx.doi.org/10.3200/JRLP.142.2.209–220

Sheridan, S. M., Buhs, E. S., & Warnes, E. D. (2003). Childhood peer relationships in context. *Journal of School Psychology, 41*, 285–292. http://dx.doi.org/10.1016/s0022–4405(03)00049–9

Shin, Y. (2007). Peer relationships, social behaviours, academic performance and loneliness in Korean primary school children. *School Psychology International, 28*, 220–236. http://dx.doi.org/10.1177/0143034307078103

Sletta, O., Val?s, H., Skaalvik, E., & S?bstad, F. (1996). Peer relations, loneliness, and self–perceptions in school–aged children. *British Journal of Educational Psychology, 66*, 431–445. http://dx.doi.org/10.1111/j.2044–8279.1996.tb01210.x

Stigler, J. W., Smith, S., & Mao, L–W. (1985). The self–perception of competence by Chinese children. *Child Development, 56,* 1259–1270. http://dx.doi.org/10.2307/1130241

Sun, X., Zhou, Z., Fan, C., & Ke, S. (2009). Loneliness in middle childhood and its relation to multi–level peer experience. *Psychological Science (China), 32,* 567–570.

Terrell–Deutsch, B. (1999). The conceptualization and measurement of childhood loneliness. In K. J. Rotenberg & S. Hymel (Eds.), *Loneliness in childhood and adolescence* (pp. 11–33). New York: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511551888.002

Terry, R. (2000). Recent advances in measurement theory

and the use of sociometric techniques. *New Directions for Child and Adolescent Development*, 2000(88), 27–53. http://dx.doi.org/10.1002/cd.23220008805

Uruk, A. C., & Demir, A. (2003). The role of peers and families in predicting the loneliness level of adolescents. The Journal of Psychology: *Interdisciplinary and Applied, 137,* 179–193. http://dx.doi.org/10.1080/00223980309600607

van Dulmen, M. H. M., & Goossens, L. (2013). Loneliness trajectories. *Journal of Adolescence, 36*, 1247–1249. http://dx.doi.org/10.1016/j.adolescence.2013.08.001

Vanhalst, J., Goossens, L., Luyckx, K., Scholte, R. H.J., & Engels, R. C.M.E. (2012). The development of loneliness from mid– to late adolescence: Trajectory classes, personality traits, and psychosocial functioning. *Journal of Adolescence, 36*, 1305–1312. http://dx.doi.org/10.1016/j.adolescence.2012.04.002

Wassermann, S., & Faust, K. (1994). Social network analysis: *Methods and applications.* New York: Cambridge University Press.

Weiss, R. S. (1973). Loneliness: *The experience of emotional and social isolation.* Cambridge, MA: MIT Press.

Yu, G., Zhang, Y., & Yan, R. (2005). Loneliness, peer acceptance, and family functioning of Chinese children with learning disabilities: Characteristics and relationships. *Psychology in the Schools, 42,* 325–331. http://dx.doi.org/10.1002/pits.20083

Zhou, Z., Sun, X., Xiang, Y., & Liu, J. (2007). Peer interaction and loneliness in middle childhood: A cross–lagged analysis. *Psychological Science (China), 30*, 1309–1313.

Zhou, Z., Sun, X., Zhao, D., & Hsueh, Y. (2005). The test of the mediator variable between peer relationship and loneliness in middle childhood. *Acta Psychologica Sinica (China), 37*, 776–783.

Zhou, Z., Zhao, D., Sun, X., & Ding, X. (2006). Children's experiences with peers and loneliness: A two–year longitudinal study. *Acta Psychologica Sinica (China), 38*, 743–750.

Zhu, Y., Zhang, L., Fan, J., & Han, S. (2007). Neural basis of cultural influence on self–representation. *Neuroimage, 34*, 1310–1316. http://dx.doi.org/10.1016/j.neuroimage.2006.08.047

心理辅导研究所团队成员

江光荣，博士，教授，华中师范大学心理辅导研究所所长，主要研究领域为心理咨询与治疗的过程与效果、青少年自杀与自伤行为、国民心理健康素养与心理求助行为、学校心理健康教育。

陶嵘，博士，副教授，华中师范大学心理辅导研究所副所长，主要研究领域为情绪异常的心理生理机制及心理治疗，团体心理治疗、中国文化下的心理治疗、梦的工作等。

吴才智，博士，副教授，主要研究领域为心理健康教育、心理辅导的组织与管理、危机干预与自杀预防。

夏勉，博士，副教授，主要研究领域为中国人的求助行为与助人行为、心理咨询与治疗。

孙启武，博士，副教授，主要研究领域为基于依恋理论的心理病理与治疗改变机制、创伤心理学、心理咨询与治疗中的文化维度、心理咨询与治疗中的评估与诊断及咨询伦理等。

朱旭，博士，讲师，主要研究领域为心理咨询的过程与效果、咨询关系、团体治疗。

于丽霞，博士，讲师，主要研究领域为心理咨询与治疗、青少年自杀与自伤行为。

段文婷，硕士，讲师，主要研究领域为心理健康教育、心理咨询与治疗、大学生求助决策。

暴力电子游戏的短期脱敏效应：两种接触方式比较 *

郭晓丽[1,2]，江光荣[1]，朱　旭[1]

([1] 华中师范大学心理学院 , 人的发展与心理健康湖北省重点实验室 , 武汉 430079)

([2] 华中科技大学学生工作处 , 武汉 430074)

摘　要　比较主、被动接触暴力电子游戏的脱敏效应 , 以 44 名男性大学生为被试 , 利用生物反馈仪测量被试主动参与游戏或被动观看游戏录像前后 , 及随后观看暴力视频过程中皮电与心率的变化 (脱敏效应的生理指标)。结果表明 :(1) 暴力电子游戏可以产生脱敏效应。接触游戏 15 分钟后 , 暴力游戏组观看暴力视频过程中皮电的增加值明显小于非暴力游戏组 ;(2) 游戏的接触方式对于脱敏效应的程度无显著影响 , 但主动参与组对于游戏内容知觉到更高的愉快与更低的沮丧。

关键词　暴力电子游戏 ; 脱敏效应 ; 接触方式

分类号　R395

1　问题提出

人们相信娱乐媒体中的暴力对于现实生活中的攻击行为有不可推卸的责任。美国 3 个科学委员会的调查——外科综合委员会 (SGC) 报告 , 国家心理卫生研究院 (NIMH)10 年跟踪 , 以及美国心理学会暴力与青少年委员会调查——断定观看暴力可以增加攻击行为 (Mullin & Daniel, 1995)。

日常生活中 , 人们通过电影、电视等基于屏幕的媒体接触到的暴力远远大于在真实生活中接触到的暴力。过去 , 电视与电影是主要的基于屏幕的媒体 , 接触方式主要是观看 , 但是当游戏出现后 , 人们接触暴力的方式就从"观看"发展到"参与"了。在过去 30 年中 , 电子游戏产业发展迅速。伴随着游戏产业的暴发式增长 , 公众对于接触电子游戏的负面效应

的担忧也与日俱增。担忧的原因之一是现有游戏中暴力内容的普遍性 (VanMierlo & VandenBulck, 2004)。一项对于游戏内容的分析表明 , 89% 的游戏含有部分暴力内容 , 50% 的游戏含有对于其他人物的严重暴力内容 (Dill & Dill, 1998)。这个百分比足以证明"暴力"是游戏的一个普遍特点。另一个原因是大量研究揭露接触暴力媒体的确有着消极效应。一项元分析研究发现接触暴力电子游戏导致攻击行为、攻击情感、攻击认知与生理唤起的增加 , 以及亲社会行为的减少 (Anderson & Bushman 2001)。美国心理学会暴力与青少年委员会还指出 , 攻击行为的增加不是观看暴力的唯一消极后果 , 观看暴力还可能会增加对于暴力的脱敏 , 导致对于暴力麻木的态度 , 从而降低暴力发生时采取行动帮助受害者的可能性 (Mullin & Daniel, 1995)。

脱敏 (desensitization) 一词的概念源于临床心理学中的系统脱敏治疗，指反复面对一个会导致焦虑、恐惧的刺激时，焦虑、恐惧等负面情绪反应逐渐消退的现象。暴力脱敏则特指持续暴露于暴力刺激时的情绪反应钝化现象。这种脱敏在生理上通常表现为皮肤电、心率的降低或血流量的减少。在有关媒体暴力脱敏的研究中，以上3种指标被公认为是反映脱敏的有效生理指标。脱敏并非简单的生理变化，它伴随着一系列认知与行为改变。暴力脱敏效应不仅可以导致暴力行为的增加，还有可能减少助人行为。暴力脱敏效应还可能是导致攻击人格形成的重要途径之一 (Carnagey, Anderson, & Bushman,2007)。

Anderson 和 Bushman(2002) 提出了一般攻击模型 (General Aggression Model) 来解释暴力媒体对于攻击行为的影响机制。根据一般攻击模型，个体变量 (例如：敌意特质、对暴力的态度) 和情境变量 (例如：媒体暴力) 交互影响个体的内部状态，认知、情感与生理唤起彼此作用，整体上影响个体对攻击行为的解释，并进一步影响后续的决策过程与行为表现 (Anderson, Carnagey, & Eubanks,2003;Gentile & Anderson, 2003;Kirsh, 2003)。一般攻击模型指出，暴力电子游戏对于攻击行为可以产生短期与长期效应 (Anderson, 2004; Carnagey & Anderson, 2005; Gentile & Lynch,2004)。在短期效应中，游戏作为情境变量导致攻击认知、生理唤起水平的提高，以及攻击情感的增加。在长期效应中，游戏通过促进形成攻击信念和攻击态度，促进产生攻击图式、攻击行为与攻击期望，以及降低个体对攻击的敏感性来影响攻击行为。相应的，这些因素也促进了个体攻击性人格的发展 (Anderson & Griffiths, 2004)。

一般攻击模型提供了一个有用的社会——认知框架来理解脱敏过程。人类对于暴力内容的原始反应是害怕与焦虑，当暴力刺激在积极情绪反应伴随下重复呈现时，这种原始的焦虑反应会由于抗条件作用 (counterconditioning) 而减弱。暴力脱敏发生的表现之一是个体重复接触媒体暴力后，在接触真实暴力时生理唤起水平降低。一旦脱敏发生，真实暴力就不再能引发原始的恐惧与焦虑。个体在认知上也发生相应的改变。例如，可能更少注意到攻击事件，知觉到较少或较轻的伤害，对于暴力受害者的同情降低等。

国外早在六七十年代就已经有人开始研究电视或电影暴力内容的脱敏效应。早先关于暴力媒体潜在脱敏效应的研究以被试的皮肤电为指标，发现被试对于电影中流血场面的观看降低了之后对于类似场面的生理唤起，表现出脱敏反应 (Cline & Croft,1973;Thomas & Horton, 1977)。同时，相关研究的结果也表明，过去对暴力影视的接触与所呈现的暴力视频的脱敏效应有正相关 (Funk & Baldacci,2004;Funk & Bushman, 2003)。

2000 年之后，开始出现少量针对暴力电子游戏脱敏效应的研究。相关研究的结果表明过去对于游戏暴力的接触与低同情、高亲暴力态度有关。而实验研究的结果表明暴力电子游戏不但可以导致对于媒体暴力的脱敏，对于真实暴力同样可以表现出脱敏。以往反映脱敏效应的指标为问卷或皮电、心率等传统生理指标，最近的研究开始尝试利用事件相关电位分析脱敏发生的过程 (Bartholow, Bushman, & Sestir, 2006)。

暴力媒体的负面效应也开始引起国内研究者的重视，目前国内的相关文献一部分集中于暴力媒体影响机制的理论探讨 (辛自强，池丽萍，2004；郑宏明，孙延军，1997)，少量的实证研究则关注媒体暴力对内隐攻击认知的影响 (陈美芬，陈舜蓬，2005; 崔丽娟，胡海龙，吴明证，2006)。虽有一些类似的提法 (张镇，刘月霞，张建新，2006)，但没有看到暴力媒体情绪脱敏效应的实证研究。

与传统的影视媒体相比，电子游戏作为一种新型媒体具有新的特点：参与性与互动性。游戏接触者会与虚拟的对手发生互动，做出暴力行为，而不仅仅是被动地观看暴力，因此在游戏过程中注意与情感卷入的水平可能更高，对于暴力角色有更多的认同，暴力行为也得到更有力的强化，因此玩暴力游戏对于他们的负面影响可能比观看暴力影视更为严重。

这种"主动参与"与"被动接受"的区别已经有不少研究者提及 (Anderson, Berkowitz, & Donnerstein,

2003; Anderson & Bushman, 2002;Anderson & Carnagey, 2004;Griffths, 1999), 但到目前为止, 还缺乏有力的实证比较。这里可能存在一个技术上的困难, 即保证"主动参与暴力游戏"与"被动观看暴力游戏"除自变量外无关变量 (如画面内容、暴力程度、图像质量等) 的等值性。

综观国内外对于暴力媒体的研究, 可以发现存在这样一些特点及问题 :(1) 相关研究设计较多 (探索暴力媒体的暴露量与另一些变量如攻击认知、情感与行为的相关), 缺乏对改变机制 (如脱敏效应) 的实验研究 ;(2) 脱敏效应的研究多集中于暴力影视, 针对暴力电子游戏的研究较少 ;(3) 有关暴力电子游戏由于主动参与的特点而有更严重脱敏效应的看法仅仅是理论假设, 未经实证检验。

本研究聚焦于暴力电子游戏所导致的短期心理脱敏效应, 并加入"暴露方式"这一新的自变量 ("主动参与"对"被动接受"), 以实验研究设计检验有关暴力游戏脱敏效应的以下假设 :

H1: 暴力电子游戏能够产生脱敏效应———接触游戏 15 分钟之后, 暴力游戏组观看视频过程中生理指标的增加值显著低于非暴力游戏组。

H2: 主动参与暴力游戏的脱敏效应程度大于被动接受暴力游戏———主动参与组观看视频过程中生理指标的增加值显著低于被动接受组。

假设一着眼于暴力电子游戏的暴力内容所产生的即时效应, 如果得到验证则说明暴力电子游戏的心理后效符合一般攻击模型, 与已有的暴力影视的效应有共同的理论依据。假设二则基于暴力接触方式, 探索"主动参与式"与"被动接受式"暴露于暴力媒体心理效应的区别。

2　研究方法

2.1　被试

通过广告招募大学生被试, 为控制性别变量, 被

_____

[1] 被试的身体攻击水平可能对自变量的效应有影响。

试全部为男性。要求被试有一定电脑使用的经验。共有 44 名学生参加实验, 年龄从 19 到 26 岁, 平均年龄为 21 岁。被试被随机分配到各实验组。实验结束后付给被试一定的报酬。

2.2　测量问卷

2.2.1　Buss-Perry 量表 Buss-Perry 量表 (Buss-Perry Scale) 包含 4 个组成部分 : 身体攻击分量表, 言语攻击分量表, 愤怒分量表与敌意分量表 (Buss & Perry, 1992)。本研究使用了其中的身体攻击 (Physical Aggression) 分量表测量被试的身体攻击性水平。该分量表包含 9 个条目, 7 点记分, α 系数为有影响, [1] 因此该量表得分作为协变量进入方差分析。

2.2.2　游戏使用习惯问卷 游戏使用习惯问卷 (Video Game Questionnaire) 要求被试列出他们最喜欢的五种游戏, 然后就每种游戏的内容与画面的暴力程度以及他们玩游戏的频率评定等级, 用以反映被试过去对暴力游戏的接触程度 (Anderson & Dill, 2000)。被试的游戏使用习惯可能会影响自变量的效应, 因此所得结果作为协变量进入方差分析。

2.2.3　游戏评定量表 游戏评定量表 (Video Game Rating Sheet) 要求被试对所接触游戏的画面暴力水平、内容暴力水平、动作快慢水平、游戏激动水平、游戏沮丧水平、游戏愉快水平、游戏难度水平进行评定 (Anderson & Dill, 1986)。该量表为 7 点记分, α 系数为 0.78。被试在该量表上的得分反映被试对于游戏的感知, 并作为自变量操作有效性的证据。具体就本研究设计来说, 如果暴力接触组和非暴力接触组对游戏画面暴力水平和内容暴力水平的评价有差别且前者高于后者, 说明对媒体暴力程度的差别性操作有效 ; 如果主、被动接触组在画面暴力水平和内容暴力水平的评价上无差异, 说明对媒体暴力程度的等值性操作有效。

2.3　实验材料

2.3.1　暴力电子游戏　游戏的选择遵循以下标准 : 难度适中 (指经过 5 分钟的练习之后, 有一定电脑使用经验的人都能够顺利玩游戏)、画面逼真、游戏人物为人类、暴力为主要特征 (而非恐怖或色情)、游

戏进度与难度可控。根据以上标准，通过试玩比较，认为《喋血街头Ⅱ》(单机版角色扮演动作射击类游戏)最符合以上标准，将其用作暴力游戏的实验材料。

2.3.2 非暴力电子游戏 非暴力游戏的选择关键在于避免游戏中有任何形式的暴力行为，同时考虑到动态抓屏软件的运行效果，最后选定一款弹球休闲游戏作为实验材料。以往的研究中也曾使用过这类游戏 (Anderson, Carnagey, & Flanagan, 2007;Carnagey, Anderson, & Bushman, 2004)。

2.3.3 暴力视频 选取电影《杀死比尔Ⅰ》10分钟的片段，该片断包含十分血腥的暴力镜头。该影片定级为R级。该视频资料呈现给所有被试，作为引发情绪唤起的暴力刺激源。

2.4 实验设备

2.4.1 生物反馈仪 该套多参数生物反馈仪由加拿大 Thought 公司生产，品牌为 ProCompInfiniti。

2.4.2 动态抓屏软件 利用虚拟视频软件 V3.0.2(CamtasiaStudioV3.0.2) 录制主动组被试玩游戏的全过程，保存为视频文档作为被动组的实验材料。

2.5 实验设计

实验采用 2×2 完全随机设计，包括两个自变量：游戏暴力程度 (暴力游戏 VS 非暴力游戏)、游戏接触
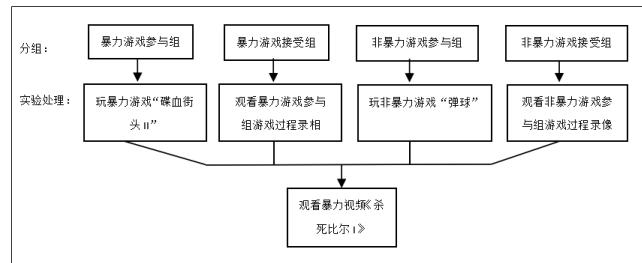
图 1 实验设计

方式 (主动参与游戏 VS 被动接受游戏录像)。因变量为被试的皮电与心率值。具体实验设计见图 1。

游戏接触方式是在以往研究的基础上新加入的变量，分为主动参与组与被动接受组。主动参与组是指被试自己亲身玩游戏，被动接受组是被试观看主动参与组游戏的录像。即一位主动参与组的被试在玩游戏时，将其游戏过程用软件 (CamtasiaStudioV3.0.2) 录制并保存下来，而这段游戏的录像就作为被动接受组

某位被试的实验材料。这样安排是为了使主动组与被动组的实验材料在画面内容与质量上等值，从而实现二组接触的暴力程度等值。经比较，两组被试对于各自游戏暴力水平的主观评定结果无差异 (详见 3.1)。

2.6 实验程序

被试个别进行实验，被告知研究目的是评估不同类型的媒体。具体程序包括以下几步：

（1）签订协议书、填写问卷：协议书中告知被试实验内容与注意事项，之后被试完成《身体攻击分量表》与《游戏使用习惯问卷》。实验结束后，被试完成《游戏评价问卷》。

（2）测量基线：进行 5 分钟心率与皮电的基线测量。

（3）接触游戏：将被试随机分配到四种实验条件中，接触游戏的时间为 15 分钟 (Anderson, Carnagey, & Flanagan, 2007)。为控制被试对于游戏的熟练程度，游戏的难度都设置为最低。

（4）后测：所有被试再次进行 5 分钟心率与皮电的测量。

（5）观看视频：所有被试观看《杀死比尔Ⅰ》的 10 分钟片段。在观看过程中，持续测量被试的心率与皮电。

2.7 统计方法

采用 SPSS11.5 进行描述性统计与方差分析。

3 研究结果

3.1 游戏评价问卷结果

2×2 随机方差分析的结果表明，对于游戏画面暴力水平的评定，游戏暴力程度的主效应显著，$F(1, 40)=21.20, p<0.001$; 对于游戏内容暴力水平的评定，游戏暴力程度的主效应显著，$F(1, 40)=28.47, p<0.001$; 对于游戏的沮丧水平，游戏暴力程度的主效应显著，$F(1, 40)=13.24, p<0.05$, 接触方式的主效应显著，$F(1, 40)=13.24, p<0.05$; 对于游戏的愉快程度，游戏暴力程度的主效应显著，$F(1, 40)=4.32, p<0.05$, 接触方式的主效应显著，$F(1, 40)=16.45, p<0.001$; 对于游戏的难度水

表 1　各实验组游戏评价问卷项目的平均数与标准差

问卷项目	暴力游戏参与组		暴力游戏接受组		非暴力游戏参与组		非暴力游戏接受组	
	M	SD	M	SD	M	SD	M	SD
画面暴力水平	5.45	1.70	5.27	1.74	2.82	2.27	2.55	1.97
内容暴力水平	5.45	1.29	5.45	1.70	2.64	2.11	2.55	1.92
动作快慢水平	4.55	1.64	3.82	1.78	3.82	1.66	3.00	1.48
游戏激动水平	3.36	1.57	3.36	1.50	3.91	1.76	2.91	1.64
游戏沮丧水平	3.45	1.70	5.36	1.36	2.00	1.10	3.45	1.86
游戏愉快水平	3.55	1.64	2.00	1.41	4.85	1.60	2.64	1.43
游戏难度水平	1.82	1.25	3.64	1.21	2.00	1.00	2.36	1.29

注：每组被试均为 11 人。

平，接触方式的主效应显著，F(1, 40)=9.23, p<0.05。所有交互作用均不显著。

以上结果表明，被试对暴力游戏的暴力与血腥程度的评价均显著高于非暴力游戏，说明实验对于暴力程度这一自变量的操作是有效的。除此之外，暴力游戏与非暴力游戏在沮丧程度与愉快程度上也有显著差异。在后续的分析中，将这两个评分作为协变量进行方差分析，结果显示无显著效应。

对于另一个自变量———游戏的接触方式，发现对于游戏画面与内容暴力水平的评定，主动参与组与被动接受组无显著差异，这说明对实验材料在暴力程度上等值性的操作是有效的。结果显示主动参与组与被动接受组在游戏难度、愉快程度与沮丧程度三个维度上均有显著差异。主动参与组认为游戏更简单，感受到更低的沮丧和更高的愉快。这说明对于相同的游戏内容，不同的接触方式会使得被试的感受有所区别，主动参与者对游戏过程有更积极的感受。

3.2 皮电结果

3.2.1 描述性统计　以基线值作为因变量，以接触方式和暴力程度作为自变量，进行 2 ×2 随机方差分析，主效应和交互效应均不显著，说明四个组别的基线值在统计上无显著差异。非暴力游戏接受组的标准差较高，则可能与样本较小或该组实验刺激的单调

性有关，个别被试有可能没有集中注意于实验刺激，而是受到其它心理活动的影响，造成数据的不稳定。

3.2.2　皮电值随时间的变化趋势　以皮电为因变量，进行 2(游戏暴力程度：暴力，非暴力）×2(接触方式：主动参与，被动接受）×3(测量时间：基线阶段，后测阶段，观看视频阶段) 三因素混合设计方差分析。游戏暴力程度与接触方式为组间变量，测量时间为组内变量。结果表明，三向交互作用不显著，测量时间与游戏暴力程度的两向交互作用显著，F(2, 40)=4.54, p<0.05, 测量时间的主效应显著，F(2, 40)=14.68, p<0.001。

由于测量时间与暴力程度的两向交互作用显著，进一步考查其简单效应，结果表明：暴力情境下，游戏后的皮电值显著低于基线阶段与视频观看阶段，而基线阶段与视频观看阶段无显著差异 (M 基线 =6.09, M 后测 =4.85, M 观看视频 =6.90); 非暴力情境下，视频观看阶段的皮电值显著高于基线阶段与游戏后阶段，而游戏前后无显著差异 (M 基线 =4.89, M 后测 =5.77, M 观看视频 =7.33)。

3.2.3　暴力电子游戏的脱敏效应　用视频观看阶段与基线阶段皮电值的差值作为因变量———脱敏发生的指标，以身体攻击性分量表与游戏使用习惯问卷得分作为协变量，进行 2(游戏暴力程度：暴力，非暴力）×2(接触方式：主动参与，被动接受) 随机方差分析。

表 2　皮电值的平均值与标准差

实验处理	基线		后测		观看视频	
	M	SD	M	SD	M	SD
暴力游戏参与组	6.04	3.53	5.24	3.11	6.93	3.66
非暴力游戏参与组	3.75	1.69	4.04	2.50	5.42	2.96
暴力游戏接受组	6.14	3.55	4.47	2.71	6.87	2.85
非暴力游戏接受组	6.04	8.82	7.49	12.86	9.23	12.13

结果表明个体身体攻击水平与游戏使用习惯均没有任何效应显著，即这两个个体变量对于自变量的效应均无影响。游戏暴力程度与接触方式无交互作用，$F_{(1, 40)}=1.11$, $p>0.05$。游戏暴力程度的主效应显著（$M_{暴力}=0.81$, $M_{非暴力}=2.43$, $F_{(1, 40)}=4.19$, $p<0.05$），说明暴力组在观看暴力视频的过程中皮电的增加值明显小于非暴力组，表现出更低的生理唤起。H1得到验证，即暴力电子游戏可以产生脱敏效应。游戏接触方式的主效应不显著（$M_{主动}=1.28$, $M_{被动}=1.97$, $F_{(1, 40)}=0.75$, $p>0.05$），主动参与组与被动接受组在观看视频的过程中皮电的增加值没有差异，H2没有得到验证，接触方式对于脱敏效应的影响不显著。

看阶段，而基线阶段与视频观看阶段无显著差异；非暴力情境下，视频观看阶段的皮电值显著高于基线阶段与游戏后阶段，而游戏前后无显著差异。根据一般攻击模型，被试基线阶段无生理唤起，皮电值应较低；接触游戏刺激后，无论是否为暴力游戏，皮电值都应有所增加。但进入暴力视频观看阶段，皮电值的具体变化则无法预测，唯一能假设的是非暴力游戏组的生理唤起应高于暴力游戏组。本研究发现在暴力情境下，游戏后的皮电值反而下降，而前人有研究发现皮电值随时间无显著变化 (Carnagey, Anderson, & Bushman, 2007)，这些结果均在一定程度上与一般攻击模型相

表3 心率的平均值与标准差

实验处理	基线		后测		观看视频	
	M	SD	M	SD	M	SD
暴力游戏参与组	83.44	8.80	78.86	8.55	77.38	7.37
非暴力游戏参与组	79.72	11.63	80.20	12.10	77.97	14.43
暴力游戏接受组	75.74	12.16	74.86	9.06	73.09	7.84
非暴力游戏接受组	79.84	8.30	77.46	13.15	74.50	9.18

3.3 心率结果

3.3.1 描述性统计 见表3。

3.3.2 方差分析结果 以心率为因变量，进行2(游戏暴力程度：暴力，非暴力)×2(接触方式：主动参与，被动接受)×3(测量时间：基线，后测，观看视频)三因素混合设计方差分析。结果表明，只有测量时间的主效应显著，心率在基线阶段、游戏后与视频阶段呈递减的趋势（$M_{基线}=79.68$, $M_{后测}=77.85$, $M_{观看视频}=75.73$, $F=8.44$, $p<0.05$）。这一变化趋势与以往的研究结果正好相反 (Anderson & Bushman, 2001; Carnagey, Anderson, & Bushman, 2007)，与一般攻击模型也有矛盾之处。

4 讨论

4.1 皮电值随时间的变化趋势

以皮电值作为因变量的三因素混合设计的方差分析表明测量时间与游戏暴力程度有交互作用——暴力情境下，游戏后的皮电值显著低于基线阶段与视频观

违。考虑其中原因，一个可能是一般攻击模型不完全适用于暴力电子游戏；但也可能与生理指标的测量误差有关。本实验所用的生物反馈仪对于环境温度、湿度的变化较为敏感，有可能由此造成测量误差。

4.2 个体变量的调节作用

有关媒体暴力的文献中，经常受到关注的一个主题是个体对于媒体暴力易感性的差异。根据一般攻击模型，个体变量与情境变量交互影响个体的内部状态，包括认知过程、情绪变化与生理唤起。本研究将身体攻击性与游戏使用习惯这两个个体差异变量作为控制变量纳入了研究。结果并未发现这两个个体变量对于暴力游戏的脱敏效应有调节作用。近期其他人的一些研究也得到了类似的结果 (Bartholow, Bushman, & Sestir, 2007; Carnagey, Anderson, & Bushman, 2006)，提示游戏的脱敏效应有可能在一定程度上独立于个体变量。

4.3 游戏的接触方式

本研究加入了一个新的自变量：游戏的接触方式——主动参与游戏或被动观看游戏录像，以此类比

游戏媒体与传统影视媒体，从而尝试比较这两种不同类型媒体暴力脱敏效应的差异。这样操作实际上假定参与性和互动性是这两种媒体接触方式的根本区别，所以实验要使参与性和互动性之外的其他方面，例如游戏内容、画面激动程度、特别是暴力程度等保持等值。本研究采取用主动组画面录像给被动组观看的办法，来保证这种等值性。被试对于游戏的评价结果也支持了此实验操作的有效性。

本研究的结果并不支持假设二，游戏的接触方式的不同并没有导致脱敏效应的差异。这个结果可以有两种解释，一是主、被动接触方式的确不影响脱敏程度。另一可能是，实验处理中，暴力接触时间长度不够。实验处理为 15 分钟，可能这个时间太短，主、被动接触在情感、注意卷入上的不同还未能积累到产生差异效应的程度。

5　结论与建议

5.1　结论

由本实验结果得出以下结论：

（1）暴力电子游戏能够产生短期脱敏效应。在接触游戏 15 分钟后，与非暴力游戏组相比，暴力游戏组观看暴力视频过程中生理唤起的水平更低。

（2）游戏的接触方式 (主动参与游戏和被动接受游戏录像) 对于脱敏效应没有显著影响；但对于内容完全等值的游戏，主动组知觉到更低的沮丧与更高的愉快程度。

5.2　建议

对于本研究所得出的主动参与与被动接受暴力电子游戏，其脱敏效应没有不同的这一结果，应持谨慎态度。后继的研究在实验处理中对于暴力接触时间的把握、样本量的设定以及非暴力游戏的选择方面需要有所改进，再检查两种接触方式的脱敏效应是否有差别。另外，本研究中用作检验脱敏效应的实验材料为一段暴力影片，在以后的研究中可使用包含真实暴力内容的实验材料来更为有效地检验脱敏效应。

参考文献

Anderson, C. A., Berkowitz, L., & Donnerstein, E. (2003). The influence of media violence on youth. *Psychological Science in Public Interest, 4,* 81‑106.

Anderson, C. A., & Bushman, B. J. (2001). Effects of violent video games on aggressive behavior, aggressive cognition, aggressive affect, physiological arousal, and prosocial aehavior: A meta-analytic review of the scientific literature. *Psychological Science, 12,* 353‑359.

Anderson, C. A., & Bushman, B. J. (2002). Media violence and the American public revisited. *American Psychologist, 57,* 448‑450.

Anderson, C. A., & Dill, K. E. (1986). Affect of the game player: Short-term consequences of playing aggressive video games. *Personality and Social Psychology Bulletin, 12,* 390‑402.

Anderson, C. A., & Dill, K. E. (2000). Video games and aggressive thoughts, feelings, and behavior in the laboratory and in life. *Journal of Personality and Social Psychology, 78,* 772‑790.

Anderson, C. A., & Griffiths, M. (2004). Cotemporary issues in adolescent video game playing: Brief overview and introduction to the special issue. *Journal of Adolescence, 27,* 1‑3.

Anderson, C. A. (2004). An update on the effects of playing violent video games. *Journal of Adolescence, 27,* 113‑122.

Anderson, C. A., Carnagey, N. L., & Eubanks, J. (2003). Exposure to violent media: The effects of songs with violent lyrics on aggressive thoughts and feelings. *Journal of Personality and Social Psychology, 84,* 960‑971.

Anderson, C. A., Carnagey, N. L., & Flanagan, M. (2004). Violent video games: Specific effects of violent content

on aggressive thoughts and behavior. *Advances in Experimental Social Psychology, 36,* 199‒246.

Bartholow, B. D., Bushman, B. J., & Sestir, M. A. (2006). Chronic violent video game exposure and desensitization to violence: Behavioral and event‒related brain potential data. *Journal of Experimental Social Psychology, 42,* 532‒539.

Bushman, B. J., & Anderson, C. A. (2002). Violent video games and hostile expectation: A test of the General Aggression Model. *Personality and Social Psychology, 28,* 1679‒1686.

Buss, A. H., & Perry, M. P. (1992). The aggression question. *Journal of Personality and Social Psychology, 63,* 452‒459.

Carnagey, N. L., & Anderson, C. A. (2005). The effects of reward and punishment in violent video games on aggressive affect, cognition, and behavior. *Psychological Science, 16,* 882‒889.

Carnagey, N. L., Anderson, C. A., & Bushman, B. J. (2007). The effect of video game violence on physiological desensitization to real‒life violence. *Journal of Experimental Social Psychology, 43,* 489‒496.

Chen, M. F., & Chen, Y. P. (2005). The effect of aggressive games on individual implicit aggressiveness. *Psychological Science, 28,* 458‒460.

[陈美芬 , 陈舜蓬 . (2005). 攻击性网络游戏对个体内隐攻击性的影响 . *心理科学 , 28,* 458‒460.]

Cline, V. B., & Croft, R. G. (1973). Desensitization of children to television violence. *Journal of Personality and Social Psychology, 27,* 360‒365.

Cui, L. J., Hu, H. L., & Wu, M. Z. (2006). A study on the implicit aggressiveness of gaming addicts. *Psychological Science, 29,* 570‒573.

[崔丽娟 , 胡海龙 , 吴明证 . (2006). 网络游戏成瘾者的内隐攻击性研究 . *心理科学 , 29,* 570‒573.]

Dill, K. E., & Dill, J. C. (1998). Video game violence: A review of the empirical literature. *Aggression and Violent Behavior, 3,* 407‒428.

Funk, J. B., & Buchman, D. D. (2003). Playing violent video games, desensitization, and normal evaluation in children. *Applied Developmental Psychology, 24,* 413‒436.

Funk, J. B., & Baldacci, H. B. (2004). Violence exposure in real‒life, video games, television, movies, and the internet: Is there desensitization? *Journal of Adolescence, 27,* 23‒29.

Gentile, D. A., & Anderson, C. A. Violent video game: The newest media violent hazard. Retrieved December 20, 2008, from http://www.psychology.iastate.edu/~dgentile/106027_07.pdf

Gentile, D. A., & Lynch, P. J. (2004). The effect of violent video game habits on adolescent hostility, aggressive behavior, and school performance. *Journal of Adolescence, 27,* 5‒22.

Griffths, M. (1999). Violent video games and aggression: A review of the literature. *Aggression and Violent Behavior, 4,* 203‒212.

Kirsh, S. J. (2003). The effects of violent video games on adolescents: The overlooked influence of development. *Aggression and Violent Behavior, 8,* 377‒389.

Mullin, C. A., & Daniel, L. (1995). Desensitization and resensitization to violence against women: Effects of exposure to sexually violent films on judgments of domestic of domestic violence victims. *Journal of Personality and Social Psychology, 69,* 449‒459.

Thomas, M. H., & Horton, R. W. (1977). Desensitization to portrayals of real‒life aggression as a function of exposure to television violence. *Journal of Personality and Social Psychology, 35,* 450‒458.

Van Mierlo, J., & Van den Bulck, J. (2004). Benchmarking the cultivation approach to video game effects: A comparison of the correlates of TV viewing and game play. *Journal of Adolescence, 27,* 97‒111.

Xin, Z. Q., & Chi, L. P. (2004). The mechanism of violence

in the virtual on the aggressive behaviors of children. *Journal of the Chinese Society of Education*, (5), 41 – 44.

[辛自强 , 池丽萍 . (2004). 虚拟世界的暴力对儿童攻击行为的影响机制 . *中国教育学刊* , (5), 41 – 44.]

Zhang, Z., Liu, Y. X., & Zhang, J. X. (2006). The theory model of the effect of violent media on the youths aggressiveness. *Chinese Journal School Health, 27*, 915 – 916.

[张镇 , 刘月霞 , 张建新 . (2006). 媒体暴力影响青少年攻击行为的理论模型 . *中国学校卫生 , 27*, 915 – 916.]

Zheng, H. M., & Sun, Y. J. (1997). Violent video games and aggression. *Advances in Psychologicial Science, 14*, 266 – 272.

[郑宏明 , 孙延军 . (1997). 暴力电子游戏对攻击行为及相关变量的影响 . *心理科学进展 , 14,* 266 – 272.]

Short-term desensitizing effects of violent video games: Comparison between two exposure ways

GUO Xiaoli[1, 2] , JIANG Guangrong[1] , ZHU Xu[1]

([1]School of Psychology, Huazhong Normal University, Hubei human beings development and mental health key laboratory, Wuhan, 430079, China)

([2]Student Work Department, Huazhong University of Science and Technology, Wuhan, 430074, China)

Abstract This study examined the desensitizing effect of active or passive exposures to violent video games on male college students' physiological arousal when viewing a violent film. The study employed a 2 (active or passive exposure) x 2 (violent or nonviolent video game) factorial design. Half of the forty-four participants were randomly assigned to either playing a violent video game or watching the records of someone else playing the violent game for 15 minutes, and the other half assigned to playing or watching a nonviolent video game. Then all the participants were presented with a 10-minute long violent film segment while their heart rate (HR) and galvanic skin response (GSR) were being recorded. The result showed that participants who previously played or viewed a violent video game had lower GSR while viewing the violent film than those who previously played or watched a nonviolent video game. This result demonstrated a physiological desensitization effect of exposure to violent video games on physiological arousal toward violence. However, the way of exposure, active or passive, to violent video games failed to show any influence on the degree of desensitization, although actively-playing group reported more enjoyment and less frustration than did passively-viewing group. Results were interpreted and discussed using the General Aggression model.

Keywords violent video games; desensitization; exposure ways

当事人对领悟的看法：质化分析

胡姝婧 [1,2]，江光荣 [2]，鲁艳桦 [2]，张莎莎 [2]，陈锐娟 [2]，于丽霞 [2]，杜　睿 [2]

([1] 海军工程大学理学院，武汉 430033)

([2] 青少年网络心理与行为教育部重点实验室，华中师范大学心理学院，湖北省人的发展与心理健康重点实验室，武汉 430079)

摘　要　为了从当事人的视角理解领悟，采用协商一致的质化研究方法对 15 位当事人的访谈结果进行分析。结果发现 7 个与领悟有关的域：领悟的内容，领悟的效果，影响领悟产生的因素，评估领悟质量的依据，领悟出现时的反应，领悟的来源和阻碍领悟发挥作用的因素。形成了领悟的概念界定：领悟是对自己和他人(主要是自己)的新认识，对自己的认识内容包括，自己的问题模式，心理困扰或问题模式的原因、影响和解决办法，以及自己内在的心理状态。

关键词　心理治疗；领悟；质化研究

分类号　R395

1　前言

心理咨询中有一个变量从咨询诞生时就受到重视，但直到最近才获得了越来越多的研究关注，这就是领悟 (insight)。弗洛伊德十分重视领悟，追求领悟是精神分析的标志性特征。其他咨询取向对领悟重要性的认识在早期有差异，但发展至今，领悟的重要性已经得到了比较一致的认同。在咨询理论越来越趋向于整合的今天，在较有影响力的跨理论咨询模型中，领悟也被提到了相当的高度。如 Hill(2009) 提出的三阶段咨询模型，探索—领悟—行动，领悟就是关键的第二阶段。

除了理论对领悟的重视之外，研究也一再证实了领悟在咨询中的重要性。对咨询会谈中重要事件类型的研究，代表性的是 Elliott 的助益事件 (helpfulevents)

研究和 Mahrer 的好的时刻 (good moments) 研究。Elliott, James, Reimschuessel, Cislo 和 Sack(1985) 通过访谈当事人在会谈结束后的体验，得到当事人所认为的会谈中的助益事件，8 类助益事件领悟是其中之一。Mahrer 和 Nadler (1986) 根据理论和前人的研究，归纳出 11 个会谈中好的时刻，领悟也是其中之一。此后其他各种对重要事件的研究，领悟的重要性都一再被证实：如格式塔治疗中好的时刻 (Boulet, Soulière, Sterner, & Nadler, 1992)，抑郁症治疗中当事人对治疗获益的觉知 (Gershefski,Arnkoff, Glass, & Elkin, 1996)，团体咨询中的重要事件 (Moreno, Fuhriman, & Hileman, 1995)，当事人在咨询中的获益体验 (Paulson, Truscott, & Stuart,1999)，当事人认为的咨询中的重要内容 (Levitt,Butler, & Hill, 2006)，等等，这些研究结果无一例外都包含领悟。

通讯作者：江光荣，E-mail: grjiang@yeah.net

理论和实证研究都表明领悟是咨询中一个十分重要的因素，然而对领悟重要性的认识，却与对领悟的研究不成正比，专门的研究还很缺乏。Gibbons, Crits-Christoph, Barber 和 Schamberger(2007) 在对领悟的相关实证研究进行文献综述后感叹道："鉴于领悟在理论文献中的重要性，过去 40 年中，在这一概念的操作化和与治疗效果关系的检验上所付出的努力如此之少，是很令人惊讶的"。

心理咨询中的领悟指的是什么，这是研究的出发点。以下将总结西方主流咨询取向对领悟的看法，对领悟的跨理论界定，以及国内学者对领悟的理解，来梳理现有的对领悟基本概念的认识。

1.1 主流咨询取向的观点

主流咨询取向对领悟的认识各不相同。在《InsightinPsychotherapy》(Castonguay&Hill,2007) 一书中，主流的心理咨询取向对各自所认为的领悟进行了界定和诠释，概括如下。

精神分析是领悟概念最早出现也最重视领悟的取向。(1) 弗洛伊德的着作中很少直接提及领悟一词，但是其理论构想和实践中贯穿着领悟。在他那里，领悟是对潜意识动机和防御机制的发掘，对痛苦真相的寻找，对创伤性经历与当前心理痛苦关系的了解。(2) 在自我心理学中，领悟有两种涵义：一种是指内观的过程及其发现，另一种是承认自己的问题，而这种承认预示着治疗的成功。领悟既可以是治疗的手段也可以是最终目标，当其作为目标时，是将以前无意识的驱力、愿望、幻想、冲突和其他非理性的斗争整合到现实自我中。这两种理论都把领悟看作心理改变的原因。(3) 在关系理论中，领悟被视为心理改变的结果，改变发生的证据。它是在安全的治疗关系中努力澄清当事人的困扰后所产生的结果。当领悟发生时，治疗双方会察觉自己在咨访互动中表现着当事人生活中的主题。(Messer &McWilliams, 2007) 总体来看，在心理动力学取向中，领悟是无意识的冲突、驱力、愿望、动机等的意识化，它既是治疗的手段也是治疗的目标，既可以是心理改变的原因也可能是改变的结果。

在体验疗法中 (包括当事人中心、格式塔、过程 – 体验和某些存在疗法)，领悟通常被等同为觉察 (awareness) 和元觉察 (meta-awareness)。觉察是指明确地关注当下体验的某一方面，而元觉察是对感知事物、信息加工或建构个人体验的方式的特殊觉察。(1) 当事人中心疗法中，罗杰斯把领悟描述为"当事人达到的一种体验"，以联结和接纳的方式，是一种感觉到的，而不是理智上的体验。在罗杰斯那里，领悟似乎可以和觉察、感觉到的体验、符号化等互换。(2) 强调体感聚焦 (focusing) 的体验治疗，把领悟视为在当下的觉察过程中解释和创造新涵义的产物。(3) 存在疗法中的领悟是存在性的领悟，通常是在面对终极关怀 (死亡、孤独、无意义和自由) 时获得的觉察，是以情感为基础的对生命和生活的看法。(4) 格式塔疗法也认为领悟就是觉察和元觉察，是发现一个人的体验和行为，以及行为的方式。(Pascual-Leone & Greenberg, 2007) 可以看出，体验疗法十分强调领悟中的体验成分，领悟是内在体验和体验方式的意识化，领悟发生时必定伴有体验。

与前两种取向相比，认知 – 行为疗法最初对领悟重要性的强调最弱。但随着实践中领悟不断地偶然出现，以及对治疗产生的促进作用，逐渐引起了理论家的重视；与此同时，实证研究的结论也加速了该流派对领悟的接纳。(1) Ellis 在其理 – 情行为治疗中区分了理性领悟 (intellectual insight) 和情绪性领悟 (emotional insight)，认为后者造成的信念和行为改变的强度要超过前者。(2) 在其他的认知疗法中，领悟的涵义与认知改变、认知重构、理性重构、认知调整、理性再评价、发现非理性等近似。Beck 认为，认知改变过程由对自己想法的觉察、识别不适当的想法和用更适当的想法替换组成，领悟则包含在识别非理性的自动化想法和觉察替代性的认知中。Meichenbaum 认为认知重构是行为改变的关键，它既是治疗改变的手段也是改变的目的。认知重构反映着图式的改变，图式改变与单纯的理性领悟不同，它包含着心理机能的多重维度。(3) 从图式理论的观点来看，领悟是自我和他人图式的改变。在此过程中，个体有意识地觉

察到两个或多个图式的联结，而该联结是之前不存在或以特殊方式联结在一起的。图式作为在长时记忆中储存的心理表征，其表征水平与储存位置在外显或内隐的记忆系统有关，理性领悟是在外显水平上建立图式间新的联结，而情绪性领悟则需整合外显和内隐的表征 (Holtforth et al., 2007)。综上所述，认知－行为疗法重视领悟中的认知成分，同时也强调认知改变时所伴随的情绪体验，它在心理表征的层面上进一步阐明了领悟的实质。

总体来看，心理动力学把领悟看作联结的形成，联结的内容包括过去与现在、内在冲突与外在表现、依恋关系与移情，等等，当两两间内在的联系贯通时，领悟就发生了。体验疗法把领悟等同于觉察和元觉察，当内在的体验意识化时，领悟就出现了。认知－行为疗法认为领悟是认知重构或图式改变，发现以前的非理性信念或图式，用新的理性信念或新图式取代，这就是领悟。而图式改变的含义亦即在图式间建立新的联结。由此看来，心理动力学和认知－行为疗法都认为领悟的实质是建立新联结，但是对联结的内容、联结形成方式等的看法有所不同。

Pascual-Leone 和 Greenberg (2007) 从领悟加工方式的角度抽取了两个维度，来对不同取向中的领悟进行比较，见图 1。这两个维度是抽象程度和加工类型。

抽象是指提取跨情境的具体的稳定因素并内化的过程，跨越时间和空间。抽象程度越高，归纳的范围越宽，抽象的来源越广。加工类型是指情感和认知加工的相对分量，加工一种体验既可以采用对知觉和情绪即时化的方式，也可以通过概念和理性思考的方式。

该模型认为主流取向中的领悟在两个维度上都有各自的位置，相应地表现出一些典型特征：(1) 大多数体验疗法中的领悟主要是觉察，抽象水平最低，聚焦于此时此地，加工方式是知觉－情绪 (如 "我现在感到对父亲很愤怒"); (2) 存在疗法和某些体验疗法中的领悟主要是体验性的元觉察，抽象水平比觉察高，加工方式既有知觉－情绪，也有概念－理性，但以知觉－情绪为主 (如 "我现在有种把整个世界都看作我的对立面的感觉"); (3) 认知疗法中的领悟主要是理性的元觉察，抽象水平比体验－存在性领悟更高，加工方式既有知觉－情绪，也有概念－理性，但以概念－理性为主 (如 "我现在意识到我是害怕失败所以不愿意尝试"); (4) 心理动力学中的领悟主要是概念联结，抽象水平最高，跨越时间和空间，加工方式是概念－理性 (如 "现在我明白，从小就缺乏安全感，所以我一直不敢亲近任何人")。这四类领悟，抽象程度依次递增，加工类型逐渐从知觉－情绪向概念－理性转移，一起共同构成了完整的领悟概念。
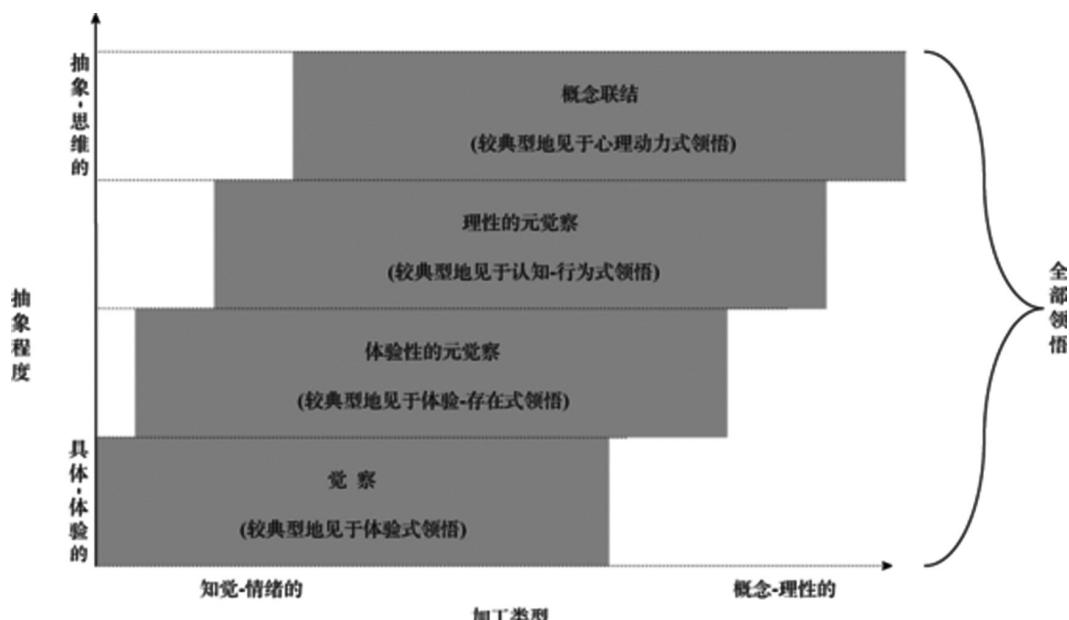
图 1　领悟的二维模型 (Pascual-Leone & Greenberg, 2007)

1.2 跨理论的观点

APA 心理学词典 (VandenBos, 2006) 对领悟的定义是：在心理治疗中，对自己或他人的情绪、认知或行为困难的潜在来源的觉知。Hill 等 (2007) 曾与 30 位不同取向的临床心理学家进行讨论，最后对领悟的概念得出了较为一致的意见：领悟是包含新联结的有意识的意义转变 (即"这个与那个相联系"或某些因果的感觉)。

在跨理论的实证研究中，研究者对领悟给出了不同的界定。Elliott (1985) 在对会谈中"助益事件" (helpful events) 的研究中，对领悟的界定为：当事人描述认识到一些和自己有关的新东西，包括获得认知领悟，看到与自我或人际关系中的自我的一些新的联结。Mahrer 和 Nadler (1986) 在对会谈中"好的时刻"(good moments) 的研究中，对领悟的界定为：当事人表达或陈述一个重要的领悟 – 理解，有 3 个特点：(1) 表达情绪唤起的感受；(2) 在看待 (认识和 / 或建构) 自己和自己的世界的方式上表现出确实的改变；(3) 对自己的生活和个人 / 人际行为具有重要含义。Hill (1992) 在对会谈中的当事人行为进行分类的《当事人行为系统》(Client Behavior System) 中，对领悟的界定是：当事人表达出对自己的理解，可以明确说出行为、想法或感受的模式或原因；领悟通常包含一个"啊哈"的体验，当事人以一种新的方式知觉自己和这个世界；当事人承担适当的责任而不是责怪他人、使用外界强加的"应该"或合理化。Gelso, Kivlighan, Wine, Jones 和 Friedman (1997) 在研究领悟、移情与咨询效果的关系时，对领悟的界定为：当事人对正在探索的材料表达出正确理解的程度。理解可能是针对关系、当事人在咨询室外的机能状况、或者当事人的心理动力和行为方面。Gelso 和 Harbin (2007) 在精神分析理论的基础上，对领悟给出了一个较宽泛的定义，将领悟界定为对以下内容的理解和觉察：(1) 潜在的感受，想法和行为；(2) 它们之间的内在联系以及与早期事件的关系；(3) 内部事件 (想法和感受) 和外部事件 (行动) 的关系。他们强调领悟最基本的特点是对以前无意识的感受、想法和行动的觉察和理解。它并不需要直接和个人遥远的过去有关 (如童年早期)，也不需要反映无意识的加工，尽管往往会这样。

虽然表达各异，但仔细比较可以发现，不同学者对领悟的认识是大同、小异。大同表现为对领悟本质特征的认识是一致的，都认为领悟是当事人对自己有了一些新的认识和理解，表述为"认识到一些和自己有关的新东西"、"在看待 (认识和 / 或建构) 自己和自己的世界的方式上表现出确实的改变"、"表达出对自己的理解"、"表达出正确理解"，"对以前无意识的感受、想法和行动的觉察和理解"，都是在说当事人对自己的新理解。小异主要表现在两个方面：其一，对领悟内容的认识不尽相同，换句话说，对当事人对自己的什么有了新的认识和理解看法不同，有的认为是"自我或人际关系中的自我"、有的是"自己和自己的世界"、有的认为是"行为、想法或感受的模式或原因"、还有的是"关系、在咨询室外的机能状况、或者心理动力和行为"以及"以前无意识的感受、想法和行动"。对这些内容进行总结，可以归纳为两部分，一部分是自己的心理机能，包括认知、情绪、行为倾向和它们之间的相互影响；一部分是自己与他人的关系。其二，对领悟特征的认识不尽相同。有的界定中表述了领悟的特征，有的没有，如"领悟通常包含一个'啊哈'的体验"、"表达情绪唤起的感受"，等等。这些有关特征的表达在不同的界定中没有共性，因此有可能反映的是研究者对领悟的个人看法，不一定是领悟的必要特征。如"表达情绪唤起的感受"，会在情绪性领悟中出现，但不会在理性领悟中出现，但理性领悟仍然是领悟的一种类型。

理性领悟和情绪性领悟的区分是对领悟的一种比较重要的划分，得到了大多数临床心理学家的认同。心理动力学早先并没有对领悟进行分类，直到自我心理学家 James Strachey 在 1934 年提出"突变解释" (mutative interpretation) 的概念，这一概念强调整合了情感和认知的解释，要大大优于只有认知的解释。在此之后，大多数心理动力学家就区分了理性领

悟和情绪性领悟 (引自 Messer &McWilliams, 2007)。 Albert Ellis 最早在理 – 情行为治疗中将领悟区分为理性领悟和情绪性领悟。两种领悟中当事人都能认识到错误的信念、自我挫败的行为，也都会体验到改变信念和行为的愿望。但是，两种领悟在影响程度上存在差别，在影响行为类型的数量、影响效力和承诺等方面，情绪性领悟都要优于理性领悟 (Ellis, 1963)。值得注意的是，心理动力学和理 – 情行为疗法所说的情绪性领悟，尽管使用的是同一个词，但二者的内涵并不一样，心理动力学认为的情绪性领悟是"伴随着理解的宣泄过程" (Gelso et al., 1997)，而理 – 情行为疗法认为的情绪性领悟是"伴随着理性领悟的确信感" (Ellis,2001)。Wachtel (1997) 对理性领悟和情绪性领悟的区分进行了综述，将二者概括为：理性领悟可以看作一个认知过程，帮助当事人掌握内在冲突的因果关系；情绪性领悟包含情感，当事人不但在理智上掌握了某种内部事件，而且体验到与那些事件有关的感受，这些感受是之前难以获得或没有体验过的。对如此区分的临床心理学家来说，情绪性领悟对促进行为改变更有效。Gelso 和 Harbin (2007) 对"情绪性领悟"这一名称提出了意见，认为这一命名并不恰当，因为容易和情绪宣泄相混淆，而且尽管其含义里包含认知成分，但这一命名本身却并没有体现出认知元素。因此，他们建议将其改为"综合性领悟" (integrative insight)，它是认知理解和情绪体验的综合。当事人获得综合性领悟时，可以在认知上掌握自己的冲突和问题的原因，同时能体验到之前没有觉察和联系到这一认知理解的感受。

以上的文献回顾显示，目前获得较多认同的研究结论有：(1) 领悟的实质是建立新联结。尽管主流理论取向对领悟的界定不一致，但基本都认同领悟的实质是建立新联结，这一观点也得到了大多数临床心理学家的认可。(2) 领悟是当事人对自己的新认识和理解。对领悟的跨理论界定基本都持这一观点，新认识和理解的内容包括两部分，一是自己的心理机能，包括认知、情绪、行为倾向和它们之间的相互影响，二是自己与他人的关系。(3) 领悟有两种类型，理性领

悟和情绪性领悟。这一区分得到了大多数临床心理学家的认同。临床观察表明，在促进行为改变上，情绪性领悟比理性领悟更有效。

1.3 国内学者的观点

国内学者对领悟的关注不多，且大多是在个案报告中提及。最有影响力的研究当属钟友彬教授的认识领悟疗法。该疗法是在精神分析的理论基础上，植根于中国本土的一种心理动力疗法，因此，它也秉承了精神分析的理论精髓，强调领悟的重要性。钟友彬认为，领悟的本质是要"暗示病人认识并厌弃那些过时的或幼稚的感情和行为模式而代之以较为成熟的、健康的行为模式" (钟友彬，1985)，而其本土性就表现在病人领悟的内容，是"结合了中国的历史文化背景和病人的生活经验的"(钟友彬，1985)。还有一些学者用中国古诗词来描述他们对心理治疗中出现领悟的理解，如胡岚和周和玲 (2006) 认为领悟的出现会经过三个境界。第一境界是"昨夜西风凋碧树。独上高楼，望尽天涯路"，意指咨询师清楚地看到患者的许多问题，对患者的问题及相关因素尽收眼底，理清来龙去脉。第二境界是"衣带渐宽终不悔，为伊消得人憔悴"，意指咨询师在准备领悟到来的过程中，冥思苦想，呕心沥血，仔细研究患者症状轻重程度，认真分析患者问题的因果关系，通盘考虑与患者问题有关的因素，然后慎重选择有针对性的理论和技巧，用于最好解决的问题。第三境界是"众里寻他千百度，蓦然回首，那人却在，灯火阑珊处"，意指灵感到来，领悟出现。咨询师达到前两个境界，为第三境界的到来做好准备，而第三境界才是心理治疗的真正结果，要达到这一境界，咨询师要善于用心倾听，分析综合，联想推理，奇思妙想，经过漫长的准备，才能引导患者看到自己症状的根源及因果关系，最终引导患者领悟，达到"那人却在，灯火阑珊处"的境界。此外，还有研究者在分析比较西方主流理论取向对领悟看法的基础上，提出了领悟的跨理论界定："当事人在治疗师的辅助下采取各种方式，对自己和自己的世界形成新的觉知，表现为在个人意义系统中建立新的联结" (胡姝婧，江

光荣，2010)。

尽管该领域的研究已经取得了一些成果，但总体来看还存在一些问题：(1) 缺乏跨理论的研究。当今心理治疗的大趋势是走向整合，因此，跨理论治疗中的领悟研究就显得愈发需要。但目前这类研究一方面还比较少，另一方面也缺乏统一的概念界定，相应地也缺乏在概念界定基础上编制的专门的测量工具，在这些基本问题得到解决之后，才能对跨理论治疗中领悟的作用大小、作用机制等问题进行更深入的研究。(2) 缺乏对领悟的系统分类。对领悟进行分类，有助于深化对领悟现象的认识，也有利于考察不同类型领悟在咨询中所起的作用。目前只有理性领悟和情绪性领悟这一种区分得到了比较多的认同。这种区分来自于临床实践，在临床上具有一定的价值，但在研究中往往很难区分，也难以进一步研究 (Gelso & Harbin, 2007)。(3) 缺乏当事人视角的研究。现有研究几乎都是站在咨询师或临床研究者的角度来对领悟进行界定，但这一视角和当事人的是否一致？当咨询师致力于促进当事人获得领悟时 (注意是咨询师认为的领悟)，当事人是否真的从这样的领悟中获益，还是他们认为的领悟与此不同？现有研究无法回答这样的问题。(4) 缺乏本土的研究。从前面的文献回顾可以看出，现有的实证研究都是西方的，中西方当事人在不同的文化背景中，在不同的咨询氛围中，对领悟的认识是否会有所不同，钟友彬的研究从实务层面提供了一些证据支持，但还缺乏实证研究的归纳提炼。

基于以上分析，该研究将以国内的跨理论咨询为背景，对中国当事人眼中的领悟进行质化研究。质化研究的意义在于：首先，通过质化研究对领悟进行概念化，它为之后的测量工具编制和其他量化研究提供了基础。其次，有助于全面了解跨理论咨询中出现的领悟，对当事人在领悟前、中、后的体验有清晰的认识。再次，质化研究可以发现一些重要的问题，以后用量化研究的方法进一步检验和分析。最后，质化研究将以领悟的亲身体验者——当事人为访谈对象，它有助于我们从另一个视角丰富现有的认识。

2 研究方法

2.1 受访者

15 位当事人接受了访谈，其中男性 1 人，女性 14 人，年龄在 19 至 27 岁之间 (M=22.67, SD=2.09)。其中 14 人为某大学的学生，接受咨询时均为在校生，在该校心理咨询中心进行咨询，另有 1 人已工作，在社会心理咨询机构接受咨询。咨询次数最少 8 次，最多 56 次 (M=12.73, Md=9, SD=12.34)。参加研究时 13 人已经结案，2 人尚未结案。咨询的问题基本为发展性问题，包括自我探索与个人成长、人际关系、学习压力、亲密关系、人生规划等。15 位咨询师为这些当事人提供咨询，没有单一取向的咨询师，都是以某一种取向为主，同时兼用其他取向作为辅助，咨询取向主要是以人为中心疗法、认知疗法和心理动力疗法。

2.2 数据收集

在征得该校心理咨询中心负责人的书面同意后，获得了该中心部分接受心理咨询的当事人名单。从中筛选出会谈次数在 8 次及以上的心理咨询个案。通过手机短信告知这些当事人研究的相关信息，邀请其参与研究。将研究的访谈提纲通过电子邮件发送给同意参与研究的当事人，使其了解访谈内容并有所准备。然后和当事人约定好时间进行访谈。访谈形式有面谈、网络访谈和电话访谈，根据现实条件和当事人的意愿进行选择。访谈前向当事人介绍研究目的、访谈录音、文字记录、保密等相关事宜，征得当事人同意并签署知情同意书后开始访谈。访谈一般在 1 个小时左右。访谈结束后向当事人表示感谢，对面谈的当事人赠送了小礼品。通过这种方式获得了在该中心咨询的 14 位当事人的访谈资料。另有一位当事人在社会心理咨询机构接受了咨询，由他人推荐参与研究，通过相同的流程也对其进行了访谈。这样一共获得 15 份访谈材料。将这 15 份材料全部转成逐字稿，略去了可识别当事人和咨询师身份的信息，共计 11 万余字。

2.3 访谈提纲

访谈提纲由三部分组成。第一部分为当事人的

基本信息，包括年龄、参加咨询的时间、咨询次数、是否结案等。第二部分是当事人咨询的主要问题。第三部分是咨询中所获领悟的相关问题，主要包括三个方面，一是对领悟的界定，认为领悟是什么；二是领悟的内容及对其质量评价；三是按照时间顺序，领悟出现前、中、后咨询的相关情况，包括咨访互动、当事人的想法感受、咨询的变化、当事人及其生活的变化等。该访谈为半结构式访谈，访谈问题均为开放式问题，访谈内容基本固定，但访谈过程中可根据当事人的回答适当增加或调整问题，以使答案更清晰。访谈提纲中的问题主要由本文第一作者拟定，在问题基本形成后，请两位咨询方向的博士生进行了审阅，根据其提出的意见进行了修改。

2.4 资料分析

对访谈资料的分析采用协商一致的质化研究方法 (Consensual Qualitative Research, CQR)。该方法由 Hill 等提出 (Hill, Thompson, & Williams, 1997; Hill et al., 2005)，是一种在心理咨询的研究中应用较广泛的质化研究方法。该方法与其他质化研究方法最大的区别在于，访谈资料的分析者不是一个人，而是一个小组；对访谈资料的分析要在组员间达成一致；并且会对结果的代表性进行检查。这样的分析过程有助于克服传统质化研究中存在的一些问题，如单一研究者在分析时可能存在偏见，研究结果难以被重复，研究过程达不到科学研究要求的规范和严谨等 (Hill et al., 1997)。

CQR 要求受访者至少 8 至 15 人，研究小组由 3 至 5 人组成，另外包括 1 至 2 名审核者。研究步骤主要是三步：(1) 将开放式问题的回答划入不同的域 (domain)，即不同的话题范围。(2) 将每个域中的内容提炼核心观点 (core idea)，即写摘要或简短的总结。(3) 交叉分析，将所有个案的同一个域的核心观点放在一起，对其进行归类 (category)，并对类别的代表性予以评定。每一步都必须通过研究小组的讨论，在组员间达成一致。审核者对每一步的结果予以审查，提出修改意见，供研究小组参考。

2.4.1 研究小组

本研究小组共 6 名成员，均为咨询方向的博士生或硕士生，其中博士生 4 人，硕士生 2 人，均为女性。所有组员事先对 CQR 的操作规程进行了学习和讨论，并阅读了采用 CQR 方法所做研究的相关文献。在研究开始前，按照 CQR 的要求，为了避免个人偏见和期待对研究过程的影响，每位组员先按照访谈提纲，根据自己的认识陈述对领悟的看法。主要有：领悟是对感受的澄清，了解自己的感受是什么及感受的来源；领悟是对带入咨询中的问题、在咨询中涉及的问题的新认识；领悟是对解决问题的方法途径的认识；领悟是对自己或自己的人际交往的新认识；领悟是将无意识意识化。组员表达了自己的看法后，被提醒尽量搁置这些看法，忠实数据本身。在此后的数据分析过程中，组员间时常会相互提醒其分析是否受自己看法的影响。

2.4.2 研究流程

按照 CQR 的操作要求，研究主要包括三步，划域，提取核心观点和交叉分析。具体操作如下。

（1）划域。阅读每一份访谈材料，将当事人的回答中和领悟有关的内容，划入相应的域中，并对每个域给予命名。组员先独立地对个案进行划域，然后一起讨论，逐渐确定下每个域的含义、命名以及逐字稿的划分方法。用确定下来的域再去划分之后的个案，并增补前面个案中未出现的域，或根据新出现的内容对域或域名进行修改调整。最终，每个域的含义和命名，以及对每一个个案的划域都在小组成员间达成一致。所有个案的划域结束后，组员再对每个个案独立进行检查，看是否需要修改，然后在小组中讨论，达成一致。

（2）提取核心观点。对每一个个案的每一个域里的内容，组员先独立地提取核心观点，即概括中心思想。核心观点要忠实原意，能用原文表达时最好用原文。然后对每个核心观点的内容和表述进行讨论，在小组中达成一致。在全部个案讨论结束后，组员独立检查每个个案，结合上下文考察核心观点的表述是否准确，并对同一个个案重复的核心观点予以合并。修改的结果经小组讨论，再次达成一致。

（3）交叉分析。以域为分析单位，将不同个案

的同一个域的核心观点全部放在一起。组员先独立地对每一个域的核心观点进行分类，按照某一个分类标准将一个域里的核心观点分成几类。然后在小组中讨论，对分类标准、类别的命名以及核心观点的归类达成统一。先对15个个案中的13个进行分析，保留2个个案用于检查分类的稳定性，包括类别是否适用，类别的代表性是否会发生变化。结果这2个个案并没有改变原有结果，因而认为结果是稳定的。

（4）审核。将研究结果送给4位咨询方向的博士生审核，审核内容包括域的划分和命名是否准确，分类标准及其命名是否恰当，核心观点的归类是否合适等。研究小组对审核的结果进行讨论，确定是否需要修改和如何修改。

（5）类别的代表性评定。根据最终的结果对类别的代表性进行评定。按照Hill等(2005)的标准，如果类别适用于所有的个案或只有一个个案不符合，则该类别是普遍的(general)，在本研究中是15或14个个案；如果适用于一半以上的个案，则该类别是典型的(typical)，在本研究中是8至13个个案；如果适用于至少2个个案，则该类别是变异的(variant)，本研究中是2至7个个案；如果只适用于一个个案，则将该类别放入杂类中，在结果中不予报告。

3　结果

和领悟有关的内容被划分成7个域：(1) 领悟的内容；(2) 领悟的效果；(3) 影响领悟产生的因素；(4) 评估领悟质量的依据；(5) 领悟出现时的反应；(6) 领悟的来源；(7) 阻碍领悟发挥作用的因素。表1呈现了所有的域，每个域里的类别(包括母类和子类)，以及类别的代表性评定。

3.1 领悟的内容

当事人在咨询中获得的领悟主要是围绕自己的心理困扰或问题模式，认识到自身存在的问题模式、心理困扰或问题模式的原因、造成的影响和解决办法。此外，还有少数领悟是对他人的理解或对自己内在心理状态的觉察。

问题模式是当事人惯常采用的会造成不良影响的应对方式，它常常被运用却往往没有自觉。在本研究中，80%的当事人对自己的问题模式有所领悟。当事人领悟的问题模式可以归纳为四种：（1）达到外界标准，限制真我表达。当事人认识到自己为了达到外界的要求，得到外界的肯定和接纳，而隐藏了真实的自我，把自己变成外界希望的样子。如当事人领悟到，"自己很多时候表现的不是真我，其实自己是特立独行和清高的，但做出来的样子是大家喜欢的、合群的"，"我一直通过达到外在的标准来获得价值感，因为无法自己肯定自己的价值，所以要通过控制别人，和有性格缺陷的异性交往来获得价值感"。（2）被动回避倾向。当事人认识到自己在生活中存在被动回避面对问题的倾向，如"我还像一个没有长大的小孩，生活中总是被动应对，没有主动操控，遇到困难会逃避面对，减少和外界的接触"，"我以前过于被动，总是等待机会或回避困难"。（3）不当的防御方式。当个体的需求没有得到满足而出现负面情绪，或潜意识中预感到负面情绪将要出现时，不直接面对和处理，而采用一些防御方式。如"在情感上不接受不好的经历，总会合理地解释它，从而不能客观地看待问题"，"我通过别人(姑姑、男友)来发泄自己的负面情绪。"（4）不良的人际交往模式。除了上述三种对自己问题模式的认识，还有一种对人际交往中自己问题模式的认识，它将焦点从个体自身转移到个体与他人的互动中。如"我对朋友太依赖，我让朋友很窒息"，"虽然我很想和别人建立好的关系，但是没有发自内心地想要去信任或者尊重别人"。

近3/4的个案对自己的心理困扰或问题模式的形成原因有所领悟，可以概括为三种：（1）完美主义倾向。因为追求完美导致对自己和他人的不满意、不接纳，如"一系列问题表现(如不自信、对自己不满意等)的原因是太追求完美，对自己要求太高"，"我的道德标准很高，我认为应该达到的道德要求，别人达不到时我会很困扰。这其实是要求别人和我一样。"（2）未获满足的积极关注的需要。被关注、肯定和爱的需要没有得到充分的满足，为了获得他人的积极

表 1　领悟的 CQR 研究结果

域	类别（母类 / 子类）	代表性评定
1. 领悟的内容	1.1 问题模式	12T
	1.1.1 达到外界标准，限制真我表达	4V
	1.1.2 被动回避倾向	4V
	1.1.3 不当的防御方式	3V
	1.1.4 不良的人际交往模式	5V
	1.2 心理困扰或问题模式的原因	11T
	1.2.1 完美主义倾向	4V
	1.2.2 未获满足的积极关注的需要	8T
	1.2.3 人际互动中的不良倾向	4V
	1.3 心理困扰或问题模式的影响	5V
	1.3.1 对自己的影响	2V
	1.3.2 对人际交往的影响	3V
	1.4 心理困扰或问题模式的解决办法	10T
	1.4.1 降低完美主义倾向	3V
	1.4.2 提高对负性事件的接纳度	2V
	1.4.3 增强主动性	3V
	1.4.4 改变人际交往的观念	4V
	1.4.5 调整人际界限	4V
	1.5 对他人的理解	2V
	1.6 觉察	3V
2. 领悟的效果	2.1 自己的变化	15G
	2.1.1 态度	12T
	2.1.1.1 接纳自己	6V
	2.1.1.2 成为自己	4V
	2.1.1.3 趋于成熟	2V
	2.1.1.4 接纳他人	5V
	2.1.2 情绪感受	8T
	2.1.2.1 平静轻松	5V
	2.1.2.2 开朗积极	3V
	2.1.2.3 存在感、清晰感、掌控感	5V
	2.1.3 行为	6V
	2.1.3.1 人际行为的变化	5V
	2.1.3.2 个体行为的变化	3V
	2.1.4 自我认识与调节	11T
	2.1.4.1 自我理解和觉察	8T
	2.1.4.2 自我调整	10T
	2.1.5 整体变化	5V
	2.2 人际关系的变化	5V
	2.3 咨询的变化	8T
	2.3.1 更信任喜欢咨询师	4V
	2.3.2 咨询更清晰流畅	4V
	2.3.3 考虑结束咨询	3V
	2.3.4 无变化	5V
	2.4 其他（改变的状态）	4V
3. 影响领悟产生的因素	3.1 咨询师方面的促进因素	15G
	3.1.1 倾听	4V
	3.1.2 反馈	11T
	3.1.3 解释	5V
	3.1.4 即时化	2V
	3.1.5 接纳、非指导的态度	5V
	3.2 当事人方面的促进因素	10T

续表 1

域	类别（母类／子类）	代表性评定
	3.2.1 内省思考	7V
	3.2.2 信任咨询师	2V
	3.2.3 阅读书籍	3V
	3.2.4 咨询中尝试行动	2V
	3.2.5 咨询中的状态	3V
	3.3 阻碍因素	2V
4. 评估领悟质量的依据	4.1 领悟对当事人及其生活的影响力	13T
	4.2 领悟的特性（重要性、深刻性、完善性）	5V
	4.3 领悟时的感受强度	3V
5. 领悟出现时的反应	5.1 豁然开朗	7V
	5.2 希望感	5V
	5.3 惊喜震撼意外	5V
	5.4 释放放松	4V
	5.5 感动	2V
	5.6 与个人问题相关的感受	2V
6. 领悟的来源	6.1 咨询师	11T
	6.2 当事人	7V
	6.3 咨访双方	5V
7. 阻碍领悟发挥作用的因素	7.1 旧有习惯的惯性	4V
	7.2 现实因素的制约	4V
	7.3 改变动力的不足	2V

注：代表性评定中，G 代表 general（普遍），T 代表 typical（典型），V 代表 variant（变异）；数字代表个案数；如 "7V" 表示该类别在 7 个

关注，而形成问题模式或产生心理困扰，如 "为了在重男轻女的家庭环境里得到更多关注，让自己向男性性别角色认同"，"小时候受冷落的自己希望通过更完美、更乖、学习更好来引起别人的注意和爱，但又达不到，所以自卑。"（3）人际互动个案中出现，属于变异的。中的不良倾向。人际互动中不良的心理或行为倾向，导致人际交往中出现心理困扰或产生问题模式，如 "只关注自己的紧张情绪，没把精力投入到对方身上，是导致我人际交往困难的原因"，"和父亲关系不好的一个原因是我总是用教训的口吻和他说话"。

1/3 的个案对于心理困扰或问题模式对自己造成的影响有所领悟。有两种影响：一种是对自己的心态和生活造成的影响，如 "自卑让我在做很多事情之前就把结果想得很糟糕，所以不敢做"，"超越界限使我做事收效不好"；另一种是对人际交往造成的影响，如 "表现得像小孩子能逗大家喜欢，但不一定能获得尊重"，"被动使我不能被别人理解"。2/3 的个案对如何解除自己的心理困扰或问题模式有所领

悟。可概括为 5 种方式：(1) 降低完美主义倾向。调整自己对于完美的认识，接纳自己的不完美，如 "世界不完美，我也不可能做到完美"，"人本身就是不完美的，正是这种不完美，才让人不断努力不断追求，所以要接纳自己的不完美"。(2) 提高对负性事件的接纳度。在认知上引导自己接纳消极负面的事件，如 "要客观地看待父亲对我的培养方式，不能为了不破坏父亲的形象，就找借口来掩饰"，"只有跟过去和解了才能更好地活在当下"。(3) 增强主动性。调整心态，积极主动地面对生活，如 "我要用积极的而非退缩的方式处理问题"，"生活中，困难和问题会时时存在，主动解决会带来好心情、自信心"。(4) 改变人际交往的观念。对人际交往的方式增加了一些新的认识，或者改变了原有不正确的观念，如 "人际交往中表达和理解同样重要，学会表达能获得主动权。适时适度的表达会得到真正的理解"，"我值得父母关心，不一定要很优秀才能得到父母的爱，对他们对我的关爱不用内疚。父母关心子女是其价值的一种体现"。(5) 调整人际界限。从认识上或行动意向上调整人际距离，

保持适当的人际界限，如"我和弟弟是两个独立的个体，不应要求弟弟走和自己一样的路"，"自己应该独立了，不应再依赖别人的支持"。

绝大部分当事人的领悟属于以上四种类型，是对自身存在的问题的新认识。但也有个别当事人对他人的心理状态或行为方式有所领悟，获得了新的理解，本研究中有两个个案谈到这种领悟，如一个当事人认识到"即将高考的弟弟压力很大，除了自身的压力和父母的期望，还有超越姐姐和父亲的压力，所以用表面看起来漫不经心、对高考不屑的方式来应对压力"，还有一个当事人谈到"站在爷爷的角度去考虑问题，我了解了他那样做的原因"。另外有 3 位当事人获得的领悟可以概括为觉察，是对自己内在心理状态的体认，如"其实我对父亲是有不满的"，"有些情绪和事情其实是对自己有影响的"，"我的力量增强了，敢于面对情敌了"。

3.2 领悟的效果

在获得领悟后，当事人自身、其人际关系和咨询都会发生一些变化，将这些变化统称为领悟的效果。

所有的当事人自身都有变化，可归纳为五个方面：(1) 态度的变化。80% 的个案在对待自己和他人的态度上有所变化。对于自己态度的变化，一是要接纳自己，如"要接受自己，扬长避短"，"比较能接受自己的不足了"；二是要成为自己，做自己而不再被外界所塑造，如"不想再为别人而活，想要坚持自己的路"，"要让本真的自我力量增强，不受外界的打压"；三是要让自己的为人处世更成熟，如"觉得自己处事应更成熟、大度，不能还像小孩一样任性"，"不再采用旧的行为模式，要展现自己成熟的一面"。对于他人态度的变化主要是接纳他人，对他人更理解、包容和关爱，如"对他人更包容"，"觉得父亲挺不容易的，感激父亲，想跟父亲走得更近"。(2) 情绪感受的变化。60% 的个案在情绪感受上有所变化。主要表现在三个方面：一是感到平静轻松，以前的消极情绪消除或缓解，如"烦躁降低"，"内心平静了很多"；二是更加开朗积极，心情更豁然，对待生活更积极，如"豁然很多"，"认识到生活有希望，困难可以解决，

不再过于畏难了"；三是体验到存在感、清晰感或掌控感，在生活中体验到真我的存在，对自己的认识更清晰，更能掌控自己的问题，如"日常生活中体验到真实的自己的存在"，"思想理顺了，通透了"，"比以前更有能力处理情绪问题，不再困在情绪问题里不知所措"。(3) 行为的变化。近一半的当事人在行为上有变化，有两种变化：一种是在人际交往方面，主动与外界接触沟通，如"敢和父母说出内心真实的感受了"，"跟外界的接触更多"；另一种是当事人自身行为的变化，如对他人的抱怨减少，或为了解决自己的问题进行一些有针对性的行为练习，如"上课主动发言锻炼自己，主动与同学结组参加每周活动，以提高自信心"，"在生活中尝试拒绝"。(4) 自我认识与调节。近 90% 的个案在自我认识与调节方面出现变化。在自我认识上，对自我的理解和觉察能力增强，对与自身问题相关的认知、情绪和行为的理解和觉察力提高，如"对自己认识更清晰，知道想法产生的原因和如何应对"，"能从生活细节中觉察自己退缩的模式"；在自我调节上，面对自己的问题和会引发自己问题的情境时，能够从心态上进行调整，更好地应对问题，如"不再理睬以前让自己很纠结的小事情，平静地做事，试着依靠自己而非掌控环境来获得安全感"，"在看到他人缺点时，会回想起领悟的内容，从而更倾向于包容他人"。(5) 整体变化。1/3 的个案谈到了自己整体性的变化，如"自卑降低"，"更有力量生活，生活更积极主动，心情好，更能承担责任"。

1/3 的个案的人际关系出现一些变化，大部分是积极的变化，如"朋友关系更和谐，朋友更多，自己很开心"，"和同学、男友的关系改善"；但也有个别个案出现消极的转变，如"和父母的关系有些纠结"，或者是没有变化，如"想过能否根据领悟改善人际关系，但觉得难就放弃了，所以没有变化"。

80% 的个案在领悟后咨询会出现一些变化，可概括为四个方面：(1) 更信任喜欢咨询师。在领悟后对咨询师更信任喜欢，对咨询师更开放，暴露更多更重要的个人信息，如"信任咨询师，告诉她从未说过的对自己影响很大的事情"，"更信任咨询师，暴露更

重要的事"。(2) 咨询更清晰流畅。领悟提供了线索，使接下去的探索更清晰流畅，如"咨询有了重点和线索，自己的思路更清晰"，"在以后的咨询中涉及人际方面时，会谈到领悟的内容"。(3) 考虑结束咨询。在领悟后感到问题基本解决，咨询可别个案出现消极的转变，如"和父母的关系有些纠结"，或者是没有变化，如"想过能否根据领悟改善人际关系，但觉得难就放弃了，所以没有变化"。80% 的个案在领悟后咨询会出现一些变化，可概括为四个方面：(1) 更信任喜欢咨询师。在领悟后对咨询师更信任喜欢，对咨询师更开放，暴露更多更重要的个人信息，如"信任咨询师，告诉她从未说过的对自己影响很大的事情"，"更信任咨询师，暴露更重要的事"。(2) 咨询更清晰流畅。领悟提供了线索，使接下去的探索更清晰流畅，如"咨询有了重点和线索，自己的思路更清晰"，"在以后的咨询中涉及人际方面时，会谈到领悟的内容"。(3) 考虑结束咨询。在领悟后感到问题基本解决，咨询可以结束，如"知道如何面对问题，愿意自己应对，咨询中发现的新问题少了，可以考虑结束咨询"，"清楚了自己的问题，不再迷茫，没有要咨询的困惑了"。(4) 咨询没有变化。也有个案表示领悟后咨询没有变化。

有 4 位当事人谈到了在领悟后自己改变的状态，表达了较为一致的信息：领悟后的改变不是很快很容易发生的，而是慢慢的一点点的改变，需要时间和努力，如"让自己更清晰，但距离实际的生活改变还有很长的距离"，"明确了改变的大方向，但具体到生活中的小事不是很容易改变，会尝试"。

3.3 影响领悟产生的因素

影响领悟产生的因素包括促进因素和阻碍因素两部分。促进领悟产生的因素可以分为两个方面，一是咨询师方面，一是当事人方面。所有的当事人都谈到了咨询师方面的促进因素，包括五种：(1) 倾听。近 1/3 的当事人表示咨询师的倾听促进了他们的自我探索，并获得领悟。(2) 反馈。近 3/4 的当事人获得领悟是经由咨询师的反馈，咨询师陈述他们对于当事人问题的看法。(3) 解释。1/3 的当事人通过咨询师给予

解释获得领悟。(4) 即时化，有 2 位当事人通过咨询师在咨询过程中做即时化的反应获得领悟。(5) 接纳、非指导的态度。1/3 的当事人感到咨询师对自己的接纳促进了其对自己的接纳，在这样放松、没有限制的环境中可以充分地自我探索，并获得领悟。1/3 的当事人感到咨询师对自己的接纳促进了其对自己的接纳，在这样放松、没有限制的环境中可以充分地自我探索，并获得领悟。2/3 的当事人谈到了自己方面的促进因素，包括五种；(1) 内省思考。近一半的当事人谈到自己的内省思考对领悟形成的作用，他们会通过各种方式，包括对咨询师反馈的思考、和身边的人讨论以及自己的反省，来获得领悟。(2) 信任咨询师。1 位当事人谈到对咨询师的信任会促进自己充分的自我表露，从而获得领悟；另 1 位当事人谈到对咨询师的信任使自己愿意接受咨询师的分析和反馈，从而获得领悟。(3) 阅读书籍。3 位当事人在咨询过程中看书受到了启发获得领悟。(4) 咨询中尝试行动。2 位当事人在咨询过程中，于日常生活中尝试了一些与以前不同的行为，进而获得了领悟。(5) 咨询中的状态。3 位当事人谈到自己在咨询中无拘无束地表达的这样一种状态，以及迫切地想摆脱困扰的状态，促进了领悟的出现。还有 2 位当事人谈到了阻碍领悟产生的因素，主要是在咨询关系方面，咨询双方互不接纳，或者联结不够紧密，当事人不能够完全开放。

3.4 评估领悟质量的依据

当事人对领悟质量的评估主要是从三个方面：(1) 领悟的影响力。近 90% 的当事人认为领悟对自己和生活的影响力的大小，是评估领悟质量的最重要的依据。如果领悟能够对自己和生活产生积极的影响，当事人对领悟质量的评分较高，反之则较低。这是从领悟效果的角度对领悟的评估。(2) 领悟的特性。1/3 的个案从领悟的重要性、深刻性、完善性的角度对领悟质量进行评估，领悟内容对自己的重要程度，接近核心问题的程度，以及完善全面的程度，是他们评估领悟质量时的依据。(3) 领悟时的感受强度。3 个个案从领悟时自己的感受强度来评估领悟质量，感受包括两种，一种是和问题相关的个人化的情绪感受，另

一种是领悟出现时感到的惊讶 (震撼)。当情绪感受强烈的时候，当事人倾向于对领悟质量评高分。

3.5 领悟出现时的反应

在领悟出现时，当事人会有一些感受和反应，可概括为 6 种：(1) 豁然开朗。近一半的当事人谈到这一感受，如"打开一条通路的感觉，拨云见日，豁然开朗"，"眼前一亮，找到出口的感觉"。(2) 希望感。领悟后 1/3 的当事人的希望感增强，感到生活有了希望，有了力量和信心去应对问题。如"突然有了信心，觉得一切都不可怕了"，"感到有力量，有信心应对问题"。(3) 惊喜震撼意外。1/3 的当事人对领悟感到惊喜、震撼或意外，如"对自己表达的想法和表达时思路的清晰感到震惊"，"感到意外"。(4) 释放放松。4 位当事人谈到自己在领悟后哭了，尽情地宣泄，感到放松，如"尽情地哭和宣泄，感觉很放松"，"一下哭出来，压抑的情绪顿时释放"。(5) 感动。2 位当事人谈到对于咨询师对自己的关注和帮助感到感动，如"对咨询师的帮助很感动"，"为咨询师对我的关注感动"。(6) 与个人问题相关的感受。2 位当事人谈到领悟后对自己的问题产生了一些个人化的感受，如"觉得委屈，心疼以前的自己"，"羞愧"。

3.6 领悟的来源

研究还考察了领悟的来源：来源于咨询师的有近 3/4 的个案，咨询师对当事人反馈后，当事人一下子感到说得很对、很准，从而有所领悟；来源于当事人自身的有近一半的个案，他们可能是在咨询探索的过程中自己逐渐领悟，也可能是在生活中通过内省、看书、行动获得领悟；来源于咨访双方的有 1/3 的个案，当事人在咨询师的引导下逐渐获得领悟，或者是在咨询师反馈后当事人在会谈中或会谈后继续思考，确认咨询师反馈的正确性后再获得领悟。

3.7 阻碍领悟发挥作用的因素

本研究中，有一些个案谈到在咨询中虽然有领悟，但是领悟并没有对自己造成很大的影响，自己还是没有多大的改变，或是改变起来感到很困难。总体来说，有三个因素阻碍了领悟发挥作用：(1) 旧有习惯的惯性。4 个个案认为旧有的习惯或模式已经养成，

形成了一种惯性，尽管它会对自己造成不良的影响，但一方面习惯的影响力很大，改变起来很困难，另一方面维持惯性很舒服，而如果要改变，会感到不舒服。(2) 现实因素的制约。4 个个案认为现实因素或外部环境限制了自己的改变，如面临现实生活中的压力，没有精力考虑如何改变；旧有的人际关系已经形成，难以突破等。(3) 改变动力的不足。2 个个案谈到改变动力不足，如生活忙碌了就不把问题放在心上；以后会离开现在的人际环境，所以不用为了改善现在的人际环境而去改变，等等。

4 讨论

4.1 领悟的界定

从最一般的意义上说，心理咨询是一个解决问题的过程，当事人带着自己的问题而来，在咨询师的协助下解决。在咨询中要解决的问题既可以是当事人自诉的心理困扰，也可以是潜伏在心理困扰之下，悄悄发挥作用的更根本的问题。当对这些问题进行工作时，当事人会获得一些新的认识，这些新认识包括自身存在的问题模式，造成自己的心理困扰或者问题模式的原因，心理困扰或者问题模式对自己造成的影响，以及消除心理困扰或问题模式的办法，这四个方面共同构成了当事人在咨询中获得的领悟的主要类型。除此之外，当事人也可能会对他人有所领悟，更能理解他人，或者是对自己内心的状态有所体察。因此，若从这一研究结果来界定咨询中的领悟，可以将其定义为：领悟是对自己和他人 (主要是自己) 的新认识，对自己的认识内容包括，自己的问题模式，心理困扰或问题模式的原因、影响和解决办法，以及自己内在的心理状态。

将这一定义和前人从咨询师角度的跨理论定义进行比较，可以发现一些异同。共同点在于，都认为领悟是当事人对自己的新认识，这是对领悟本质特征的最基本认识。而区别主要有两点：其一，对领悟内容的认识层面不同。前人的界定中对内容的描述有的是在比较抽象宏观的层面，如"自我或人际关系中

的自我"(Elliott, 1985),"自己和自己的世界"(Mahrer & Nadler, 1986),这样的表述无疑是正确的,但因为过于宏观,所以提供的信息量不大。有的是在比较微观的层面从心理机能的角度来描述领悟内容,如"行为、想法或感受的模式或原因"(Hill, 1992),"关系、在咨询室外的机能状况、或者心理动力和行为"(Gelso et al., 1997),"以前无意识的感受、想法和行动"(Gelso & Harbin, 2007)。而本研究可以说是在二者之间从中观的层面对领悟内容的描述,围绕心理困扰或问题模式来建构领悟。这与当事人的身份及其对领悟的直观感知是一致的,他们报告的是自己知觉到的领悟内容,因而是与其咨询问题密切相关的,自己的问题是什么,是如何产生的,对自己造成了什么影响,以及如何解决,这就是他们所领悟的主要内容。而咨询师受过心理学专业训练,因此在看待领悟的内容时难免刻上心理学的烙印。双方视角不同,对领悟内容的解读也就存在差异。这一结果提示咨询师在咨询中可以多一个视角,尝试从当事人的角度来看待和促进领悟。其二,对领悟内容的认识不尽相同。尽管认识层面不同,但当事人和咨询师认为的领悟内容绝大部分还是一致的,这主要是指问题模式、原因、影响和觉察。除此之外,当事人认为的领悟内容还包括心理困扰或问题模式的解决办法,前人的界定中没有这一部分。这说明当事人和临床心理学家在这一点上有分歧。按照经典精神分析的观点,领悟后症状会自然消失,但有些持跨理论观点的学者对此提出了异议,认为领悟后仍然需要行动的跟进,才能消除症状 (Hill, 2009)。本研究表明,对有些当事人来说,获得领悟并不一定能使症状消除,他们需要在理解自己问题的基础上,继续探索并对如何解决问题有所领悟,然后依此行动或做出调整,才能有所改善。此外,当事人的领悟还包含对他人的新认识,尽管绝大多数领悟是对自己的新认识,但当对他人有新的认识和理解时,当事人也认为是领悟。而前人的界定中,除了 APA 心理学词典 (VandenBos, 2006) 的定义中涉及了他人,其他都只包括对自己的认识。

4.2 领悟的来源、出现时的反应和作用

领悟既可以来自于咨询师的反馈,也可以来自于当事人自身的思考,还可以来自于咨访双方的合作,在本研究中,分别是 11 个、7 个和 5 个个案。从这个结果来看,咨询师是最主要的领悟来源,当事人完全通过自己获得的领悟相对少一些,但差别不是很大。完全通过一种来源获得领悟的个案有 8 个,通过两种或三种来源获得领悟的个案有 7 个。采用哪种方式使当事人获得领悟,在几种传统的主流咨询取向中表现出比较大的差异,精神分析和认知疗法中主要的领悟来源是咨询师,体验疗法中主要的来源是当事人自己。这也说明领悟可以有不同的来源。在跨理论取向的治疗中,领悟究竟来自咨询师还是当事人也许并不太重要,既然是领悟,那说明它们最终都是被消化吸收而成为了当事人自己的东西,尽管殊途却是同归,"白猫黑猫,抓住老鼠就是好猫"的理论也许在这里同样适用。但从另一方面来说,不同来源的领悟在质量上是否存在差异,对咨询过程和效果的影响是否有所不同,这些问题还有待研究检验。

在领悟出现时,当事人相应地会有一些感受和反应,西方的研究中将主要的反应称为"Aha",类似于本研究中的豁然开朗,都是指一种突然通了、亮了、清楚了、明白了的感受。除了这一种常与领悟伴随出现的感受外,本研究表明,还会有一些其他的感受或反应。如果从理性领悟和情绪性领悟的角度来看这几种反应,可以发现,豁然开朗、希望感、惊喜震撼意外、感动是两种领悟都可能导致的反应;而释放放松、与个人问题相关的感受却只和情绪性领悟有关。由此看来,可以将这些反应分为两类,一类是由领悟导致的积极情绪反应,一类是由领悟引起的具有个人意义的情绪反应,前者可以由理性领悟或情绪性领悟引发,而后者的出现才表明获得的是情绪性领悟。这样,可以通过了解当事人在领悟出现时的反应来判断获得的是理性领悟还是情绪性领悟。当然,这倚赖于当事人的自我报告,如果当事人在咨询中没有明确的言语说明或非言语表示 (如哭泣),咨询师和观察员很难凭观察来判断其是否获得了情绪性领悟。这或许

可以为理性领悟和情绪性领悟的测量提供一点启示。

领悟与咨询效果的关系一直是研究者关心的一个问题，领悟对咨询效果起多大作用，是如何起作用的，对这些问题的回答有助于深化对咨询机制的理解。本研究虽然不能从量上回答这些问题，但研究结果反映出领悟的确会导致当事人发生改变，而这些改变中有一部分是咨询效果本身，如态度、情绪感受和行为的变化，还有一部分是导致咨询产生效果的中间变量，如加深了自我认识，增强了自我调节的能力。此外，结果显示领悟会促进工作同盟质量的提升，如当事人更信任喜欢咨询师——增强了情感联结，咨询更清晰流畅——增强了目标和任务的一致性。这一结果支持 Wampold, Imel,Bhati 和 Johnson-Jennings (2007) 提出的理论模型，他们认为领悟会促进工作同盟的进一步巩固，而工作同盟与咨询效果的关系已得到大量研究的确证，因此领悟会通过促进工作同盟来促进咨询效果的实现。以后的研究可以采用定量的方式进一步检验这些假设，并考察领悟的影响力大小。

虽然大多数当事人认为领悟对自己有积极的影响，但也有个案谈到虽然有领悟，但改变并不容易。这说明，领悟并不一定能直接改善当事人的症状或改变其生活，仍然需要在对自己的问题有所认识后，付出努力才能产生切实的改变。如果因为惯性、现实因素、缺乏动力就不付出努力和行动，也就不会有很大的改变。这与 Hill (2009) 领悟并不一定能直接导致行动的看法是一致的，也正因为如此，她将行动放在领悟之后，作为咨询的最后一个阶段，而不是在领悟之后就结束咨询。

4.3 启示

对于临床工作，该研究提示：(1) 咨询师应该重视促进当事人在咨询过程中获得领悟，因为领悟既可以给当事人带来积极的改变，也有助于巩固工作同盟，使咨访双方更好地合作，以实现咨询目标。但也应注意，并非领悟后咨询就可以结束，而应在领悟的基础上进一步陪同当事人制定改变计划，并监督其付诸实施，否则有可能当事人在回归现实环境后因为较大的困难而不改变。(2) 咨询师可以转换视角，从当事人的角度来看待和促进领悟，按照问题解决的思路，思考当事人心理困扰的前因后果，进行概念化，然后在咨询中和当事人讨论，进行验证，使当事人获得领悟。(3) 促进领悟形成的因素很多，咨询师既可以适时采用合适的咨询技术，也可以促动当事人调用自身的因素。同时应注意不良的咨询关系会阻碍领悟形成，因此在咨询关系欠佳时，咨询师应首先致力于修复咨询关系，而不应急于促成当事人的领悟，否则有可能会适得其反。

对于后续的研究工作，该研究提供了概念化支持，可以在此基础上编制领悟测量工具。还可以根据本研究的一些发现，采用定量的方式进一步研究，如理性领悟与情绪性领悟的区分，及其与后续咨询过程和效果的关系；不同的促进领悟形成的因素对领悟的影响力大小；领悟与咨询效果的关系，以及工作同盟、自我调节等在其间所起的作用；等等。

从文化角度审视该研究的结果，笔者认为，总体来说并未显示出十分强烈的文化特异性。尤其是在领悟的内容方面，大多数内容与精神分析或人本主义理论的基本观点是一致的，如对防御方式、价值条件的领悟等。导致这一状况的原因，既可能是因为这些理论观点的确是具有普适性的，也有可能是由于该研究中的咨询师，都受过这些理论取向的训练，因此在咨询中解释或引导时都会受到理论的影响，从而当事人的领悟也都有明显的理论特色。正如钟友彬 (1985) 指出的："病人的领悟内容和医生的理论观点有密切关系，显然医生的解释起了重要作用"。Wampold 等 (2007) 也认为，领悟的真实性对于当事人接受解释或解释发挥作用并不必要，不存在一种解释比另一种解释更接近真实，解释的真实性也与效果无关，重要的是一个解释是否被当事人接受，而不在于解释是否精准。因此，领悟的内容是本土还是舶来也许并不是那么重要，关键在于当事人是否能够接受。咨询师在帮助当事人获得领悟时，既无需和理论牵强附会，也不必刻意强调本土性而抵制西方理论，而应该贴近实际，灵活处置。

该结果中有一处是否体现出文化特异性尚值得

探讨：影响领悟产生的因素中，来自咨询师方面的促进因素，其中占比最大的是咨询师的反馈，15 个个案中有 11 个获得领悟是由于有咨询师的反馈。这说明，很多当事人会通过咨询师陈述对其问题的看法而获得领悟，尽管也有当事人从咨询师的倾听和接纳、非指导的态度中获得领悟，但当事人也并不排斥咨询师直接表达自己的观点，只要恰当，他们也愿意接受。许多论述心理咨询本土化问题的文章都提到，中国当事人相较西方当事人，更愿意接受咨询师的权威角色（如吴垠，2011; 孔德生,2007)。咨询师在反馈时，的确是扮演着权威角色的，而当事人获得了领悟，说明他们认同了咨询师的反馈，也就是接受了咨询师的权威。如果中国当事人的确更乐意咨询师扮演权威，而咨询师直接反馈也的确能使当事人获得领悟，且这样的领悟与当事人自己探索获得的领悟在效果上并无差异，那么这种方式可能是更适合于中国当事人的方式。但这诸多的前提条件是否能满足，还需要更多的研究检验。

4.4 局限与展望

本研究呈现了跨理论心理咨询中当事人对领悟的基本认识，但研究存在一些局限，以后可以改进。首先是样本问题，访谈对象中绝大多数为女性，接受咨询时都是大学生，均为非临床样本，大多在一个咨询中心接受咨询，咨询的都是发展性问题，除了一个个案有 56 次会谈之外，其余均为短程个案，这虽然符合 Hill 提出的样本同质性的要求，但同样因此而减少了代表性。在与此不同的样本中，结果是否同样适用，还需要研究检验。其次，研究小组的成员均为女性，尽管在结果审查时有男性参与，但仍然不确定完全同性的成员组成是否会对结果造成实质性的影响。最后，该研究对领悟内容的分类，可进一步通过量化研究，如聚类分析或因素分析来检验其可靠性。

参考文献

Boulet, D. B., Soulière, M. D., Sterner, I., & Nadler, W. P.(1992). Development of a category system of goodmoments in Gestalt therapy. *Psychotherapy, 29*, 554 – 563.

Castonguay, L. G., & Hill, C. E. (Eds.) (2007). *Insight inpsychotherapy*. Washington, DC: American PsychologicalAssociation.

Elliott, R. (1985). Helpful and nonhelpful events in briefcounseling interviews: An empirical taxonomy. *Journal ofCounseling Psychology, 32*, 307 – 322.

Elliott, R., James, E., Reimschuessel, C., Cislo, D., & Sack, N.(1985). Significant events and the analysis of immediatetherapeutic impacts. *Psychotherapy, 22*, 620 – 630.

Ellis, A. (1963). Toward a more precise definition of "emotional" and "intellectual" insight. *PsychologicalReports, 13*, 125 – 126.

Ellis, A. (2001). "Intellectual" and "emotional" insightr evisited. *NYS Psychologist, 13*, 2 – 6.

Gelso, C. J., & Harbin, J. (2007). Insight, action, and thetherapeutic relationship. In L. G. Castonguay & C. E. Hill(Eds.), *Insight in psychotherapy* (pp. 293 – 311). Washington, DC: American Psychology Association.

Gelso, C. J., Kivlighan, D. M., Wine, B., Jones, A., &Friedman, S. C. (1997). Transference, insight, and thecourse of time-limited therapy. *Journal of CounselingPsychology, 44*, 209 – 217.

Gershefski, J. J., Arnkoff, D. B., Glass, C. R., & Elkin, I.(1996). Clients' perceptions of treatment for depression: I.helpful aspects. *Psychotherapy Research, 6*, 233 – 247.

Gibbons, M. B. C., Crits-Christoph, P., Barber, J. P., &Schamberger, M. (2007). Insight in psychotherapy: Areview of empirical literature. In L. G. Castonguay & C. E.Hill (Eds.), *Insight in psychotherapy* (p. 159). Washington,DC: American Psychology Association.

Hill, C. E. (1992). An overview of four measures developed totest the Hill Process Model: Therapist intentions, therapistresponse modes, client reactions, and client

behaviors.*Journal of Counseling & Development, 70,* 728–739.

Hill, C. E. (2009). Helping skills: *Facilitating exploration,insight, and action* (3rd ed.). Washington, DC: AmericanPsychology Association.

Hill, C. E., Castonguay, L. G., Angus, L., Arnkoff, D. B.,Barber, J. P., Bohart, A. C., ⋯ Wampold, B. E. (2007).Insight in psychotherapy: Definitions, processes,consequences, and research directions. In L. G. Castonguay& C. E. Hill (Eds.), *Insight in psychotherapy* (p. 442).Washington, DC: American Psychology Association.

Hill, C. E., Knox, S., Thompson, B. J., Williams, E. N., Hess,S. A., & Ladany, N. (2005). Consensual qualitativeresearch: An update. *Journal of Counseling Psychology, 52,*196–205.

Hill, C. E., Thompson, B. J., & Williams, E. N. (1997). Aguide to conduct Consensual Qualitative Research. *TheCounseling Psychologist, 25,* 517–572.

Holtforth, M.G., Castonguay, L. G., Boswell, J. F., Wilson, L.A., Kakouros, A. A., & Borkovec, T. D. (2007). Insight incognitive–behavioral therapy. In L. G. Castonguay, & C. E.Hill (Eds.), *Insight in psychotherapy* (pp. 57–80).Washington, DC: American Psychology Association. Hu, L., & Zhou,

H. L. (2006). The comprehension inpsychological consultation and treatment. *Medicine &Philosophy (Humanistic & Social Medicine Edition), 27,*41–43.

[胡岚 , 周和玲 . (2006). 心理咨询与治疗中的领悟 . *医学与哲学 (人文社会医学版)*, 27, 41–43.]

Hu, S. J., & Jiang, G. R. (2010). Definition of "insight" inpsychotherapy. *Advances in Psychological Science, 18,*1489–1495.

[胡姝婧 , 江光荣 . (2010). 心理咨询中的"领悟"概念辨析 . *心理科学进展 , 18*, 1489–1495.]

Kong, D. S. (2007). Exploration and practice for localizationof psychological counseling in China.

Academic Exchange,(12), 31–35.

[孔德生 . (2007). 我国心理咨询本土化的探索与实践 . *学术交流 , (12)*, 31–35.]

Levitt, H., Butler, M., & Hill, T. (2006). What clients findhelpful in psychotherapy: developing principles forfacilitating moment–to–moment change. *Journal of Counseling Psychology, 53,* 314–324.

Mahrer, A. R., & Nadler, W. P. (1986). Good moments inpsychotherapy: A preliminary review, a list, and somepromising research avenues. *Journal of Consulting andClinical Psychology, 52,* 10–15.

Messer, S. B., & McWilliams, N. (2007). Insight inpsychodynamic therapy: Theory and assessment. In L. G.Castonguay & C. E. Hill (Eds.), *Insight in psychotherapy* (pp.9–29). Washington, DC: American Psychology Association.

Moreno, J. K., Fuhriman, A., & Hileman, E. (1995).Significant events in a psychodynamic psychotherapygroup for eating disorders. *Group, 19,* 56–62.

Pascual–Leone, A., & Greenberg, L. S. (2007). Insight andawareness in experiential therapy. In L. G. Castonguay, & C.E. Hill (Eds.), *Insight in psychotherapy* (pp. 31–56).Washington, DC: American Psychology Association.

Paulson, B. L., Truscott, D., & Stuart, J. (1999). Clients'perceptions of helpful experiences in counseling. *Journalof Counseling Psychology, 46,* 317–324.

VandenBos, G. R. (2006). *APA Dictionary of Psychology* (pp.484–485). Washington, DC: American PsychologicalAssociation.

Wachtel, P. L. (1997). *Psychoanalysis, behavior therapy, andthe relational world.* Washington, DC: AmericanPsychological Association.

Wampold, B. E., Imel, Z., Bhati, K. S., & Johnson–Jennings,M. (2007). Insight as a common factor.

In L. G. Castonguay,& C. E. Hill (Eds.), *Insight in psychotherapy* (pp. 119–139).Washington, DC: American Psychology Association.

Wu, Y. (2011). The tension between the Chinese "relationship–oriented" social psychology and thelocalization of psychological counseling. *Social SciencesJournal of Universities in Shanxi, 23*, 91–93.

[吴垠 . (2011). 中国人"关系取向"的社会心理特征与心理咨询本土化之间的张力 . *山西高等学校社会科学学报* , 23,91–93.]

[Zhong, Y. B. (1985). "Insight" in psychodynamic therapy. *Journal of International Psychiatry, (3)*, 137–140.

[钟友彬 . (1985). 论动力学心理疗法中的"领悟". 国外医学 . *精神病学分册* , *(3)*, 137–140.]

Chinese psychotherapy clients' perspectives on insight: A qualitative examination

HU Shujing[1,2], JIANG Guangrong[2], LU Yanhua[2], ZHANG Shasha[2], CHEN Ruijuan[2], YU Lixia[2], DU Rui[2]

([1] College of Science, Naval University of Engineering, Wuhan 430033, China)

([2] Key Laboratory of Adolescent Cyberpsychology and Behavior (Ministry of Education); School of Psychology, Central China Normal University; Key Laboratory of Human Development and Mental Health of Hubei Province, Wuhan 430079, China)

Abstract　Insight is a crucial phenomenon in counseling with its importance been confirmed by counseling theory andempirical studies. However, there is a lack of specific studies on this topic, and most researchers focused on psychodynamic area whereas little attention has been laid on pantheoretical counseling. There is no unified definition of insight available at the moment. Psychodynamic therapy regards insight as the formation of connections, e.g. between past and present, inner conflict and external performance, attachment relationships and transference. Experiential therapy equates insight to awareness and meta−awareness. Cognitive‒behavioral therapy views insight as a cognitive restructuring or change of schema; insight occurs when previous irrational beliefs or schemas are recognized and replaced by new, rational beliefs or schemas. Meanwhile, different tresearchers also gave different operational definitions in cross−theory empirical research. As a result, the lack of a unified definition of insight, in either theoretical or empirical research, has made it impossible to integrate different results and prevented the development of a specialized and reliable measurement of insight. In addition, nearly all definitions were proposed by researchers or clinicians. Whether or not clients have a different tperspective is unknown. Moreover, all the definitions currently in use were developed by western researchers focusing on western psychotherapy. Would Chinese psychotherapy present a different perspective? This study sought to delineate insight from Chinese clients' perspective via a qualitative approach.

Fourteen clients counseled in a university counseling center and a client counseled in a social counseling institution participated in the study. They were interviewed with a semi−structured protocol, either face−to−face, by phone, or through internet, about their opinions and experiences of insight during counseling. Their responses were analyzed by a research team using Consensual Qualitative Research (CQR) method.

Analysis revealed seven key issues: contents of insight, effects of insight, factors influencing the appearance of insight, basis for evaluating insight quality, feelings when insight appears, source of insight, and factors that hinder insight from having an effect. Based on interviewees' responses, a definition of "insight" from the client perspective is proposed: "insight is a new understanding of oneself and others (mainly oneself), the contents of which include problematic patterns, their reasons, their effects, solutions for psychological distress or problematic patterns, and awareness of one's internal mental state."

Results suggested that most contents of insight from client and counselor's perspectives are in concordance. However,

the description level is different and some contents which haven's t been thought as insight before was confirmed by clients. It reminds counselors that they could change their angle of view in order to understand insight from client's s point. The current study also provides the foundation of future study's s development of scal efor insight.

Keywords psychotherapy; insight; qualitative research

班级环境与学生适应性的多层线性模型

江光荣[1]，林孟平[2]

（[1] 华中师范大学心理学院，武汉，430079）

（[2] 香港中文大学教育心理学系，香港）

摘　要　此研究探索中国中小学体制下的班级社会心理环境对于学生的适应性的关系。以江光荣和林孟平所编制的《我的班级》问卷测量班级环境，选择学生的学校适应（由 Teacher－Child Rating Scale(T－CRS) 测量）、主观幸福感（以 Student's Life S atisfact ion Scale(SLSS) 测量）和焦虑（用 St ate－Trait Anxiet y Invent ory for Children(STAIC) 测量）作为适应指标，以多层线性模型 (H LM) 方法进行分析，结果显示：学生个体所知觉到的班级环境，对其适应水平有相当肯定的解释力，而一个班级学生整体适应水平的高低，与这个班的班级环境有极大关联。此结果表明，中国学校体制下的班级社会心理环境对于学生的发展和适应状况，具有举足轻重的作用。

关键词　班级环境；学生适应；多层线性模型

1　引言

学校教育环境与学生发展之间的关系一直是发展心理学和学校心理学领域里的研究主题之一。当前的理论比较强调生态学的或系统论的观点，这种观点的核心，是把发展视为一个"个人—环境"相互作用的过程。个人和其周遭的环境共同构成一个生态系统。不同环境成份对于个人的影响力有着远—近、直接—间接的层次区别。研究者认为，班级/课堂环境是与家庭环境处在同一层次，较为直接的环境。近年来 Eccles 等学者在前人观点的基础上，提出所谓"阶段－环境适配"的理论模型 (S tage－Environment Fit)。其核心观点是，学校教育环境应该跟不同年龄段孩子的发展性需要相配合，对学生的发展需要有敏感的反应，这样才能促进学生发展。

传统上对学校环境与学生发展的研究受重智主义的影响，片面集中于探讨学校环境与学生学科成绩和学科态度之间的关系，很少关注学校环境对学生人格和社会性发展的影响。而探讨家庭环境对儿童情绪、人格和社会性方面发展的影响的学者，一般更关注出生到学前这一时段的儿童。这样，有关学校环境与学生人格和社会性发展关系的研究就成了一个较为薄弱的研究领域。这一研究偏向近年受到一些学者的批评，情况有所改变。例如当前不少研究已经转向关注学校对学生情绪和社会性发展的影响。诸如学生的自我概念、情绪适应性、人际关系以及主观幸福感等，越来越多地出现于近年的文献中。

西方学者对课堂环境的结构及其测量方式已经有较充分的研究。在过去 30 余年里，Moos, Walberg 和 Fraser 等人发展了一系列用于测量课堂环境的工具。江光荣对中国大陆初中和小学高年段的班级环境结构进行了研究。结果表明，这个学段的班级环境至少包括 5 个有意义的维度：师生关系、同学关系、秩序和纪律、竞争，以及学习负担。江光荣还发展出了一个测量班级环境的工具，名为《我的班级》。

本研究的目的，是探讨中国学校体制下的班级环境与儿童和青春期少年的学校适应（侧重于情绪和社会性适应）之间的关系，以初步了解班级环境对学生发展的意义。

2 方法和程序

2.1 分析单位和样本

本研究有两种分析单位。一种分析单位是学生个体，另一分析单位是班级。二种分析单位之间呈两层（水平）嵌套关系，即学生嵌套在班级水平之下。

样本来自两所省会城市的城区学校，以分层取样方式抽取。样本包括 25 所中小学，3 个年级，105 个教学班。其中小学 5 年级 7 个班，小学 6 年级 46 个班，初中 2 年级 52 个班。从每个班随机抽取 20 名学生。全样本在个体分析水平共有 2064 名学生，在班级分析水平有 105 个班。根据其中一所城市教育局提供的统计资料计算样本与总体的比例，结果是：学生样本数约占总体的 1%，班级样本数约为总体的 2 %。小学段学生平均年龄 11.5 岁 (SD=0.70)，初中段学生平均 13 .68 岁 (SD=0 .60)。

2.2 工具和测量

班级环境的测量采用江光荣编制的《我的班级》问卷。该问卷由 5 个分量表构成，分别测量班级环境的 5 个维度：师生关系、同学关系、秩序和纪律、竞争和学习负担。

学生的适应性选择了三组变量，包括学校适应、主观幸福感和焦虑。其中学校适应采用 Hig htwer 等人所编制的 Teacher –Child Rating Scale(T –CRS) 为

测量工具，主观幸福感采用 Student' s Life Satisfaction Scale (SLSS) 为测量工具，而焦虑测量则采用 Spielberger 的 State–Trait Anxiety Inventory for Children (STAIC)。

Teacher–Child Rating Scale(T–CRS): Teacher –Child Rating Scale(T–CRS) 由 Hightower 等人在前人量表的基础上精炼修订而成。包括两个部分，一部分测量学生适应上的问题（或障碍），一部分测量学生的适应能力。适应问题部分包括三个分测验：

行动表达 (Acting –Out)——— 过度活动，寻衅，捣乱等行为；

羞怯 – 焦虑——— 退缩，羞怯，胆小，焦虑等行为；

学习问题——— 不良学习习惯，缺乏动机，成绩问题。

适应能力部分也包括三个分测验：

挫折耐受力——— 承受挫折和失败的能力，情绪平衡性等品质；

自表技能——— 自信、合适的社交行为；

任务取向——— 独立、主动、有条理有计划地完成学习任务。

T –CRS 是一个教师评定量表，每个分量表仅含 6 个条目，简单易用。量表由本研究作者译为中文。由本研究的样本得到该中文版的测量学指标为：各分量表 alpha 系数在 0.86 –0.93 之间；因素分析结果表明 6 个分量表分别测量一个独立的因子，适应问题部分三个因子共解释总方差的 67 %，适应能力部分三个因子解释了总方差的 72 %。各项测量学指标跟美国版非常接近。

Student' s Life Satisfaction Scale(SLSS): 这是一个测量总体生活满意度的工具，是一个包含 7 个条目的单一维度测验。量表亦由研究者译成中文，由本研究样本得到的信度系数 (alpha 系数) 为 0.73 。

State–Trait Anxiety Inventory for Children (STAIC): 这是 Spielberger 基于成人的状态 – 特质焦虑问卷 (STAI) 而发展的适用于儿童的工具。问卷包括两部分，一部分测量状态焦虑，一部分测量特质焦虑。本研究

只测量其中特质焦虑部分，即该问卷的 Form C −2，包括 20 个条目。本研究所使用的 Form C–2 系作者根据英文版翻译。以本研究的样本为依据，计算出此部分的内部一致性信度 (alpha 系数) 为 0.92 。因素分析表明全部项目可由一个单一因子解释，可被解释的总方差为 40.1 ％。

以上三个工具系直接使用国外版本。其中 T−CRS 此前已有国内学者使用经验，表明其可用于中国学生的评定 ;S LSS 是一个较流行的青少年生活满意感测量工具，条目所测内容符合中国学生实际 ;STAIC 是世界上通用的儿童焦虑评定量表，文化敏感性小。以本研究的样本得到的三个工具的测量学指标跟国外版的指标均非常接近。

2.3　分析方法

由于数据的嵌套性质，传统的回归分析存在问题。当我们用学生个人为分析单位，忽略学生的班级身份，会导致两个问题。一是这时我们不再是用"班级均值"为班级环境的测量，而是用"学生个人知觉"为班级环境的测量，这样得到的班级环境与学生变量的关系，与用班级均值为分析单位所得到的关系在逻辑上性质是不同的 ;二是这样的分析忽视了"班级身份"对班级环境与学生变量间关系的影响，实际是把班级之间在学生变量均值上的差异歪曲成了班级环境对于学生变量的效应。而真实的情况可能是这种班级均值的差异系由另外的原因所造成。从回归分析的逻辑来说，这会造成对估计标准误的低估并增加犯第一类错误的机会。而此时如果只用班级为分析单位，分析以"班级均值"为单位的班级环境与同样以班级均值为单位的学生变量的关系，则虽然没有逻辑上的问题，但要付出损失学生在班级内个别差异的信息的代价。

近年发展出来的"多层线性模型"分析方法，为这个问题提供了很好的解决途径。这种分析方法不仅较好地处理了上述数据嵌套的问题，而且它允许在不同的分析水平引入不同的预测变量，从而可以处理更复杂的模型。近年在国外已经有课堂环境的研究运用了这种分析方法。本研究采用多层线性模型

处理班级环境跟学生变量的关系，分析工具采用 SSI 公司的软件 HLM4.04 。

3　结果

3.1　班级环境对学生学校适应的影响

根据多层线性模型的分析原理，分别以 T−CRS 的 6 个分测验为因变量，班级环境 5 个维度为自变量 (自变量和因变量均使用标准分数)，构成以下二层级 HLM 方程：

第一层级 (学生个体水平) 方程：

$$Y_{(1-6)ij}=\beta_{0j}+\beta_{1j}(X1_{ij})+\beta_{2j}(X2_{ij})+\beta_{3j}(X3_{ij})+\beta_{4j}(X4_{ij})+\beta_{5j}(X5_{ij})+r_{ij} \tag{1}$$

第二层级 (班级水平) 方程：

$$\beta_{0j}=\gamma_{00}+\gamma_{01}(x1_j)+\gamma_{02}(x2_j)+\gamma_{03}(x3_j)+\gamma_{04}(x4_j)+\gamma_{05}(x5_j)+\mu_{0j} \tag{2}$$

$$\beta_{1j}=\gamma_{10}+\mu_{1j} \tag{3}$$

$$\beta_{2j}=\gamma_{20}+\mu_{2j} \tag{4}$$

$$\beta_{3j}=\gamma_{30}+\mu_{3j} \tag{5}$$

$$\beta_{4j}=\gamma_{40}+\mu_{4j} \tag{6}$$

$$\beta_{5j}=\gamma_{50}+\mu_{5j} \tag{7}$$

$Y_{(1-6)}$ 为 T−CRS 的 6 个分测验得分，下标" j "为班级标号，" i "为学生标号，" ij "代表第 j 个班级第 i 名学生，β_0 为 Y 的班级均值，$X1-X5$ 代表个体的对班级环境 5 个维度的知觉，$\beta_1-\beta_5$ 为对应于 $X1-X5$ 的 5 个班级平均斜率，代表班级水平预测变量对于因变量的效应，r 是残差项。方程 (1) 表示，第 j 个班级第 i 名学生在 T−CRS 六个分测验之一的得分，可以分解为三部分，一是该项得分的班级均值 (β_{0j})，二是可由班级环境各维度的班级平均效应量来解释的部分，三是残差。

方程（1）的 6 个统计量 β_0，$\beta_1-\beta_5$ 分别由方程（2）–（7）来估计。γ_{00}，$\gamma_{10}-\gamma_{50}$ 分别为 β_0，$\beta_1-\beta_2$ 在班级总体中的期望值。在方程 (2) 里，x1j – x5j 代表第 j 个班级的 5 个班级环境维度平均知觉，$\gamma_{01}-\gamma_{05}$ 为对应于 x1 – x5 的班级总体斜率，$\mu_0-\mu_5$ 为班级水平的误差项。

表 1 班级环境与学生学校适应：二层级线性模型

Dependent Variables		Fixed Effects				Random Effects	
		Coefficient	SE	T		V	X^2
行动表达	γ_{10}	−0.105	0.033	−3.219**	μ_1	0.019	117.37
	γ_{20}	−0.104	0.031	−3.343**	μ_2	0.012	75.21
	γ_{30}	0.111	0.033	3.281**	μ_3	0.025	108.83
	γ_{40}	0.056	0.026	2.09*	μ_4	0.007	104.42
	γ_{50}	0.113	0.032	3.476**	μ_5	0.035	115.49
	γ_{01}	−0.067	0.076	−0.884			
	γ_{02}	0.040	0.090	0.444			
	γ_{03}	−0.101	0.066	−1.529			
	γ_{04}	−0.054	0.063	−0.862			
	γ_{05}	−0.011	0.057	−0.184			
羞怯－焦虑	γ_{10}	−0.008	0.032	−0.276	μ_1	0.021	120.78
	γ_{20}	−0.086	0.030	−2.855**	μ_2	0.012	93.58
	γ_{30}	0.037	0.029	1.305	μ_3	0.002	83.33
	γ_{40}	−0.039	0.025	−1.549	μ_4	0.005	101.06
	γ_{50}	−0.007	0.028	−0.283	μ_5	0.015	120.85
	γ_{01}	−0.063	0.094	−0.673			
	γ_{02}	0.010	0.111	0.093			
	γ_{03}	0.119	0.083	1.453			
	γ_{04}	−0.162	0.078	−2.076*			
	γ_{05}	0.063	0.070	0.903			
学习问题	γ_{10}	−0.134	0.032	−4.085**	μ_1	0.019	122.06
	γ_{20}	−0.128	0.031	−4.016**	μ_2	0.016	99.87
	γ_{30}	0.196	0.035	5.587**	μ_3	0.034	124.98
	γ_{40}	−0.004	0.027	−0.131	μ_4	0.012	115.87
	γ_{50}	0.071	0.029	2.472*	μ_5	0.015	125.48
	γ_{01}	−0.042	0.068	−0.626			
	γ_{02}	−0.061	0.080	−0.759			
	γ_{03}	0.084	0.059	1.418			
	γ_{04}	−0.205	0.056	−3.608**			
	γ_{05}	0.113	0.051	2.201*			
自信的社交技能	γ_{10}	0.078	0.033	2.318*	μ_1	0.025	139.186**
	γ_{20}	0.101	0.034	2.951**	μ_2	0.034	133.76*
	γ_{30}	−0.119	0.033	−3.596**	μ_3	0.022	117.18
	γ_{40}	0.040	0.029	1.378	μ_4	0.021	137.75*
	γ_{50}	−0.024	0.031	−0.783	μ_5	0.028	142.81**
	γ_{01}	−0.038	0.081	−0.473			
	γ_{02}	0.105	0.095	1.100			
	γ_{03}	−0.072	0.070	−1.025			
	γ_{04}	0.189	0.067	2.803**			
	γ_{05}	−0.058	0.060	−0.963			
任务取向	γ_{10}	0.146	0.034	4.236**	μ_1	0.032	138.51*
	γ_{20}	0.117	0.034	3.397**	μ_2	0.034	124.75
	γ_{30}	−0.141	0.034	−4.091**	μ_3	0.031	107.13
	γ_{40}	−0.001	0.028	−0.061	μ_4	0.020	137.41*
	γ_{50}	−0.041	0.031	−1.298	μ_5	0.032	157.90**
	γ_{01}	0.016	0.078	0.205			
	γ_{02}	0.054	0.092	0.59			
	γ_{03}	−0.035	0.068	−0.522			
	γ_{04}	0.196	0.065	3.005**			
	γ_{05}	−0.048	0.058	−0.820			
挫折耐受力	γ_{10}	0.086	0.034	2.492*	μ_1	0.042	166.66**
	γ_{20}	0.104	0.030	3.373**	μ_2	0.021	134.15*
	γ_{30}	−0.068	0.031	−2.137*	μ_3	0.025	118.70

续表 1

Dependent Variables	Fixed Effects			Random Effects		
	Coefficient	*SE*	*T*		*V*	X^2
γ_{40}	−0.019	0.028	−0.678	μ_4	0.027	163.49**
γ_{50}	−0.023	0.031	−0.744	μ_5	0.038	181.45**
γ_{01}	0.013	0.095	0.143			
γ_{02}	−0.011	0.112	−0.095			
γ_{03}	−0.041	0.083	−0.491			
γ_{04}	0.246	0.079	3.117**			
γ_{05}	−0.104	0.071	−1.467			

（1）**p<.01 ;*p<.05 ;（2）γ_{10}– γ_{50} : 个人知觉为预测指标时的效果量；μ_1–μ_5 : 个人知觉为预测指标时 , 效果量 (回归系数) 的班级差异及其检验；γ_{01}– γ_{05} : 班级均值为预测指标对班级均值的因变量的效果量 (回归系数);（3）估计参数所对应的自变量为: γ_{10}, γ_{01}, μ_1—师生关系；γ_{20}, γ_{02}, μ_2—同学关系；γ_{30}, γ_{03}, μ_3—秩序纪律；γ_{40}, γ_{04}, μ_4—竞争；γ_{50}, γ_{05}, μ_5—学习负担。

统计结果见表 1。

由表 1 可看出 , 当我们以学生对班级环境的个别化知觉来预测个人的学校适应 (个体分析水平), 在适应问题的三个方面 , 以学习问题跟班级环境 (除了其中的竞争维度) 的关联较大。羞怯 – 焦虑跟班级环境的关联最小。行动表达问题跟 5 个班级环境维度全部都有关系 , 与师生关系和同学关系的关联为负 , 表示学生知觉到的师生关系和同学关系越积极 , 行动表达的行为越少。与秩序纪律、学习负担以及竞争则出现正的关系 , 表示班级环境中这三方面越强 , 学生外显性的行为问题越多。羞怯 – 焦虑与同学关系有负的关联。学习问题跟师生关系和同学关系呈负的关联 , 显示这两种关系较好的学生学习问题较少。而秩序纪律因素和学习负担因素则跟学习问题同方向消长。

在适应能力的三个维度 , 可看出只有师生关系、同学关系及秩序纪律三个环境维度与之有关联。师生关系和同学关系对三个适应能力维度均有正的关系 , 表示这两种关系越好 , 适应能力越好。而秩序纪律维度与适应能力的关系为负 , 可能表明过分强调纪律的班级对学生的适应能力有一定抑制作用。

再看以平均知觉的班级环境与班级平均的学校适应的关联 (班级分析水平), 总的倾向是 , 单个班级环境维度与班级水平的学校适应关联较低。不过其中有一个比较明显且特别的现象 : 班级环境中竞争维度跟学校适应 6 个方面中的 5 个有肯定联系 , 而且这种联系的性质是 , 较强的竞争与较低的羞怯 – 焦虑、较少的学习问题、较好的社交表现、较明显的任务取向、以及较强的挫折耐受能力相关联 , 低的竞争则与上述关联方向相反。按照一般对学校适应的界定 , 此处竞争与学校适应的联系是 : 竞争程度强 , 学校适应好 ; 竞争程度低 , 学校适应差。这个结果有些费解 , 将在后面讨论。除此而外 , 平均知觉的班级环境中仅

表 2　班级环境与学生生活满意感 : 二层级线性模型

	Fixed Effects			Random Effects		
	Coefficient	*SE*	*T*		*V*	X^2
γ_{10}	0.102	0.036	2.822**	μ_1	0.036	137.70*
γ_{20}	0.197	0.031	6.177**	μ_2	0.012	109.66
γ_{30}	0.097	0.032	2.988**	μ_3	0.014	101.33
γ_{40}	−0.030	0.026	−1.156	μ_4	0.002	98.32
γ_{50}	−0.218	0.029	−7.487**	μ_5	0.012	107.88
γ_{01}	0.076	0.037	2.051*			
γ_{02}	0.138	0.044	3.125**			
γ_{03}	−0.063	0.032	−1.945*			
γ_{04}	0.041	0.031	1.321			
γ_{05}	−0.116	0.028	−4.120**			

表注同表 1 。

学习负担与学校适应中的学习问题有关联。学习负担较大者，学习问题较多。

3.2 班级环境对学生生活满意感的影响

以生活满意感分数 (标准分数) 为因变量，以班级环境的 5 个维度 (标准分数) 为自变量，构成如前以学校适应为因变量一样的二层级 HLM 模型。因为此方程组仅第一层级 (学生个体水平) 的因变量改为生活满意感，其余各项均与前一组方程完全相同，故从略。分析结果见表 2 。

由表 2 可看出，除了竞争一项对于学生生活满意感的效应未达到统计显著性水平而外，其余 4 个班级环境维度对于学生生活满意感的作用都比较肯定。仍然是个体分析水平上预测变量的效果量较大，班级分析水平上预测变量的效果量较小。在两种水平上，各维度对于因变量的相对效应比较一致。例如，两种水平上都是同学关系和学习负担对于生活满意感的作用较大，竞争的效应最小。另外，师生关系的作用也很明显。

3.3 班级环境对学生之一般性焦虑的影响

重复上一问题的处理程序，以标准化后的特质焦虑分数为因变量。以同样的二层级线性模型来估计班级环境对焦虑的效应。结果见表 3。

总的来看，班级环境跟学生的一般性焦虑反应有相当肯定的关联。但是班级环境中的不同维度对于焦虑的关系并不一致。无论是在个体分析水平还是在班级分析水平，班级环境中跟焦虑关系密切的都是同学关系、竞争和学习负担这三个维度。其中又以同学关系和学习负担的效应量较大。其作用方向与一般认识一致：同学关系越积极，焦虑反应越低；而学习负担越重，焦虑水平越高。在个体知觉水平，秩序和纪律一项也有较明显的效应。但在班级均值为预测指标时，秩序和纪律对于焦虑不起什么作用。另外，当以个体知觉为预测指标时，各项效果量 (斜率) 在班级之间不存在差别，说明这种关联跨班级的一致性相当高。

3.4 班级环境对不同学生变量的解释力比较

对于一个估计模型，HLM 可以计算出整个模型

表 3 班级环境与焦虑：二层级线性模型

	Fixed Effects				Random Effects	
	Coefficient	SE	T		V	X^2
γ_{10}	−0.039	0.033	−1.185	μ_1	0.024	117.14
γ_{20}	−0.223	0.030	−7.420**	μ_2	0.010	97.76
γ_{30}	−0.189	0.032	−5.865**	μ_3	0.020	101.39
γ_{40}	0.070	0.029	2.45*	μ_4	0.021	116.83
γ_{50}	0.240	0.026	9.247**	μ_5	0.002	99.52
γ_{01}	−0.045	0.039	−1.132			
γ_{02}	−0.195	0.047	−4.142**			
γ_{03}	−0.013	0.034	−0.369			
γ_{04}	0.069	0.033	2.081*			
γ_{05}	0.161	0.030	5.378**			

表注同表 1 。

表 4 班级环境所解释学生变量的比例

因变量	被解释的方差比例	
	个体水平	班级水平
学校适应		
行动表达	0.10	0.07
羞怯－焦虑	0.04	0.02
学习问题	0.11	0.25
自信的社交能力	0.10	0.14
任务取向	0.13	0.14
挫折耐受力	0.12	0.09
生活满意感	0.20	0.82
特质焦虑	0.25	0.79

对于因变量的解释力的估计 1，即因变量的方差中，可以从模型中的全部预测变量得到解释的部分所占的比例。对于第一层级和第二层级方程中均有预测变量的模型——本研究的模型就是这种模型——HLM可以分别给出在两个水平上预测变量的解释力。根据 Bry k 和 Raudenbush 的估计公式，计算出班级环境对三组学生变量之变异的解释比例。结果见表 4。

表中结果表明，无论是以个体知觉来预测个体的学校适应水平，还是以班级均值来预测班级学生的平均适应水平，作为一个整体的班级环境都是一个相当肯定的预测源。结果还显示，对于三组学生变量，班级环境的解释能力有很大差别。对于生活满意感和焦虑，班级环境的解释力比较强，当以个体对班级环境的知觉来预测个体的生活满意感和特质焦虑的时候，两变量分别有 20% 和 25 % 的方差得到解释。而在班级分析水平，班级环境可以解释其变异量的 80 % 左右。这意味着，如果两个班级在生活满意感和特质焦虑的平均水平上表现出差别，这种差别跟班级环境的不同有极大的关联。班级环境对于学校适应的解释能力明显低于对前两个变量的解释，其中原因需要讨论。不过除了羞怯 - 焦虑问题，对于其它适应方面都还是有不可忽视的预测力，在个体知觉水平，受班级环境解释的方差从 10 % 到 13%；在班级平均水平，受班级环境解释的方差从 7% 到 25 %。

4　讨论与结论

4.1　两个分析水平效果量的差异问题

从表 4 可看出，班级环境在班级水平上对因变量的预测，大大优于在个体水平上的预测，这一结果是非常合理的。在班级内个体水平上的适应差异，除了与班级环境有关而外，还与家庭、个人自身的条件等有重要关系。这时班级环境单方面的效应比较有限。而对于班级之间平均适应水平的差异，学生家庭、个体自身原因等等因素的影响力较小，班级环境的作用相对增大。这个结果也正好说明班级环境对于学生发展的重要。因为只有在这个水平上的效应量，才是对于班级环境之价值比较"纯粹"的估计。

4.2　因变量之间效果量的差异的问题

在三组学生变量中，学校适应（T –CRS 所测量者）与班级环境的关联度（效果量）显著低于另两组因变量。这一结果不易理解。因为从常理推测，在三组因变量中，受班级环境影响较大者恰恰应该是学校适应。以研究者目前的猜想，认为最有可能是 T –CRS 测量的信度有问题。误差的来源不在量表本身，而在量表的评定方式。本研究中的三组学生变量中，生活满意感和焦虑系采用自陈量表测量，而学校适应则是采取教师评定的方式。怀疑主要来自这样一个事实，T –CRS 评定的前提是：教师对学生有充分的了解。在美国，小学实行包班制，小班教学，因而教师对每个学生比较了解。但中国学校是采取大班分科教学制度（本研究样本中班级人数平均约 60 人），教师对学生个别特点的了解比较差。故推测这是导致评定误差的主要原因。

4.3　竞争维度的效应问题

如前述，本研究的结果似乎表示：竞争程度越高，适应问题越少，适应能力越强。而前人研究中有关竞争的一般结论是：竞争与较高的学业成就相联系，但高竞争通常导致情绪和社会性适应方面的问题。

对此一个非常可能的解释是：竞争与学校适应关系的上述结果乃是一种假象。它源于另一隐蔽原因。这原因就是择校和快慢分班。择校分班造成的结果是：好校和"快班"内学生成绩好，行为问题少，社会性发展亦较好，但学业竞争激烈。而一般校和"慢班"内，"差生"占多数，"差生"的主要表现就是学校适应差。但由于升学期望小，"慢班"、"差生"的学业竞争非常缓和。这样，如果不对择校和分班进行控制，就会出现竞争程度与学校适应呈正相关的结果。本研究的设计中由于未曾预计到这个问题，没有取得择校分班的资料，故无法对此假设加以检验。

总体来说，本研究的结果显示：学生个体所知觉到的班级环境，对其适应水平有相当肯定的解释力，而一个班级学生整体适应水平的高低，与这个班的班

级环境有极大关联。所以，本研究的一个总体性结论是，中国学校体制下的班级社会心理环境对于学生的发展和适应状况，具有举足轻重的作用。

参考文献

江光荣. (2004). 中小学班级环境：结构与测量. *心理科学, 27*(4), 839–843.

Bronfenbrenner, U. (1999). *Environments in developmental perspective : Theoretical and operational models*. In: S . L .Friedman & T . D . Wachs (Eds.) .Measuring Environment Across the Life Span. Washington, DC: American Psychological Association, 3–28.

Bryk, A. S ., & Raudenbush , S. W . (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA:Sage Publication.

Dew, T ., & Hubner, E. S. (1994). Adolescents' perceived quality of life: An exploratory investigation. *Journal of School Psychology, 32* (2), 185–199.

Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & MacIver, D. (1993). Development during adolescence:The impact of stage–environment fit on adolescents' experiences in schools and families. *American Psychologist, 48,* 90–101.

Fraser, B. J. (1994). Research on classroom and school climate. In:D. Gabel (Ed.). *Handbook of research on science teaching and learning.* New York: Macmillan, 493–541.

Fraser, B. J. (1999). *Using learning environment assessments to improve classroom and school climates.* In: H. J. Freiberg (Ed.). School climate : Measuring , Improving and Sustaining healthy learning environments. London: Falmer Press.

Hightower, A. D., Work, W. C., Cowen, E. L., Lotyzewski, B. S., Spinell, A. P., Guare, J. C., & Rohrbeck, C. A. (1986). The Teacher–Child Rating Scale: A brief objective measure of elementary children's school problem behaviors and competencies. *School Psychology Review, 15*(3), 393–409.

Huebner, E. S. (1991). Initial development of the Student's Life Satisfaction Scale . *School Psychology International, 12*(3), 231–240.

Johnson, D. W., & Johnson, R. T. (1974). Instructional goal structure : Cooperative, competitive , or individualistic. *Review of Educational Research, 44*(2), 213–240.

Moos, R. H. (1973). Conceptualizations of human environments. *American Psychologist, 28*, 652–665.

Paterson, L. (1991). *An introduction to multilevel modeling* , In: S. W. Raudenbush & J. D. Willms (Eds.). Schools, Classrooms, And pupils: International studies of schooling from multilevel perspective . San Diego, CA: Academic Press, 13–24.

Pianta, R. C. (1999). *Enhancing relationships between children and teachers*. Washington, DC: American Psychological Association.

Spielberger, C. (1973). *Manual for the State–Trait Inventory for Children*. Palo Alto: Consulting Psychologists Press.

Class environments and students' adaptability: A hierarchical linear modeliang

JIANG Guangrong[1], LIN Mengping[2]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior , Ministry of Education; School of Psychology, Central China Normal)

([2] Departmen t of Educational Psychology, The Chinses Universi ty of Hong Kong , Hong Kong)

Abstract　　This ex plo ration highlig hted the relationships between students' adaptability and the psy chosocial class environments in theelementary and middle school systems o f China .The class environments w ere measured by a questionnaire named " My Class" developedby the first author of the paper .Three sets of students' outcome variables, school adaptation , life satisfaction and anxiety , were chosenas the indicators of adaptability .A hierarchical linear model (HLM)showed that at the individual level, the class environment had aconsiderable effect on students' adaptability , and at the class level, the class−averaged perception of the class enviro nment could interpreta large part of the variance of the dependent v ariables.So the results indicates that in the Chinese school system , the psy chosocialclass environments are very important for students' development .

Keywords　class environment; student adaptability; hierarchical linear model

核心人际图式改变、人格改变与症状缓解的关系：个案研究 *

孙启武 [1,2,3]，江光荣 [1,2,3,**]，陶 嵘 [1,2,3]

（[1] 华中师范大学心理学院，武汉，430079）

（[2] 人的发展与心理健康湖北省重点实验室，武汉，430079)

（[3] 青少年网络心理与行为教育部重点实验，武汉，430079）

摘 要 本研究旨在通过一例抑郁症病例的 52 次心理治疗和效果评估说明核心人际图式改变、人格改变与症状缓解的关系。核心人际图式采用自编我的感受问卷测量，人格改变采用 SWAP‑200 测量，症状采用 OQ45.2 测量。结果：核心人际图式的变化伴随着临床上可观察的人格改变、症状缓解及当事人总体功能和心理健康水平的提高。症状缓解主要体现为主观困扰的降低，但在社会功能和人际关系质量上改变不明显。结论：核心人际图式、人格障碍轮廓和症状缓解具有共变关系。

关键词 核心人际图式；人格改变；SWAP‑200；个案研究

1 问题提出

1.1 问题背景

临床心理学家通常会同意心理治疗性改变的实质是人格改变这一说法。但是，心理疾病的人格基础，人格改变的具体内涵，人格改变的途径等问题，临床心理学家们则有不同的回答 (Beck, 1983;Blatt, 2008; Bowlby, 1988; Horowitz, 2004; Mikulincer& Shaver, 2007, etc.)，这些不同的回答形成丰富多样的人格理论。人们以这些人格理论为基础设计了不同的治疗方法，例如，试图改变不合理信念 (Beck,1983)、病理性质的"模板"(Horowitz, 2004)、客体关系 (Blatt, 2008) 和不安全的内部工作模型 (Mikulincer & Shaver, 2007) 等，并

就这些治疗方法的疗效开展了大量的实证研究。

但总的来说，这些研究存在以下两个问题: 首先，在治疗效果的研究文献中，研究问题通常是治疗方法和治疗效果的关系，而很少有研究关心人格变量与症状缓解之间的关系这样重要的理论问题。例如，从逻辑上讲，如果理论假设是"自我贬低"导致了抑郁症，那么，研究就需要评估治疗前后当事人"自我贬低"信念的变化及与抑郁症症状改变的关系。

其次，治疗效果研究往往采用症状指标来衡量治疗效果，但实际上，人格改变本身就是治疗效果。例如，"自我贬低"信念的改变本身就是疗效。就人格障碍而言，人格改变更是治疗的主要目标之一。已有文献报告了人格障碍的治疗效果 (Leichsen‑Ring&

Leibing, 2003; Perry, Banon, & Ianni, 1999)。但令人惊讶的是，多数对人格障碍治疗效果的研究同样采用了症状测量。

这两个问题之所以存在，可能的原因是：首先，人格改变相对来说较为困难，而症状缓解相对较为容易。其次，在研究实施上，囿于临床案例严格的筛选标准、人格改变往往需要较长时间的心理治疗以及长程治疗过程中当事人的脱落等原因，大样本研究较难实施。再次，人们对症状评估已经形成了较为一致的看法，但对人格评估却有许多不同的理论取向；并且在评估工具上，临床用人格改变的评估工具，无论数量还是质量，都不如症状评估工具。

为克服以上困难，临床上常采用个案研究法。个案研究在实施上具有一定的便利性，能为是否进一步的大样本研究提供初步依据。在评估工具方面，Shedler 和 Westen 发展的 SWAP－200(Shedler &Westen, 1998, 2004a, 2004b; Westen & Shedler, 1999a,1999b) 在新发展人格评估工具中较为突出。已有研究将 SWAP－200 用于评估心理治疗中的人格改变 (Lingiardi, Shedler, & Gazillo,2006)。

1.2 核心人际图式

近年来，在依恋理论的框架下，研究者认为早期养育经验所形成的内部工作模型对个体毕生发展具有重要影响，并且不安全的内部工作模型与 DSM-IV 诊断系统轴 I 和轴 II 的心理疾病高度相关 (Meyer & Pilkonis,2005; Mikulincer & Shaver,2007)。在测量上，成人内部工作模型最有影响的概念化是"二维度－四类别"模型 (Bartholomew &Horowitz，1991; Brennan, Clark, & Shaver, 1998)。但研究者也提出了批评。其中最为重要的两点分别是：首先，"二维度－四类别"模型对工作模型中的情绪成分未能给予充分的重视 (Pietromonaco & Barrett,2000)；其次，不同亲密关系的工作模型也不同，需首先评估不同亲密关系中的工作模型，然后才能得到一般的工作模型 (Fraley, Heffernan, Vicary, &Brumbaugh, 2011)。

在以上研究基础上，我们提出了核心人际图式的概念以强调以下三点：内部工作模型中的情绪成分，内部工作模型知识的程序性特征以及不同重要人际关系中内部工作模型的稳定性和变异性。核心人际图式指与重要他人相关的诸多认知－情感的如果－那么 (ifs–thens) 单元 (孙启武，2011; Michael& Shoda, 1995)。核心人际图式包括人际场景情境记忆，互动过程的如果－那么预期，伴随的情绪反应等。核心人际图式源于早期养育关系，成为发展过程中人际认知和人际行为的"模板"，在发展过程中既保持一定的稳定性也可能发生改变（Lopez& Gormley, 2002;Waters, Merrick, Treboux, Crowell, &Albersheim, 2000 ）。

不安全的内部工作模型对亲密他人具有较多的消极情感表征，而安全的内部工作模型对亲密他人具有较多的积极情感表征 (Bowlby, 1988; Collins &Reader, 1994)。成年个体具有不同的亲密关系，这些关系所引发的情绪感受不同。在一段时间内，这些心理表征会同时影响个体的心理状态。据此，安全的核心人际图式指个体在大多数亲密关系中都能体验到积极情感并内化为相应的心理表征，而不安全的核心人际图式指个体在有些亲密关系中能体验到积极情感，而在有些亲密关系中却体验着消极情感并内化为相应的心理表征。对临床样本而言，核心人际图式具有不安全的性质，与症状表现具有正向的关联。

在测量上，核心人际图式用个体与重要他人互动时的情绪感受测量。例如，和妈妈在一起时，我会感到温暖。在不同人际关系中，情绪效价和唤起程度不同。不同重要人际关系中情绪感受的加权反映了个体一般的内部工作模型，即核心人际图式。据此，我们发展了不同重要人际关系内部工作模型的测量方法，即我的感受问卷 (MFSR，孙启武，2011)。在普通人群中的研究表明，MFSR 具有良好的校标关联效度（以亲密关系体验量表、关系问卷和父母养育问卷等为校标）；追踪研究表明，MFSR 具有良好的预测效度（以抑郁情绪、孤独感和人际问题等为校标）。在临床样本的研究中表明，MFSR 的改变能在工作同盟和治疗结果预期之外预测症状的缓解 (孙启武，2011)。

综上，本研究试图通过个案研究法，探讨核心人际图式、人格改变和症状缓解的共变关系。根据核心人际图式的概念，本研究假设治疗前后消极情感表征数量和强度的减少，该过程伴随着人格改变和症状缓解。

2　研究过程

2.1　测量工具

2.1.1　我的感受问卷 (MFSR-12)

MFSR-12 由 12 对情感色彩相反的词汇构成，是 1 到 6 分的双极语义量表 (表 1)。作答时，要求被试首先提名 3~6 名重要他人，再分别评定和这些重要他人在一起时的感受。典型条目例如"温暖"和"冷淡"(孙启武，2011)。MFSR-12 通过评定不同重要人际关系中的感受衡量核心人际图式。以"温暖"和"冷淡"为例，若评定从冷淡变为温暖，则反映了核心人际图式性质的变化；若评定从冷淡 3 分变为冷淡 1 分，则反映了核心人际图式强度的变化。

2.1.2　效果问卷 (OQ45.2)

OQ45.2 是由 Lambert 等 (1996,2004) 发展的咨询效果评估工具，测量当事人的症状困扰、人际功能和社会功能三方面的内容，共 45 题。OQ45.2 是五级记分自评量表，得分越高，表示症状越严重。OQ45.2 已经被翻译修订为中文版，初步的证据表明 OQ45.2 中文版具有良好的信度、效度 (秦佑风，胡姝婧，2008)。

2.1.3　洗德乐－魏斯顿人格评估 (Shedler-Westen Assessment Procedure, SWAP-200)

SWAP–200 由 200 条描述人格特点的句子构成。其评定过程是：由临床心理学家根据当事人的材料，如录音、录像等评估当事人是否符合以及在多大程度上符合这些句子所描述的情形（0~7 分）。录音可以由专门的 SWAP 访谈获得，也可根据治疗会谈过程录音获得。SWAP–200 是基于 Q 分类法的人格评估，

要求在 200 个项目中，得分从 7 分到 0 分的条目分别为 8、10、12、14、16、18、22 和 100 条。SWAP-200 根据 Q 分类结果计算人格诊断的得分，并绘制人格诊断轮廓图。本研究采用的 SWAP-200 评分系统 3.05-2 试用版。该版本由第一作者于 2006 年从 http://www.SWAPassessment.com 获得并译为中文。

2.2　施测程序

本个案研究发表前，已经取得当事人对个案发表的知情同意。当事人签署知情同意后参与研究，对所有治疗会谈进行录音。在开始心理治疗之前，对当事人进行精神科初始访谈 (SCID-CV,First, Spitzer, Gibbon, & Williams, 1996; SCID-II, First,Spitzer, Gibbon, &Williams, 1997)。根据 SWAP-200 的测量要求 (Shedler & Westen, 1998, 2004a, 2004b)，两位临床与咨询心理学方向博士生根据前 8 次会谈录音对当事人进行首次 SWAP-200 评估，另外两位临床与咨询心理学方向博士生根据最后 8 次会谈录音对当事人进行第二次 SWAP-200 评估。其中，不一致的评定条目经讨论后达成一致。初次访谈后，施测 MFSR-12；研究结束时，再次施测 MFSR-12。当事人每月自评 OQ45.2 一次，共自评了 12 次。所有施测过程由研究助手完成。当事人为高校本科生，寒暑假期间未进行会谈。整个研究持续时间为 1 年半（2009 年 2 月 ~2010 年 12 月）[①]，期间共进行了 52 次咨询会谈，每周会谈一次，1 次 50 分钟。

2.3　个案描述

2.3.1　个案概况

小北，男，21 岁，为某高校本科生。大二上不能正常参加学习活动，多门功课不及格，不能控制情绪，说话音量大，易激惹。精神科初始访谈结果为，在轴 I 上，当事人符合躁郁症 II 型标准；在轴 II 上，当事人符合抑郁型人格障碍标准，具有回避特点；在总体功能评估上，当事人得分为 50 分，显示个案功能受损严重。大二下学期，小北开始接受心理治疗直至大四上学期结束。大二下学期接受

---

① 　2013 年 5 月 1 日我国的精神卫生法正式实施，非心理治疗师不能进行精神障碍的治疗，非医疗机构不能进行心理治疗。本文所呈现为精神卫生法颁布之前的抑郁症个案治疗，在非医疗机构进行，时间为 2009 年 2 月 –2010 年 12 月。特此声明。

治疗时，小北同时接受抗抑郁药物治疗。大三下学期，小北逐渐减小药物服用剂量并在大三下学期停止药物治疗。

2.3.2 个案简史

小北自幼受到母亲的严格管教。母亲曾任公职，后辞职在家。父亲收入丰厚。母亲掌管家中一切，几乎所有的事情都要按照她的意志完成。如果小北和他父亲没有顺从她的意志，她会想尽办法让他们顺从，并且每次都成功。在小北眼里，父亲儒雅，有能力，但显得懦弱。

小北知道母亲会出些什么招数让他就范。他曾尝试不那么听话，但很快就在母亲的招数下乖乖地就范。小北认为他从未在母亲那里体会到过温暖和温情，也很少从母亲那里得到赞扬或者安慰；有的只是控制，而他自己就是那只"提线木偶"。

小北父母经常争吵。每次家中发生争吵时，他会躲在自己的房间对着电脑玩游戏。玩游戏时，小北时刻尖着耳朵听在发生什么。小北说他并不关心事态如何发展，而是关心自己会不会受到什么威胁。在描述这个场景时，小北一直处在担惊受怕的状态中。

在小学阶段，小北是长辈眼中听话的孩子，亲戚都赞扬小北的妈妈对小北的教育非常成功，不像他们自己的孩子那么调皮捣蛋。到小学五年级的时候，小北已经人高马大，但经常受同学，包括女同学的欺负。他是同学眼中的老好人，这种情况一直延续至初中。

高中时，当语文老师以他的作文作为"反面教材"宣读时，小北对着他讨厌的"那个老头"情绪激动地训话。小北叙述说，那种训话方式，疾风骤雨，不带一个脏字，而且他也不知道当时怎么会那么流畅，不加思考地就讲了那么久。小北认为这是他母亲的方式，他无意中学习到了。高考时，小北超水平发挥，如意来到某重点高校就读。

大一上学期，小北和所有大学新生一样，积极参加各种活动，还曾在英语的演讲比赛中获奖。大一下学期，小北某必修课考了 54 分。暑假时，小北的母亲给他安排了家教，要求小北暑假补习。大二上学期，小北出现情绪激动且不能自控，说话声音很大

等症状，以至于他的辅导员认为他有很强的攻击性。

2.4 治疗过程简述

2.4.1 治疗要点

治疗师为从业 9 年的男性治疗师，临床心理学方向博士生。治疗师采取了整合的治疗模式。按照治疗的不同进程，治疗要点如下：(1) 建立良好的治疗关系和安全的治疗氛围，与当事人讨论治疗目标与治疗任务；(2) 采用澄清、指导等方法培训当事人详细叙述事件和感受的能力；(3) 强调此时此地情绪感受的觉察，并将此时此地情绪感受的觉察与重要人际关系相联系；(4) 运用治疗过程中由当事人激发的治疗师的感受作为材料，对当事人进行同感理解和解释 (Bowlby, 1988)。

2.4.2 治疗过程

在会谈初期，小北说话声音很大，难以觉察情绪。在此阶段，治疗会谈的焦点集中于小北"此时此地"的情绪以及生活事件前后的情绪上。小北起初坚持认为他既不悲伤、痛苦，也不高兴、快乐，"没有什么情绪"。随着会谈的进行，小北逐渐地能意识到他的"狂躁"的情绪状态，并开始能觉察到高兴、冷漠、抑郁、低沉、生气等情绪。在会谈初期，小北对是否能治好不抱希望，但在会谈中，他逐步意识到自己的情绪状态与母亲对待他方式有关联。由此，小北渐渐地不再那么焦虑，觉得"有些安心"，开始对治疗康复抱有一些希望。

随后的会谈内容集中于小北与母亲互动的早期记忆以及现阶段小北与母亲的互动方式上。在会谈中，小北倾向于理智地分析事情的来龙去脉。这个时候，治疗师会向小北指出，要关心自己的情绪，从情绪了解自己的感受和行为，而不是用理智的方式思考和母亲的关系。这些时刻，治疗师会与小北共同探索当时可能的情绪感受。渐渐地，小北能把自己理性思维背后真实的情感和母亲对待他的态度联系在一起。

开始能表达他对母亲的不满，其社会功能也得到逐步提高。例如，有时对母亲"说一套做一套"，但也听母亲的话开始补习之前落下的功课、自己主

动参加考试、偷偷参加网络游戏公司的应聘等等。

但是，小北始终没有胆量直接去冲撞他的母亲，并坚持认为那样"太可怕了"。后期的会谈，治疗师和小北就此共同一起工作。小北会描述生活中与母亲互动的具体细节来说明他的无能为力感。治疗师在会谈中尝试扮演小北，而小北在这个时候总能扮演其母的反应。这种会谈模式持续了数周。小北认为他母亲无法改变，在目前的状态下，他和母亲的互动方式无法改变，所以他无法改变。治疗师由此感到无力、无助的状态。治疗师把这种状态告诉小北，指出"我可以体会到你的无能、无力感了，你看这几周的会谈，我都没有打败你所扮演的那个妈妈"。

此次解释之后，这种会谈模式没有再次出现。小北能和治疗师一起将治疗会谈集中于两点：一是向母亲表达真实想法不会有灾难性的后果。例如，他偷偷上网被母亲发现后，好像也没发生什么事情，小北渐渐地对"摆脱"母亲的控制有了信心。二是开始具有扩大其人际关系的动机后，讨论如何扩大交往的圈子。

大四上，小北和其他同学一样开始找工作并通过所有不及格功课的补考。

3 研究结果

3.1 核心人际图式的变化

治疗过程中，小北未实质性地提及父母之外的亲密关系。在 MFSR-12 上，小北提名的其他重要他人在治疗前后的评定未发生变化。表 1 呈现了前后两次测量小北对妈妈和爸爸在核心人际图式条目上的分数。从表 1 可以看到，小北核心人际图式的变化首先体现在核心人际图式消极评价的强度上，例如无忧无虑条目上对妈妈的评价从 5 分变为的 4 分；其次体现在核心人际图式评价性质的变化上，例如在和睦条目上对妈妈的评价从矛盾转变为和睦。总体上，对妈妈的评价变化向积极方向改变了 12 个分数，而对爸爸的评价在具体的条目上发生了变化，但总的变化为 0。

3.2 SWAP-200 人格诊断轮廓的变化

SWAP-200 前后两次评估结果见图 1。根据 SWAP-200 的数据，T 分数达到 60 即提示符合 DSM-IV 的人格障碍诊断标准，T 分数在 55 和 60 之间为具有某人格障碍"特征"。

在时段 1，小北在抑郁型人格障碍上的 T 分数最高，为 61.9。其次在回避型人格障碍上 T 分数为 60.1，其临床表现是拒绝与人来往，拒绝卷入他人事务。再次，小北具有分裂型人格障碍特征，T 分数为 59.8。小北还具有依赖型人格障碍的特点和被动攻击的特点。

在时段 2，小北人格诊断轮廓图中的分裂型、回避型、依赖型、被动攻击的 T 分数都降至 55 分以下，抑郁型的分数从 61.9 降低为 55.5，提示在研究

表 1 核心人际图式在心理治疗前后的变化

认知情绪单元条目	妈妈前测	妈妈后测	爸爸前测	爸爸后测
无忧无虑 --- ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ -- 忧虑	5	4	4	3
温暖 ------- ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ -- 冷淡	5	4	3	4
紧张 ------- ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ -- 轻松（R）	1	2	5	3
满足 ------- ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ -- 不满	6	5	4	3
安全 ------- ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ -- 害怕	6	5	3	3
忧心 ------- ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ -- 安心（R）	1	2	3	4
高兴 ------- ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ -- 生气	6	4	3	3
痛苦 ------- ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ -- 幸福（R）	2	2	4	4
和睦 ------- ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ -- 矛盾	6	3	4	5
开心 ------- ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ -- 压抑	5	5	4	3
自在 ------- ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ -- 拘束	5	5	3	4
沉闷 ------- ① -- ② -- ③ -- ④ -- ⑤ -- ⑥ -- 有趣（R）	1	2	3	4

注：R 表示反向记分条目。
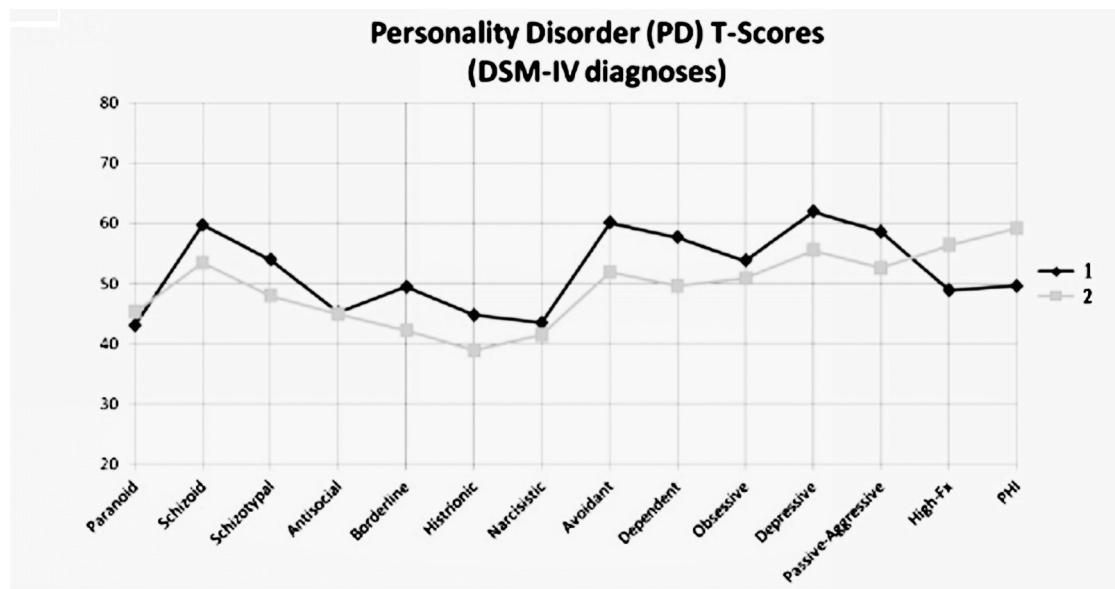
图 1　SWAP-200 人格诊断轮廓图

注：1 表示前 8 次会谈评估结果，2 表示最后 8 次会谈评估结果；Paraniod= 偏执型，Schizoid= 分裂型，Schizotypal= 分裂样型，Antisocial= 反社会型，Borderline= 边缘型，Histrionic= 表演型，Narcisistic= 自恋型，Avoidant= 回避型，Dependent= 依赖型，Obsessive= 强迫型，Passive-Aggressive= 被动攻击型，High-Fx= 社会功能，PHI= 总体心理健康水平。纵轴表示 T 分数。

结束时，小北仍然具有抑郁型人格障碍的特点。小北在时间点 2 评定的社会功能（T 分数从 48.9 增加到 56.4）和整体健康水平都有显著提高（T 分数从 49.6 增加到 59.2）。前后两次评估人格诊断轮廓图表明，小北病理性的人格特征在治疗后期产生了显著改善。

3.3　症状自评的变化

小北在 OQ45.2 上症状自评的变化见图 2。从图 2 可以看到，小北在治疗期间，反映其主观不适症状困扰的自评分数呈明显下降趋势，社会功能表现呈略微下降趋势，而人际关系质量则无明显改善。

4　讨论

一年半的心理治疗后，小北核心人际图式的变化体现在两个方面，首先是性质的变化，其次是强度的变化（表 1）。小北具有回避的人格特点，治疗初期对人际情境中的情绪觉察不足。用成人依恋的类别来讲，小北对爸爸的表征具有忽视型的特点，具有肤浅的积极情感，更多是一种理智上的认知，而非真实的情绪体验；小北对母亲的表征具有害怕型的特点，多为消极情感。值得注意的是，当事人对父亲的感受既有积极方向的变化，也有消极方向的变化（表 1）。我们认为，这种变化反映了当事人对由父亲引起的感受表征更加准确。同时表明，MFSR-12 对忽视型的工作模型和安全性的工作模型缺乏足够的区分度。

MFSR-12 的测量目标是当事人在不同重要人际关系中情绪感受的心理表征（即核心人际图式），并在理论上假设消极情绪感受的减少是人格改变和症状缓解的基础。本案例中，治疗师试图干预当事人的核心人际图式，并直接评估当事人核心人际图式的变化。结果发现，整个治疗期间，核心人际图式的微小变化伴随着较为显著的人格改变及当事人总体功能和心理健康水平的提高（表 1，图 1）。在现实功能方面，个案通过不及格功课的考试，找工作等行为表明社会功能的明显提升。但在症状测量上，当事人自评结果表明主观困扰和心理痛苦的明显降低，而社会功能和人际关系质量并无明显改变（图 2）。

为何会出现这种结果的差异呢？ OQ45 人际关系质量测量主要是恋爱（夫妻）关系、家庭关系和朋友关系等；当事人在参与治疗期间，一方面未有恋爱关系，一方面对家庭关系和朋友关系的满意度未发生变化。OQ45 社会功能主要测量学习工作的压力大小、
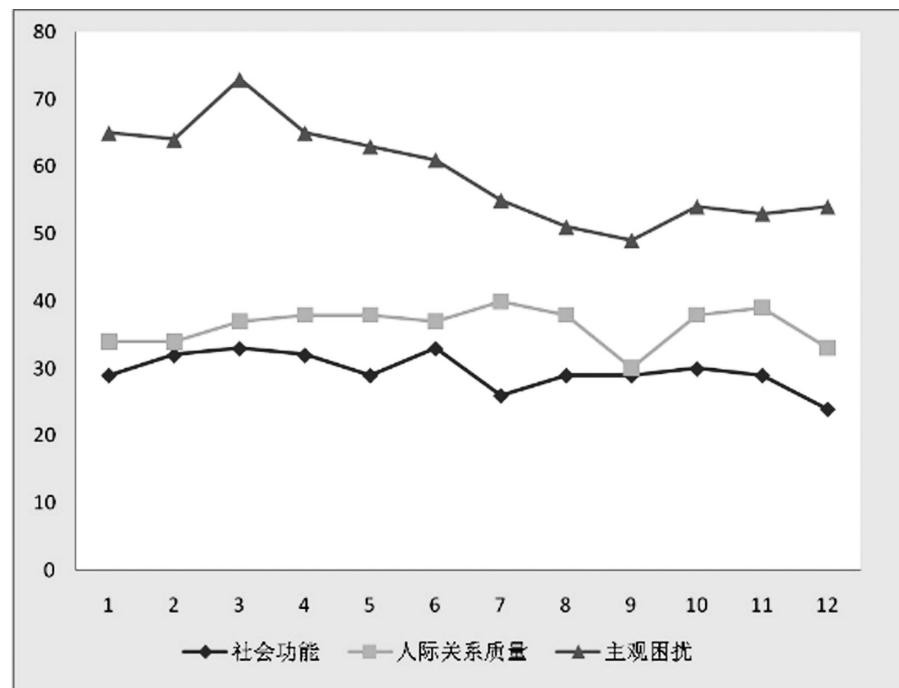
图 2 效果部卷得分在治疗过程中的变化

注：纵轴表示 OQ45.2 三维度总分，横轴表示 12 次测量。

学习工作的满意度等；在个案的主观知觉中，这些方面未有实质性的改进。SWAP-200 的总体功能和心理健康水平由"能有效的、创造性地使用他（她）的才干和能力"、"使用幽默、会意幽默"等衡量心理功能的条目测量的，而非现实意义上的婚恋关系满意度、家庭关系满意或工作学习压力的大小等，其测量对象与测量方法都与 OQ45 不同。元分析表明，在对治疗效果的评定上，不同的测量工具，自评（基于个案主观评定）和他评（基于专家客观评定）等不同的测量方法会导致测量结果的差异 (Wampold, 2001)。因此，我们认为 OQ45 和 SWAP-200 的结果的不同主要来源于测量对象的不同和测量方法的差异。

以上结果表明本案例中，直接干预核心人际图式还不足以让当事人获得社会功能和人际关系质量的明显提升。但是，核心人际图式的微小变化却可以改变病理性人格、缓解症状（主观困扰和心理痛苦），提升总体的社会功能。同时，该结果表明 OQ45 所测量的社会功能和人际关系质量与主观困扰是不同性质的治疗效果。

严格地说，MFSR-12 并未直接评估由重要他人所引发的情感，而是由重要他人所引发情感的期望。因此，MFSR-12 评估的是个体人格系统中的信念，是由重要他人引发的情绪感受的期望。Dweck(2008) 认为人格改变的本质是信念系统的改变，她认为作为人格核心的信念系统有两类，一类是自我属性的可变性，一类是关于社会接受和社会拒绝的期望，即与人际关系有关的期望。以此观之，MFSR-12 的评估对象是与关系有关信念的变化。然而，这种看法与依恋理论不符。依恋理论较为重视依恋行为及其内部表征的进化基础，因此，与其它信念系统不同，具有其它心理表征所不具有的生理起源。从理论上讲，核心人际图式还包含情绪反应方式的变化及可能生理基础的变化。目前在临床会谈中直接评估由重要他人所引发的情绪反应及生理变化在技术上还存在一定困难。这可作为今后研究的方向之一。

个案研究设计具有许多不足。首先，个案研究设计无法对心理治疗的有效性做出更多推论。本案例中，当事人同时接受了抗抑郁药物治疗，当事人的改变与药物治疗也有关联。但药物治疗的作用更多体现在主观困扰缓解上，而非体现在人格改变上。其次，

SWAP-200 的两组评定者根据录音时间和会谈内容了解会谈的先后顺序，从而对最后 8 次会谈录音产生"轻判"的倾向，这些社会期许效应都不可避免；再次，缺乏控制组，难以将当事人的改变完全归功于心理治疗；当事人生活中的其它原因，也可能最终体现在治疗效果上。个案研究虽有以上不足，但通过呈现一个完整治疗过程和治疗效果的评估，本个案研究仍具有重要价值，并能以此为基础，启示更为严谨的研究设计。从临床应用上讲，本研究启示临床实务工作者可直接针对重要人际关系中的情绪感受工作。

5　结论

在本案例中，核心人际图式、人格障碍轮廓和症状缓解具有共变关系。核心人际图式的变化伴随着临床上可观察的人格改变、症状缓解及当事人总体功能和心理健康水平的提高。症状缓解主要体现在主观困扰的降低，在社会功能和人际关系质量上改变不明显。

参考文献

胡姝婧 (2008). *心理咨询会谈中的当事人行为与咨询效果的关系* . 华中师范大学硕士论文 .

秦佑凤 , 胡姝婧 .(2008). 心理咨询效果问卷（OQ-45）在中国部分人群的试用报告 . *中国临床心理学杂志 ,16*, 138-140.

孙启武 . （2011）. *核心人际图式复杂性及影响研究* . 华中师范大学博士论文 .

于丽霞 (2009). 心理咨询会谈间体验、*工作同盟对咨询效果的影响* . 华中师范大学硕士论文 .

Baldwin, M. W. (Ed.). (2005). *Interpersonal cognition.* New York: Guilford Press.

Bartholomew, K., & Horowitz, L.M. (1991). Attachment styles among young adults:A test of a four-category model. *Journal of Personality and Social Psychology,61*, 226-244.

Beck, A.T. (1983). Cognitive therapy of depression: New perspectives. In P.J.Clayton & J.E. Barrett (Eds.), *Treatment of depression: Old controversies andnew approaches* (pp. 265-290). New York: Raven.

Blatt, S. J., (2008). *Polarities of experience: Relatedness and self-definition inpersonality development, psychopathology, and the therapeutic process.*Washington, DC: American Psychological Association Press.

Bowlby, J. (1988). *A secure base*. New York: Routledge.

Brennan, K.A., Clark, C. L., & Shaver, P. R. (1998). Self-report measurement ofadult attachment: An integrative overview. In J. A. Simpson & W. S. Rholes(Eds.), *Attachment theory and close relationships* (pp. 46-76). New York:Guilford Press.

Collins, N., & Read, S.J. (1994). Cognitive representations of attachment: Thestructure and function of working models. In D. Perlman & K. Bartholomew(Eds.), *Advances in personal relationships* (Vol.5, pp. 53-90). London: JessicaKingsley.

Dweek, C.S.(2008). Can personality be changed? The role of beliefs in personalityand change. *Current Directions in Psychological Science,17(6)* :391-394.

First, M. B., Spitzer, R. L., Gibbon, M., &Williams, J. B.W. (1996). Structuredclinical interview for DSM‐IV Axis I Disorders, *Clinician Version (SCID-CV).* Washington, DC: American Psychiatric Press.

First, M. B., Spitzer, R. L., Gibbon, M., &Williams, J. B.W. (1997). *StructuredClinical interview for DSM-IV Personality Disorders (SCID-II).* Washington,DC: American Psychiatric Press.

Frank, J. D., Hoehn-Saric, R., Imber, S.D., Liberman, B. L., Stone, A. R. (1978).*Effective ingredients of successful psychotherapy*. New York: Brunner/Mazel.

Fraley, R.C., Heffernan,M.E., Vicary, A. M., & Brumbaugh, C. C. (2011). Theexperiences in close relationships-relationship structures questionnaire:A method for assessing attachment orientations across relationships.

Psychological Assessment, 23(3), 615–625.

Horowitz, L. M. (2004). *Interpersonal foundations of psychopathology*. Washington,DC: APA.

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid,R. C., et al. (2004). *Administration and scoring manual for the OutcomeQuestionnaire–45* . Orem, UT: American .

Lambert, M., & Cattani–Thompson, K. (1996). Current findings regarding theeffectiveness of counseling:Implications for practice. *Journal of Counselingand Development, 74* , 601–608.

Leichsenring, F., & Leibing, E. (2003). The effectiveness of psychodynamic therapyand cognitive behavior therapy in the treatment of personality disorders: Ameta–analysis. *American Journal of Psychiatry, 160* , 1221–1232.

Lopez, F. G., & Gormley, B. (2002). Stability and change in adult attachment styleover the first–year college transition: Relations to self–confidence, coping, anddistress patterns. *Journal of Counseling Psychology, 49(3)*, 355–364.

Mischel, W., & Shoda, Y. (1995). A cognitive–affective system theory ofpersonality: Reconceptualizing situations, dispositions, dynamics, andinvariance in personality structure. *Psychological Review, 102*, 246–268.

Meyer, B., & Pilkonis, P. A. (2005). An attachment model of personality disorder.In J. F. Clarkin & M. F. Lenzenweger (Eds.). *Major theories of personality*. New York: The Guilford Press.

Mikulincer,M., & Shaver, P.R.(2007). *Attachment in adulthood: Structure,dynamics and change* . New

York: The Guilford Press.

Perry, J.C., Banon, E., & Ianni , F. (1999). Effectiveness of psychotherapy forpersonality disorders. *American Journal of Psychiatry, 156* ,1312–1321.

Pietromonaco, P. R., & Barrett, L. F. (2000). The internal working models concept:What do we really know about the self in relation to others? *Review of GeneralPsychology 4(2)* , 155–175.

Shedler, J., & Westen, D. (2004b). Refining DSM–IV personality disorderdiagnosis: Integrating science and practice. *American Journal of Psychiatry,161*, 1350–1365.

Shedler, J., &Westen, D. (1998). Refining the measurement of Axis II: A Q–sortprocedure for assessing personality pathology. *Assessment, 5*, 333–353.

Shedler, J., &Westen, D. (2004a). Dimensions of personality pathology: Analternative to the five factor model. *American Journal of Psychiatry, 161*,1743–1754.

Waters, E., Merrick, S., Treboux, D., Crowell, J., & Albersheim, L. (2000).Attachment security in infancy and early adulthood: A twenty–yearlongitudinal study. *Child Development, 71(3)* , 684–689.

Wampold, B. E. (2001). *The great psychotherapy debate*. Mahwah, NJ: LawrenceErlbaum and Associates.

Westen, D.,& Shedler, J. (1999a). Revising and assessing Axis II, Part I:Developing a clinically and empirically valid assessment method. *AmericanJournal of Psychiatry, 156*, 258–272.

Westen, D., & Shedler, J. (1999b). Revising and assessing Axis II, Part II: Towardan empirically based and clinically useful classification of personalitydisorders. *American Journal of Psychiatry, 156*, 273–285.

The relation of core interpersonal schema change, personality change and symptom reduction: A case study

Sun Qiwu[1,2,3] , Jiang Guangrong[1,2,3], Tao Rong[1,2,3]

([1]School of Psychology, Huazhong Normal University, Wuhan, 430079)

([2]Key Laboratory of Human Development and Mental Health of Hubei Pracince, Wuhan, 430079)

([3]Key Laboratory of Adolescent Cyber Psychology and Behavior Ministry of Education , Wuhan, 430079)

Abstract　　The nature of psychotherapeutic change was traditionally viewed as personality change. However, in the fi eld of process−outcome research,most studies reported symptom reduction instead of personality change. The proposed relationship between the change of personality variables (e.g.,the irrational beliefs) and the symptom reduction (e.g., depression) was not fully tested. Under the framework of attachment theory, the current researchintends to fi nd the change process of core interpersonal schema, and the psychotherapeutic outcome assessed not only by symptom reduction, but alsoby personality change. Core interpersonal schema means the mental representation of self and others across close relationships. Though core interpersonal schemaincludes both cognitive and emotional elements, it specifically means the emotional response elicited by the recall of a significant figure. Themeasurement for representations of two or more close relationships would be better than measuring only one close relationship in predictingindividual＇s well−being. Based on this concept, My Feelings in Close Relationships (MFSR−12; Sun, 2011) was developed to measure emotionalresponse elicited by significant figures. Theoretically, the negative emotional responses were the foundation of psychopathological personalitystructures and the overt symptoms, and the reduction of those negative emotional responses elicited by the signifi cant fi gures was the exact process ofpersonality change in effi cient psychotherapy, accompanied by clinically observed personality change and symptom reduction. The relationships among core interpersonal schema, personality change and symptom reduction were illustrated by a clinical case, Xiao Bei, whowas initially diagnosed as having depression personality disorder (SCID−II; First, Spitzer, Gibbon, &Williams, 1997) and bipolar depression II (SCIDCV;First, Spitzer, Gibbon, & Williams, 1996). My Feelings in Close Relationships (MFSR−12; Sun, 2011), Shedler−Westen Assessment Procedure(SWAP−200; Shedler & Westen, 1998) and Outcome Questionnaire (OQ45.2; Lambert et al., 2004) were administrated. Among them, MFSR−12 andSWAP−200 were tested twice, at the beginning stage and the fi nal stage respectively, and OQ45.2 was tested once per month. The results show that: (1) core interpersonal schema, personality change measured by SWAP−200, and symptom reduction were changed after52−sessionpsychotherapy; (2) The clinically signifi cant personality change , symptom reduction and the improvement of the global function wereobserved, accompanied by small changes of core interpersonal schema; (3) The score of symptoms distress was signifi cantly reduced when scores ofinterpersonal relationship and social role kept relatively stable during the period of psychotherapy.

Keywords　　core interpersonal schema; personality change; SWAP−200; case study

归因、自我效能和社会容认度对心理求助行为的影响

夏 勉，江光荣

(华中师范大学心理学院 , 武汉 430079)

摘 要 在江光荣心理求助行为的"阶段 – 决策模型"框架下 , 探查处于第三阶段的被试的求助行为 , 侧重研究(1)对心理问题的归因、(2)作为心理咨询当事人的自我效能和(3)心理求助行为的社会容认度这三个变量对实际求助行为的影响 , 同时探讨变量之间的作用机制。结果表明 , 这三个变量对求助行为均有预测作用。其中心理求助行为的社会容认度除了对求助行为有直接作用外 , 还以作为心理咨询当事人的自我效能为中介变量间接影响求助行为。整个模型对求助行为解释的变异量达到 26%。

关键词 归因 , 自我效能 , 社会容认度 , 心理求助行为。

分类号 B849:R395

1 导言

心理求助是指个体在遇到心理困扰或障碍的时候 , 向个人之外的力量寻求帮助以达到解决困扰的过程。现有的研究表明 , 许多需要心理帮助的人并不寻求专业援助 。而东方人 (包括中国人) 较之西方人对求助的态度更为消极 , 以致不少人最后出现较严重的心理疾病甚至精神崩溃。费立鹏等 (2002) 的调查发现 , 在中国的自杀死亡者中 , 仅有 17% 的人死前曾寻求过专业心理帮助。事实上 , 即使在心理咨询服务很容易获得的地方例如大学 , 人们也不愿意向心理咨询专业人员求助。因此 , 为了更好地帮助那些有潜在治疗需要的人 , 就必须了解他们不愿意求助的原因。

以往的研究从人口统计学因素、社会文化因素、心理因素三个方面考察妨碍个体求助的原因。在人口统计学方面 , 研究者发现一个非常突出的现象是男性较女性对心理求助持更消极的态度 , 并发现这跟主流文化中男性性别角色观念有直接联系。在社会和文化方面 , 研究者对受文化影响 , 与求助有关的偏见、信念和价值观等因素进行了较广泛的调查 , 并发现其中一些因素对当事人的求助行为有影响。从个体的心理因素探讨原因的研究相对较少。已有的研究探讨了当事人的情绪管理能力 , 归因方式和控制感 , 对治疗的恐惧、自我表露和自我隐藏、相互依赖和自我依赖、自我效能等因素与求助的关系。

在上述三类研究中 , 对人口统计学因素和社会文化因素研究比较多 , 而对心理因素的研究的较少。然而较近的研究表明 , 社会和文化的变量合在一起 , 对实际求助行为的解释比例仍然在一个较低的水平。这提示我们应该关注心理变量所起的作用。另外 , 国

内外的大部分研究对求助态度和求助行为没有做出恰当的区分，大部分研究考察的是求助态度，极少考察求助行为，其原因主要是观察样本难以获得。已有研究表明，尽管态度与行为可能会存在某种程度的相关，但这种相关受态度的强度、态度的稳定性、态度的可获得性、态度的醒目程度、情境压力等一系列因素的影响，因而实际相关并不能高到可以用态度代替行为进行研究的程度。如果能够直接研究实际求助行为的话，显然更有科学价值。

为了深入考察当事人求助前的心理决策过程，本研究以江光荣 (2006) 提出的心理求助行为的阶段 – 决策模型作为理论基础。该阶段模型认为求助行为包括先后相继的三个阶段：（1）心理问题自我觉察阶段——当事人是否察觉自己有心理问题；（2）自助评估阶段——当事人认为自己能不能有效地解决问题；（3）他助评估阶段——当事人决定是否求助于个人之外的力量，及向何种力量求助。本研究聚焦于第三阶段的当事人，即意识到自己有心理问题且不能自己有效解决的大学生，试图探索影响他们向专业机构求助的动机性因素。

江光荣，王铭的研究初步发现自我效能是影响个体求助的一个变量。余晓敏，江光荣进一步将自我效能分为解决心理问题的自我效能和作为心理咨询当事人的自我效能，并发现这两种自我效能对求助态度都有显著影响。但在这两项研究中，都没有考虑求助的阶段因素。从上述"阶段 – 决策模型"来考虑，个体解决个人心理问题的自我效能可能主要在第二阶段起作用，而作为心理咨询当事人的自我效能则主要影响处于第三阶段 (他助评估阶段) 的当事人。故本研究只考察作为心理咨询当事人的自我效能。

国外有少量研究涉及归因对求助行为的影响。这些研究大部分以维纳的归因理论为基础，考察归因的三个维度 (稳定性、部位、可控性) 与求助行为的关系，其中研究较多的是部位和可控性。结果发现，可控性归因对求助态度有影响，部位与求助行为的研究则得到了不一致的结果。有些研究发现，如果个体把造成心理问题的原因归于内部因素时，则倾向于不求助，当归因于外部因素时，则倾向于求助。另外一些研究则发现，内部 – 外部归因对求助行为没有显著影响。在"阶段 – 决策模型"中，对心理问题的归因也被认为是一个影响求助的因素，但这个因素在各阶段起作用的情形却不易推测，本研究拟探索归因的两个维度 (部位和可控性) 在第三阶段起作用的情况。

此前的研究者一直比较重视社会文化因素对求助行为的影响，然而实际调查结果却发现社会文化对求助行为的解释比例比较低。江光荣认为，这是因为所有外在的因素都要通过个体的内在表征起作用，因此实际上对个体产生影响的是个体知觉到的社会文化。其他研究者也提到过类似的观点，如 Fishbe in 和 A jzen 认为主观社会规范是影响行为意向的重要变量。据此，江光荣用一个新的概念——"社会容认度"——来表征个体知觉到的社会文化信念。这个概念是指在个体知觉中，公众对某种行为的接受——排斥程度。这样就将社会文化变量转化为个体的心理变量。"阶段 – 决策模型"假设：心理求助行为的社会容认度是在他助评估中起作用的一个重要变量，本研究也将探讨这个变量的影响力。

本研究除了考察上述三个变量对求助行为的影响，还准备考察变量之间的作用方式或机制。有研究提示，归因会对自我调节过程中的自我效能因素产生一定的作用，这是因为人们对其行为的自我归因，能产生一种情感体验，这种情感体验会影响后继行为。但归因与社会容认度是否存在关系我们还无法推测。至于社会容认度跟自我效能的关系，班杜拉曾指出社会文化会影响自我效能，因为文化中对于特定行为的价值规范以支持、鼓励或约束的

方式影响着文化中的个体。在经验积累的基础上，会影响到个体的自我效能感。心理求助行为的社会容认度体现着文化中有关求助行为的价值规范。如果个体知觉到的心理求助行为的社会容认度较高，那么知觉到社会偏见如羞耻、丢面子等就比较低，这一高一低的经验，最终会使当事人对心理咨询专业服务及专业人员的信任程度和接受程度提高，对求助的有效性，对可预见的与咨询有关的压力也都会持比较积

极的态度。而这些都是当事人自我效能的核心成分。

因此本研究假设归因和社会容认度会影响作为心理咨询当事人的自我效能,即归因和社会容认度除直接影响求助行为外,还以自我效能为中介变量影响求助行为。

2 研究程序与方法

2.1 取样

本研究以实际求助行为为因变量,以克服此前的研究多以求助态度或求助意图为因变量的局限。为此选择了两组被试,一组是意识到自己有心理困扰且不能自己有效解决,正在向专业机构求助的被试(以下称为"主动求助组")。一组是同样意识到自己有心理困扰且不能自己有效解决,但没有向专业人员求助的被试(以下称为"未求助组")。两组样本均取自武汉市7所高校。

以往的研究发现,求助经验会影响个体求助的意愿。为了控制求助经验的影响,曾经向专业人员求助过的被试不纳入取样范围。确定未求助组被试的程序如下:被试在回答问卷的主体部分前须回答三个问题:①您是否曾经向心理咨询师求助过?②"在最近一段时间内(三个月),您在多大程度上感到自己有过心理困扰?"(以 Likert 7 点量表由低到高记分);"如果没有求助,您觉得您自己是否能有效解决。"同时满足三个条件(第一题回答"否",第二题得分 >5,第三题回答"不能自己有效解决")的学生,被作为未求助组被试。

具体取样过程如下:将 180 份问卷发放到六所高校的心理咨询中心,请经过教育部《大学生心理健康评定量表》筛查有问题的学生填答 *,这些学生均为大一学生,回收 150 份有效问卷,有效回收率 83.3%,其中符合未求助组条件者 65 份。为了取得其它年级学生的样本,另将 720 份问卷发放到七所高校的公共课堂,当堂发放,当堂回收,共回收有效问卷 700 份,有效回收率 97.2%,其中符合未求助组条件者 89 份。这样未求助组共获得有效问卷 154 份。

主动求助组被试以到 7 所高校的心理咨询中心主动求助且属首次求助的学生构成。为避免会谈经验的影响,问卷请被试在第一次咨询前填答。共发放问卷 180 份,回收有效问卷 151 份,有效回收率为 76.6%。

研究的最终样本量为 305,其中未求助组 154,主动求助组 151。被试的具体资料见表 1。

表 1 被试的基本资料 (N =305)

组别	性别		来源			年级					专业				
	男	女	大城市	中小城市	乡镇农村	大一	大二	大三	大四	研究生	文科	理科	工科	农科	医科
求助组	69	82	17	47	87	40	48	34	23	6	60	37	33	4	14
未求助组	82	72	19	40	95	42	73	32	7	0	27	54	41	8	24
总 体	151	154	36	87	192	82	121	66	30	6	87	91	74	12	38

2.2 测量工具

2.2.1 作为心理咨询当事人的自我效能 作为心理咨询当事人的自我效能包括以下一些信念:相信自己能够坚持完成咨询,能够有效地与咨询员交流自己的问题,能够处理心理咨询中的突发事件,并相信自己能从咨询中获益,等等。

此变量由问卷测量。问卷由余晓敏,江光荣(2004)编制的《作为心理咨询当事人的自我效能问卷》修订而来。原问卷包括 10 个项目,探索性因素分析结果表明,这 10 个项目较好地聚合在单一的维度上,可以解释的变异量为 41.68%,内部一致性信度为 0.84。在本研究中,用验证性因素分析进行修订,修订后的问卷由 8 个项目组成,有较好的结构效度,问卷的内部一致性信度为 0.81。自我效能问卷验证性因素分析的结果为:χ^2=74.20, df=20, χ^2/df=3.71, RMSEA=0.094, CFI=0.92, IFI=0.92, NNFI=0.89, GFI=0.94。项目举例:"我相信自己能够自如地和心理咨询员讨论我的问题"。采用 5 点 Likert 量表记分,分值越高代表自我效能越高。

2.2.2 归因 对心理问题的归因只包括部位、可

控性两个维度。由自编问卷测量，每个维度包括一个测量问题。部位维度："假如这段时间内，您经历了一些心理困扰，您觉得您的心理困扰在多大程度上是自己的原因造成的？"；可控性维度："假如这段时间内，您经历了一些心理困扰，您觉得造成您心理困扰的原因在多大程度上是自己可以控制的？"。问卷用6点Likert量表记分。

2.2.3 心理求助行为的社会容认度　心理求助行为的社会容认度是指在被试心目中，公众对于自己寻求心理帮助的接受——排斥程度。变量亦采用自编问卷测量。问卷采用语义差别量表的形式，题干是"假如这段时间内，您经历了一些心理困扰，并且想向心理咨询员求助，那么您觉得"您的求助行为"在社会上大多数人的眼里是——"。评价部分由14对双极形容词组成。如"明智的－愚蠢的"，"积极的－消极的"。两极间距离从－3到3，共分为7级。级数越高表示对心理求助行为的接受程度越高。问卷的α信度系数为0.91。

试测分析结果表明，该问卷为单一维度，单个维度可解释的变异量为49.23%，对正式施测结果进行验证性因素分析，χ^2=287.72, df=77, χ^2/df=3.71, RMSEA=0.094, CFI=0.91, IFI=0.91, NNFI=0.90, GFI=0.88, 结果表明单一维度的结构可以接受。

2.3 数据处理

本研究采用 SPSS 11.0 和 LISREL 8.30 进行统计分析，主要的分析方法为相关分析、Logistic 回归和结构方程模型。

3 研究结果

本研究的预测变量为：作为心理咨询当事人的自我效能、对心理问题的归因、心理求助行为的社会容认度。结果变量为求助行为。求助行为是二分变量，即求助和未求助。未求助组和主动求助组在各个预测变量上的平均值和标准差见表2，预测变量和结果变量之间的相关矩阵见表3。

3.1 预测变量对结果变量的 Logistic 回归结果

预测变量对结果变量的Logistic回归结果见表4。

由表4可知，可控性归因、心理求助行为的社会容认度、作为心理咨询当事人的自我效能这三个变量对求助行为影响都达到显著水平，而内部－外部归因没有进入最优回归方程。变量间的关联是：个体知觉中心理问题的原因越是可控的，就越倾向于不求助；个体作为心理咨询当事人的自我效能越高，越有可能去求助。个体知觉到的心理求助行为的社会容认度越高，越可能求助。从 Ward 统计值来看，作为心理咨询

表 2　未求助组和主动求助组的平均数和标准差

| 组别 | 心理问题归因 | | | | 求助行为的社会容认度 | | 作为心理咨询当事人的自我效能 | |
| | 内部－外部归因 | | 可控性归因 | | | | | |
	M	SD	M	SD	M	SD	M	SD
未求助组	4.09	0.94	3.83	1.06	4.92	0.93	4.96	0.79
主动求助组	4.04	1.15	3.61	1.15	5.33	0.89	5.62	0.61
总体	4.07	1.03	3.74	1.10	5.09	0.93	5.24	0.79

注：① ** P<0.01;* P<0.05;
②求助行为与其他变量的相关为点二列相关，其余为积差相关。

表 3　三个预测变量以及结果变量之间的相关矩阵 (N =305)

变量	M	SD	1	2	3	4	5
1 内部－外部归因	4.07	1.03	1.00				
2 可控性归因	3.74	1.10	0.058	1.00			
3 心理求助行为的社会容认度	5.09	0.93	0.054	0.025	1.00		
4 作为心理咨询当事人的自我效能	5.24	0.79	−0.139*	0.079	0.165**	1.00	
5 求助行为	0.50	0.50	−0.026	−0.096	0.217**	0.413**	1.00

表 4　预测变量对结果变量的 Logistic 回归结果

变量	B	SE	Wald	df	Sig.	Exp(B)
可控性归因	−0.380	0.122	9.637	1	0.002	0.684
社会容认度	0.475	0.150	9.967	1	0.002	1.608
自我效能	1.442	0.218	43.937	1	0.000	4.231
确定系数		Cox & Snell R Square			Nagelkerke R Square	
		0.239			0.319	

注：①进入回归方程的方法为 Forward Stepwise (Conditional)　②B 为非标准化回归系数

表 5　心理求助行为的结构方程模型拟合指数

	χ^2	df	χ^2/df	RMSEA	CFI	IFI	NNFI	R^2
虚模型	3090.72	276	11.20					
模型	560.19	248	2.26	0.064	0.90	0.90	0.89	0.26
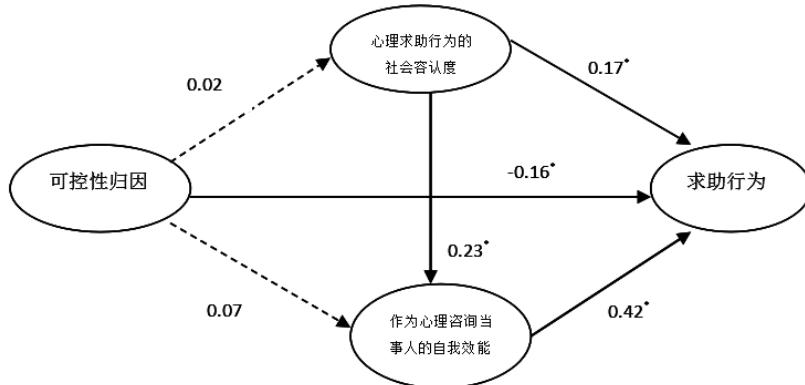
图 1　心理求助行为的结构方程模型

注：①路径系数均为标准化解②＊表示路径系数显著

当事人的自我效能作用最大，其次是心理求助行为的社会容认度，再次是可控性归因。由于自我效能、社会容认度、可控性归因的单位不同，分别为 5 级计分，7 级计分和 6 级计分，因此不再进行风险比的比较。

3.2　结构方程模型分析及中介效应检验

根据导言中的讨论，将四个变量之间的关系构建如图 1 的模型。该模型包括三个相互关联的假设：可控性归因影响求助行为（包括直接效应和通过社会容认度的中介效应以及通过作为心理咨询当事人的自我效能的中介效应）；社会容认度影响求助行为（包括直接效应和通过自我效能的中介效应）；作为心理咨询当事人的自我效能影响求助行为。结构方程的分析结果见图 1。

模型的拟合情况见表 5。

研究结果发现，可控性归因到求助行为这条路径是显著的，但是通过社会容认度和自我效能对求助行为的影响不显著。心理求助行为的社会容认度到求助行为这条路径以及社会容认度通过自我效能到求助行为这条路径都是显著的。即社会容认度除了直接影响求助行为以外，还以自我效能为中介变量对求助行为产生影响。用 LISRL 8.30 计算中介效应，发现心理求助行为的社会容认度通过"作为心理咨询当事人的自我效能"对求助行为的中介效应是 0.05，标准误是 0.01，Z 值是 3.16。社会容认度对求助行为的总效应是 0.13，标准误是 0.03，Z 值是 4.46。因为 $Z > 1.96$ ($p < 0.05$)，所以无论是社会容认度对求助行为的总效应还是通过自我效能的中介效应都是显著的，中介效应占总效应的比是 0.38。

作为心理咨询当事人的自我效能对求助行为的影响也是显著的，自我效能不但是一个重要的预测变量，而且在模型中起中介作用。总体来说模型是可以接受的，整个模型解释了求助行为变异的 26%。

4 讨论

4.1 研究实际求助行为的必要性

本研究在研究设计上的一个显著进步是研究个体实际的求助行为。由于取样及观察上的困难，以往的研究多考察求助意愿或求助态度。社会心理学的知识告诉我们，态度与行为的关系非常复杂，由态度并不能准确地预测行为。在心理求助领域，求助态度和求助行为的研究也发现了同样的情况。如 Kelly 和 Achter 的研究发现，与低自我隐藏的人相比，高自我隐藏的人求助态度更消极，但实际求助的可能性却更高。又如在对治疗恐惧的研究中，Kushner 和 Sher 发现，治疗恐惧水平的增加会引发消极的求助态度。但是 Segal, Hodges 和 Hardiman 对已经求助的被试进行研究，发现反而是那些有较高治疗恐惧的个体实际上选择了向专业人员求助。由于求助态度和求助行为并不完全一致，而我们的目标通常是想要预测行为，所以直接以实际求助行为为观测变量，其结果显然更有说服力，实践价值更大。

4.2 预测变量对求助行为的影响

本研究以认知动机理论为基础，考察了自我效能、归因和社会容认度对求助行为的影响。下面分别讨论研究得到的结果。

自我效能是指人们关于自己实现特定领域内的行为目标所需能力的信心或信念。班杜拉认为自我效能会影响个体对于未来环境和行为的选择。一般而言，个体会选择自认为他能加以有效应对的环境，而回避自感无法控制的环境。余晓敏、江光荣在此前的研究中已经发现，作为心理咨询当事人的自我效能对求助意愿有明显的预测作用。但在该项研究中，没有考虑求助阶段因素，样本中包括了处于不同求助阶段的被试。如果依阶段－决策模型，从逻辑上推测，这会低估自我效能对处于第三阶段被试求助行为的预测力。另外，该研究是以求助意愿为被预测变量，根据社会心理学的知识，实际求助行为会显著低于求助意愿。本研究一方面将求助阶段因素纳入设计考虑，另一方面改用实际求助行为为被预测变量。在此基础上比较前一研究和本研究的结果，发现当事人自我效能与实际求助行为的相关 (r=0.413) 略高于该自我效能与求助意愿的相关 (r=0.376)。可以推测，前一研究中自我效能的预测力的确被低估了，至少对处于第三阶段的被试是这样。无论如何，这两项研究表明，无论是对求助态度还是求助行为，自我效能都是一个非常重要的预测变量。

前人有研究表明可控性归因与求助意愿有关，而部位与求助意愿的关系不一致。本研究进一步考察这两个变量对实际求助行为的影响。结果发现，可控性归因对求助行为有显著影响，而内部－外部归因的作用不显著。由于考察归因与求助关系的研究积累还比较少，现在还无法推测本研究中内外归因的作用跟此前研究不一致的原因。可控性归因对求助行为的影响与以往的研究一致。个体知觉到的心理问题的原因越是可控的，就越倾向于不求助。对此的解释是：如果个体认为导致心理问题的原因是可控的，那么相对于不可控归因来说，当事人更可能产生负性反应，如自责、容易知觉到别人对自己有不良反应、掩饰行为等。作为掩饰行为的一部分，当事人可能会延迟求助或不求助。

本研究的一个尝试是解决作为群体现象的文化规范如何在个体层面上发生作用，影响个体的求助行为这个问题。这个尝试是借助提出求助行为的社会容认度这个概念实现的。文献表明，一个文化中关于心理问题的信念、该文化所赞许的心理问题表达方式和解决方式，会影响个体在遭遇心理困难时寻求专业帮助的态度。在中国文化中，患有精神障碍及因心理问题向外人求助，通常被视为羞耻的事。有关中国人性格的研究表明，中国人的一个典型心理特征是他人取向，表现为对他人的意见、褒贬特别敏感，总希望在他人心目中留下良好印象，在行为上努力与别人相一致，并尽量避免他人的责罚、讥笑、拒绝、尴尬及冲突。因此，中国人对心理求助会非常忌讳和回避。然而实际研究发现，虽然中国人的确比西方人较少利用心理卫生服务，但文化偏见所起的作用并不如预想

的那么大。问题在于，文化规范并非对同一文化中的个体都是划一地起作用。总是对有的个体影响大，有的影响小。这也许是此前的研究发现文化偏见对于求助变量的解释力并不很大的原因。Sue提出在考察与文化相关的行为时，在实践和研究中都应关注某一文化群体内的个体差异。江光荣提出的社会容认度的概念（在个体知觉中，公众对某种行为的接受——排斥程度）较好地解决了这一问题。社会容认度代表着个体对社会文化规范的内在表征，是一种个体现象。这样就可以从个体水平上考察文化变量对行为的影响。

研究结果表明，心理求助行为的社会容认度对求助行为有显著预测作用。个体知觉到的心理求助行为的社会容认度越高，越倾向于求助。这还仅仅是心理求助行为的社会容认度的影响，可以设想，如果把心理问题（罹患心理障碍）的社会容认度加进来，对求助行为的预测力还会增加。

4.3 预测变量之间的关系

有关预测变量之间关系的假设部分得到支持。结构方程模型分析表明，心理求助行为的社会容认度影响作为当事人的自我效能。进一步分析发现，社会容认度通过自我效能对求助行为产生影响的中介效应占社会容认度总效应的比例为38%，这个比例是相当高的。也就是说，如果个体知觉到的社会上大多数人对求助行为越接受，那么作为心理咨询当事人的自我效能也就比较高，个体就越倾向于求助。这个结果在实践上有一定意义：心理健康教育工作可以通过宣传，提高公众求助行为的社会容认度，从而提高潜在求助者的当事人自我效能，进而提高求助率。

但心理问题归因跟当事人自我效能之间关系的假设没有得到支持，结构方程分析表明二者之间不存在关联。分析原因，可能在于心理问题的归因跟作为心理咨询当事人的自我效能在逻辑上的不对应。按照阶段–决策模型，心理问题的归因跟求助的关系可能主要发生在第二阶段——自助评估阶段。在这个阶段，心理问题的归因可能影响着当事人独立处理个人心理问题的自我效能。但这需要另行研究。

4.4 局限

本研究以阶段–决策模型为理论基础，考察了在第三个阶段影响个体求助行为的三个认知性动机变量——社会容认度、自我效能和归因，这三个变量可以解释求助行为26%的变异。检讨整个研究设计和过程，研究者觉得这个解释比例很可能有所低估。例如问题的严重程度这个变量的控制，就不是很精确，本研究的被试在问题严重程度这个变量上，基本是以"自觉存在心理问题且这问题不能自己解决"为入选标准。相信如果引入一个较为客观精确的严重程度测量作为控制变量，上述三个预测变量的效应会有所提高。需要说明的是，本研究只选取了三个动机变量考察其预测作用。实际上在第三阶段影响个体求助行为的还有许多别的因素，如性格特点、时空便利性、支付能力等，所以我们并不预期本研究的三个变量能解释大部分求助行为的变异。另外，本研究所用的归因问卷每个维度只用一个项目测量，从测量的角度来说可能会对问卷的信度有所影响。最后，本研究所选样本是大学生，因此研究结果的可推广性会受到限制。

5 结论

（1）心理问题可控性归因对求助行为有预测作用。个体知觉到的造成自己心理困扰的原因越是可控的，就越倾向于不求助。

（2）心理求助行为的社会容认度对求助行为有预测作用，包括直接作用及以"作为心理咨询当事人的自我效能"为中介变量的间接作用。个体知觉到的心理求助行为的社会容认度越高，越倾向于求助。

（3）作为心理咨询当事人的自我效能对求助行为有预测作用。作为心理咨询当事人的自我效能越高，当事人越倾向于求助。

参考文献

Andrews, G., Hall, W., Teesson, M., & Henderson, S. (1999). *The Mental Health of Austra lians*. National

Survey of Mental Health & Well- being Report No 2. Canberra: Mental Health Branch, Common wealth Department of Aged Care.

Bryan, S. K. K., & Michael, M. O. (2003). A sian Cultural Values, Attitudes toward Seeking P rofessional Psychological Help, and Willingness to See a Counselor. *The Counseling Psychologist, 31,* 343–361.

Cheatham, H. E., Shelton, T. O., & Ray, W. J. (1987). Race, sex, causal attribution, and help–seeking behavior. *Journal of College Student Personnel, 28*(6), 559–568.

Ciarrochi, J. V., & Deane, F. P. (2001). Emotional competence and willingness to seek help from professional and nonprofessional sources . *British Journal of Guidance & Counseling, 29*(2), 233–246.

Ciarrochi, J. V., Deane, F. P., & Wislon, C. J. (2002). Adolescentswho need help themost are the least likely to seek it :the relationship between low emotional competence and low intention to seek help. *British Journal of Guidance & Counseling, 30*(2), 173–188

Corrigan, P. W., Backs– Edwards, A., Green, A., Lickey– Diwan, S., Penn, D. (2001). Prejudice, socia l distance, and familiarity with mental illness. *Schizophrenia Bulletin, 27,* 219–225.

Flisher, A. J., De, & Beer, J. P. (2002). Characteristics of students receiving counseling services at the university of Cape Town, South Africa. *British Journal of Guidance & Counseling, 30*(3), 299–310.

Kung, W. W. (2004). Cultural and practical barriers to seeking mental health treatment for Chinese Americans. *Journal of Community Psychology, 32*(1), 27–43.

Guo Benyu. (2003). *The new advance of current psycho logy.* Jinan: Shandong education press, 6–7.

[郭本禹 . *当代心理学的新进展* . 济南：山东教育出版社 , 2003 . 6 – 7.]

Jiang G. R., & Wang, M. (2003). The research about

college students' help–seeking behavior. (In Chinese) *Chinese journal of clinical psychology, 11*(3), 180–184.

[江光荣 , 王铭 . (2003). 大学生求助行为研究 . *中国临床心理学杂志* , *11*(3), 180–184.]

Jiang G. R., & Xia, M. (2006). Psychological help–seeking behavior: the research review and phase –decision model (In Chinese). *Advances in Psychological Science, 14*(6), 888–894.

[江光荣 , 夏勉 . (2006). 心理求助行为：研究现状及阶段 – 决策模型 . *心理科学进展* , *14*(6), 888–894]

Kelly, A. E., & Achter, J. A. (1995). self–conceal ment and attitude toward counseling in university students. *Journal of counseling psychology, 42*, 40–46.

Kung, W. W. (2004). Cultural and practical barriers to seeking mental health treatment for Chinese Americans. *Journal of Community Psychology, 32*(1), 27–43.

Kushner, M. G., & Sher, K. L. (1989). Fear of psychological treatment and its relation tomental health service avoidance. *Professional Psychology Research and Practice, 20,* 251–257.

Kushner, M. G., & Sher, K. L. (1991). The relation of treat ment fearfulness and psychological service utilization:an overview. *Professional Psychology Research and Practice, 22*, 196–203.

Martini, T. S. (1996). A ttributions and the stigma of illiteracy: understanding help seeking in low literate adults, *Canadian Journal of Behavioural Science, 28*(2), 121–129.

Matsuoka, J. K., Breaux, C., & Ryujin, D. H. (1997). National utilization of mental health services by Asian Americans/Pacific Islanders. *Journal of Community Psychology, 25*(2), 141–145.

Mei, J. R., & Sui, Y. J. (1998). The graduates' help – seeking tendency (In Chinese). *Chinese journal of clinical psychology, 6* (4), 210–215.

[梅锦荣 , 隋玉杰 . (1998). 大学生的求助倾向 . *中国临*

床心理学杂志 , 6(4):210 ~ 215]

Nadler, A. (1991). Help−seeking behavior: Psycholog ical costs and instrumental benefits. In: M S Clark (Ed.), *Review of persona lity and social psychology, 12,* 290–311.

Phillips, M. R., Yang, G., Zhang, Y., Wang L., Ji, H., & Zhou M. (2002). Risk factors for suicide in China:a national case−control psychological autopsy study. *The Lancet, 360,* 1728–1736.

Pugh, J. S. (2002). Help seeking and personality among college students. *The Sciences & Engineering, 62*(7 − B), 3387.

Schohn, M. (1991). *Antecedents and consequences of attributions for causes of psychopathology.* D issertation abstract international, B 52 /02, 1082.

Segal, S. S., Hodges, J. Q., & Hardiman, E. R. (2002). Factors in decisions to seek help from self−help and co−located community mental health agencies . *American journal of orthopsychiatry, 72* (2), 241–249.

Senior, V., Weinman, J., & Marteau, T. M. (2002). The influence of perceived control over causes and responses to health threats: A vignette study. *British Journal of Health Psychology, 7,* 203–211.

Sue, S.(1999). Science, ethnicity, and bias: where have we gone wrong! *American psychologist, 54*, 1070–1077.

Taylor, S. E., Replau, L. A., & Sears, D. O. (2004). Translate in Chinese by Xie X F, Xie D M, Zhang Y L. *Social psychology.* Beijing:Beijing university press, 2004. 171–178.

[Taylor, S. E., Replau, L. A., & Sears, D. O. 著 (2004). 谢晓非 , 谢冬梅 , 张怡玲等译 . *社会心理学* . 北京 : 北京大学出版社 , 171–178.]

Tsang, W. H., Tam, P. K. C., Chan, F., et al. (2003). Stigmatizing attitudes towards individualswithmental illness in Hong Kong: Implications for their recovery. *Journal of Community Psychology, 31* (4), 383–397.

Uba, L. (1994). *Asian Americans: personality patterns, identity, and mental health.* New York: Guilford.

Vogel, D. L., & Wester, S. R. (2003). To seek help or not to seek help: the risks of self−disclosure. *Journal of counseling psychology, 50*(3), 351–361.

Yang, K. S. (1993). *The social orientation of Chinese: the viewpoint of social interaction.* In Yang Kuo −Shu, Yu, A. B. The psychology and behaviorof Chinese——the theory and method. Taiwan: Guiguan press, 87–142.

[杨国枢 . (1993). *中国人的社会取向 : 社会互动的观点* . 见 : 杨国枢 , 余安邦主编 . 中国人的心理与行为——理念及方法篇 . 台湾 : 桂冠图书公司 , 1993. 87–142.]

Yeh, C. J. (2002). Taiwanese students' gender, age, interdependent and independent self − construal, and collective self−esteem as predictors of professional psycho logical help−seeking attitudes . *Cultural diversity and ethnicm inority psychology, 8*(1), 19–29.

Yu, X. M., & Jiang, G. R. (2006). The relationship of locus ofmental health, self−efficacy and help seeking of college students. *Clinical Journal School Health, 27*(5), 444–445.

Zhang, A. Q. (1999). Motivation theory:the motive research in 21st century. Wuhan: CentralChina Nor malUniversity Press, 133–134.

[张爱卿 . (1999). 动机论 : 迈向二十一世纪的动机心理学研究 . 武汉 : 华中师范大学出版社 , 133–134.]

Relationship among attribution, self-efficacy, perceived social acceptance, and help-seeking behavior

XIA Mian, JIANG Guangrong

(Schoo l of Psychology, Centra l China Norma l University, Wuhan 430079, China)

Abstract   Chinese people hesitate to seek professional help and hold negative attitudes toward help seeking. Earlier studies have examined various demographic, social−cultural and personal factors in the attempt of identifying the barriers. However, recent studies demonstrate that social−cultural variables can only explain a relatively small amount of variance in actual help−seeking behavior. Therefore, we should pay as much attention to personal factors in understanding help seeking behaviors. To understand the individual's decision−making process before help−seeking behavior, Jiang (2005) developed a Phases−Decision−Making Model (PDM). This model illustrates three phases involved in help−seeking decision: 1) the perception of having psychological problems; 2) Self−help evaluation; and 3) Other−help evaluation. To study the actual help−seeking behavior in this study, we employed this model and examined the barriers to help−seeking at the third phase. We studied those who were aware that they had psychological problems, recognized that they could not solve them effectively by themselves, and were prepared to seek help from others. In order to understand factors that may lead to help−seeking behavior, we examined attribution of psychological problems, self−efficacy of being a counseling client (the belief in one's ability of being a "good" client), and perceived social acceptance of help−seeking. Three hundred and five college students (154 female, 151 male) who were aware of having psychological problems and realized that they couldn't solve them effectively by themselves (as identified by a survey) participated in the study, with the first group being those actually sought professional help (HS) and the second never did (NHS). All the participants were enrolled from seven universities in Wuhan, China, with 39.7% sophomores, 26.9% freshmen, 21.6% juniors, 9.8% seniors, and 2% graduate students. We used questionnaires to measure the three variables. The questionnaires are The Questionnaire of Self−efficacy of being Counseling Client (Yu, Jiang, 2004), The Questionnaire of the Perceived Social Acceptance and the Questionnaire of attribution of the mental problem. The result showed that the self−efficacy of being counseling client was a significant predictor of help−seeking behavior. Higher self−efficacy led to more help seeking. Controllability predicted help seeking in that if students perceived their psychological problems as controllable they would be less likely to seek help. The locus of attribution failed to differentiate HS and NHS students. Perceived social acceptance of help−seeking predicted help−seeking behavior positively. Those who perceived higher social acceptance of help−seeking would perceive higher self−efficacy of being counseling client, higher self−efficacy of being counseling client would be positively associated with help−seeking. The self−efficacy of being counseling client and perceived social acceptance were important motivational factors that predicting help−seeking behavior in the third phase of PDM. This study supported the PDM on some extent.

Keywords   attribution; self−efficacy; perceived social acceptance; psychological help−seeking

不同情绪状态下自动表达抑制对情绪及
表情知觉敏感性的影响 *

杨　静 [1,2,3]，陶　嵘 [1,2,**]，陈锐娟 [1,2,]

([1] 青少年网络心理与行为教育部重点实验室，武汉，430079)
([2] 华中师范大学心理学院暨湖北省人的发展与心理健康重点实验室，武汉， 430079)
([3] 华中科技大学文华学院心理健康教育中心，武汉，430079)

摘　要　预备实验、正式实验的分别以 81 位、103 位大学生为被试。在正性、负性两种情绪状态下，用句子整理任务引发抑制情绪或表达情绪两种情绪调节方式，采用信号检测论的方法测量正、负性表情的表情知觉敏感性。结果表明：（1）在负性情绪状态下，表情知觉敏感性存在情绪一致性效应，即被试对负性表情更敏感；在正性情绪状态下，表情知觉敏感性不存在情绪一致性效应，即被试对正、负性表情的敏感性没有显著差异。（2）在正性（快乐）情绪状态下，三组表达方式的被试对正负表情的知觉敏感性均无显著差异。在负性（悲伤）情绪状态下，表达组、控制组的被试对负性表情的知觉敏感性高于正性表情，而抑制组对正负表情的知觉敏感性没有显著差异。

关键词　自动表达抑制；表情知觉敏感性；正性、负性情绪；情绪一致性效应

1 前言

1.1 情绪知觉及其具身模仿论模型

情绪知觉是知觉 (perceive) 或识别 (identify) 情绪，是能正确地识别人们如何感觉 (feeling) 以及识别在物体、艺术、音乐等中所包含的情绪内容的能力 (Mayer & Salovey，1997) 。识别他人的情绪有助于个体判断环境和周围人的信息，从而产生恰当的情绪和行为反应，如在特定情形下该趋近还是远离他人，是人在社会生活中一项至关重要的技能。

情绪识别的具身模仿机制是新的关于情绪识别机制的解释模型。早期 Hatfield 等人 (1993) 通过阐述情绪反馈和感染现象来解释情绪识别的过程。他们将情绪感染描述为三阶段加工过程：1. 模拟 (Mimicry)：在对话过程中，人们会自动和同步地模拟对方的各种情绪性行为。2. 反馈 (Feedback)：对对方的情绪性行为的模拟使个体的主观情绪体验和自主神经活动受到影响。3. 感染 (contagion)：个体识别出对方的情绪。后来 Hoffman(2002) 在共情的理论模型下对模拟激发情绪识别的过程进行了补充：个体通过观察对方情绪

性行为、模拟之、产生相应情绪体验、激活自身相关情绪经验的概念知识体系，最后通过结合对自身情绪体验的认识来判断出对方的情绪。而镜像神经元(Iacoboni & Lenzi，2002; Iacoboni et al.，2005) 可能是模拟现象的神经基础。在此基础上，学者提出了情绪知觉的具身模仿论模型。具身模仿是指对面部表情等的观察激发了观察者相应的运动、躯体感觉、情感和奖励系统的一种模仿状态，即对表情的观察伴随着观察者表达相应情绪的躯体和神经机制的激活，这种激活状态向观察者表征该表情的意义 (Niedenthal et al.，2010) ，这是一种共鸣式的表征。并且这种"共鸣"是无中介的，不需要意识水平上的类比 (刘亚，王振宏，孔风，2011; Gallese，2005) 。

由此可知，模拟对方表情等情绪性行为对于识别对方的情绪起着十分重要的作用，而且这种模仿很可能是在无意识水平上发生的。已有研究结果支持前者，但尚无研究在无意识水平上进行研究：

首先，对面部表情的模拟似乎是普遍存在的，Dimberg(1997) 的 EMG 研究发现模拟是非常即时和无意识的，观看情绪图片，被试面部 EMG 在数百毫秒内就会发生变化。Lishner 等 (2008) 研究发现观看表情图片时，被试会无意识地产生面部表情肌肉的运动，并产生与其一致的情绪。Bailey 和 Henry(2009) 发现老年人虽然对表情的识别能力下降，但是在面对表情图片时仍然表现出较强的模拟反应。

其次，有意识地阻止模拟面部表情会影响对情绪知觉敏感性。Niedenthal 等人 (2001) 的研究要求被试用嘴叼着一支笔用来抑制面部表情，而 Schneider(2008) 的研究则采用要求被试抑制自己情绪的表达来抑制面部表情，都发现了对情绪知觉敏感性的影响。

这些研究操作有效地阻止了对面部表情的模拟，支持模仿是情绪知觉的先决条件。但已有研究都属于有意调控，而研究发现表情知觉敏感性是受到认知负荷影响的 (Patterson & Stockbridge，1998) ，而有意抑制是需要自我监控的，需要消耗认知资源，那么有意抑制对表情知觉敏感性的影响是否是因为

抢用了认知资源或是因为抑制本身的原因呢？情绪知觉是否如具身模仿论所言，是不需要意识水平上的类比的呢，目前尚无研究对无意识阻止模拟面部表情是否会影响情绪知觉敏感性这一问题进行探索，而这一研究结果将进一步提供对具身模仿论的验证。

1.2 表情知觉敏感性及其测量

表情知觉是情绪知觉最重要的方面，本研究选择表情知觉来作为情绪知觉的指标。表情知觉敏感性受许多因素的影响，它存在着 1. 个体差异 (Martin et al.，1996; Rozin et al.，2005) 和年龄差异 (Orget et al.，2008; Montange.，2007) ; 2. 受到药物的影响 : 安定和酒精会对特定表情的识别造成一定的影响 (Borrill, Rosen & Summerfield，1987; Coupland et al.，2003) ; 3. 并与广泛的心理病理状况相关 : 精神分裂症、心境障碍、边缘人格障碍、创伤后应激障碍、焦虑障碍等患者存在特定和不同程度的表情的表达困难或识别困难 (Corcoran，2008) 。尽管这种表情表达和识别困难的性质、特点和意义尚不清楚，但推测它们起码可能与患者的人际和社会适应困难有关。

影响表情识别的另一类因素是和情绪状态和认知负荷 (Patterson & Stockbridge，1998) : ，对情绪表达信息的加工过程受到观察者当时情绪状态的影响，个体倾向于更有效的加工与自己所处情绪一致的刺激(Niedenthal et al .，2001; Niedenthal et al.，2000) ——即情绪一致性效应 (mood congruent effect) ，但这种效应的产生原因目前尚需进一步探讨。Tracy 和 Robins(2008) 发现当进行认知任务时，被试确认表情的正确率比没有进行这项任务的被试要低，但是这一区别并不显著。

对表情知觉敏感性的测量，各个研究采用的方法有所不同，大致可以分为两类 : 1. 测量辨别情绪面孔的准确性，这类研究会向被试呈现数张全情绪图片，由被试来判断分别属于哪种情绪，情绪图片一般来源于 Ekman 的情绪图片库，而研究会以情绪识别 (emotion recognition) 或者情绪确认正确度 (emotion identification accuracy) 来表示表情知觉敏感性 ;2. 测量辨认情绪面孔的速度，这类研究采用

Morphing 技术将情绪图片从中性图片向全情绪图片变化，由被试选择最先辨别为情绪图片的那一张图片。第二种测量方法提高了生态效度，毕竟在现实生活中出现更多的并不是全情绪面孔。

但在上述第二类实验范式中，其结果同时反映了被试的知觉敏感性和反应偏向。如何将两者区分开来呢？目前尚无研究对此加以区分，本研究将采用信号检测论的方法对这两种指标加以区分。信号检测论能够对在多大程度上反应由于信号加工（比如，感知水平上的敏感性）的不同而不同，还是由于反应系统（决策制定）的不同而不同作出估计。感知敏感性反映了个体如何做出正确的判断并避免错误的判断；反应偏向（比如，深思熟虑或社会赞许性），与可利用的感知信息无关，而是反映一种反应（比如，是）优于另一种反应（比如，不是）的选择偏爱。因此，表情知觉敏感性的测量更适合使用信号检测论的方法。

1.3 自动情绪调节及其研究范式

无意识目标追寻是与当前行为目标关联最为紧密的无意识自我调节形式，在无意识目标追寻的自动化模型理论（Bargh & Morsella，2008; Strack & Deutsch，2004）的基础上，Mauss 等（2007a）提出自动情绪调节（automatic emotion regulation）这个概念，将它定义为无需意识决定、注意参与和有意控制的对情绪各方面进行目标驱动改变，或说自动情绪调节是建立在对调节情绪这个目标的自动追寻基础上来改变情绪轨迹的（樊召锋，俞国良，2009）。自动情绪调节的产生需要以下三个条件：一是感觉被登记；二是被登记的感觉激活了一个图式、概念、目标或脚本（"不要表达这个情绪"或是"不要注意这个情境"）；三是情绪反应的各个方面被改变（Mauss，Bunge，& Gross，2008）。

Mauss 等人（2007）认为自动情绪调节和情绪调节一样可按照加工过程的不同分为不同种类，他们在 Gross 的情绪调节的过程模型的基础上提出了自动情绪调节整合模型，将自动情绪调节分为先行聚焦自动情绪调节（antecedent-focused AE R）和反应聚焦自动情绪调节（response-focused AE R）。反应聚

焦 AER 在情绪反应发生之后被启动，引发对情绪表达的自动调节努力。先行聚焦 AER 则发生于情绪反应发生之前，主要涉及情境选择、注意分配和认知评估转换等方面的努力。Mauss 等人认为反应聚焦 AER 与弗洛伊德提出的防御、抑制及自发抑制相关，所导致的行为调节难以减少消极体验，并会产生适应不良的生理反应；而先行聚焦 AER 通常会带来较少的消极情绪体验和适应性良好的生理反应。

实验范式方面，从 Gross 等人（2007）年提出自动情绪调节至今，大部分研究集中于两个方面，一是自动情绪调节的行为与过程，二是个体对情绪调节的内隐态度。第一方面的研究都是建立在自动目标追寻的基础上、通过句子整理任务来实现的。Mauss 等人（2007b）和 Gross(2007) 通过句子整理任务启动表情抑制的情绪调节方式，结果发现：启动组与表达组相比产生较少的愤怒情绪，而没有产生适应不良的生理反应。Williams 等人（2009）发现启动自动认知重评能够达到和有意认知重评相同的效果。第二方面的研究主要采用内隐联想测验（E R—IAT）来区分个体对情绪调节不同的内隐态度的。研究表明：对情绪调节持积极内隐态度的个体能够自动地调节和控制自己的情绪，并且体验到较少的负性情绪（刘俊升，桑标，2009; Mauss，Evers，Wilhelm， & Gross，2006）。

本研究旨在对表情的无意识模拟进行抑制，而已有研究也证明句子整理任务是可以有效启动自动抑制的，因此本研究将采用句子整理任务来启动自动表情抑制，来研究无意识模拟抑制对表情知觉敏感性的影响，这是一种不需占用认知资源的阻止面部表情的实验方式，研究结果可以进一步支持情绪识别的具身模仿机制，具有较大的理论意义。

综上所述，1. 本研究将信号检测论的方法引入表情知觉敏感性实验范式，修正了以往范式无法排除反应偏向的缺点，对表情知觉敏感性进行更准确的测量。2. 使用句子整理任务这一已被证明有效的实验范式，来启动自动表情抑制，研究无意识抑制面部表情对表情知觉敏感性的影响，扩展了继往只有对有意识抑制面部表情对表情知觉敏感性的影响研究的

现状，可以排除认知负荷对结果的影响，对表情识别的模拟机制和情绪的具身模仿论提供进一步证据。

3. 扩展以往研究中被试都是处于中性状态下的现状，对不同情绪状态下，被试无意识抑制面部表情对表情知觉敏感性的影响进行研究，探索在情绪抑制状态下是否仍然存在情绪一致性效应，为情绪一致性效应的机制提供一定的参考。

2　研究方法

2.1　研究目的

探讨不同情绪状态对不同表情情绪知觉敏感性的影响；探讨在不同情绪状态下，自动表达抑制对不同表情情绪知觉敏感性的影响。

2.2　被试

通过论坛和宣传单发放实验邀请卡，在一所重点大学招募本科及研究生共 103 人。年龄 18 ～ 34 岁，平均年龄 22.45 岁；男女比例约为 1∶3；矫正视力正常。

2.3　实验设计

本研究采用 3 × 2 × 2 的混合设计。组间自变量是自动情绪表达的方式 (抑制组、表达组、控制组) 和诱发的情绪状态 (正性、负性)。组内自变量是需要知觉的情绪种类 (正性、负性)。因变量为被试对表情的知觉敏感性。
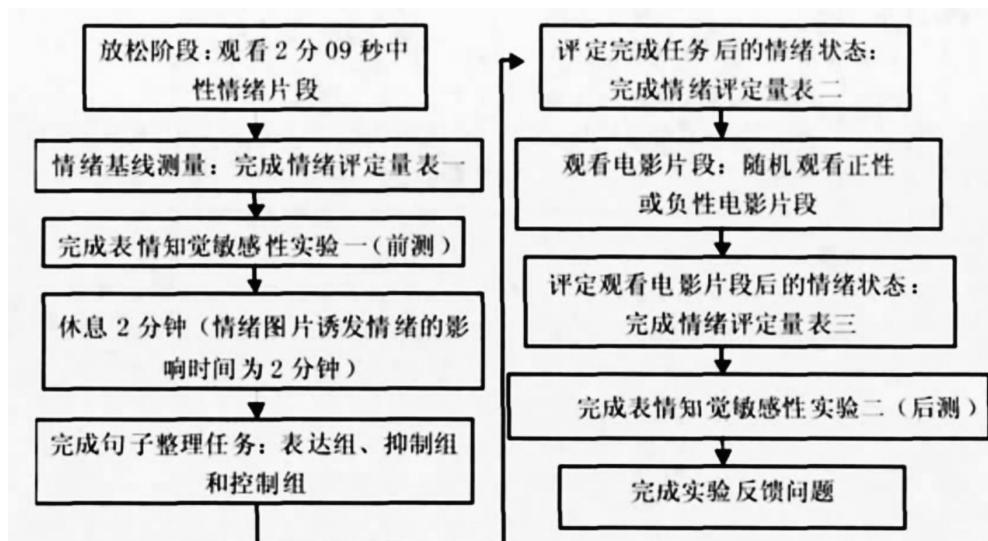
2.4　实验材料

2.4.1　情绪评定量表

采用 Plutchilk(2001) 情绪量表，见黄敏儿，郭德俊 (2001)。

2.4.2　情绪诱发电影片段

《东成西就》，《惊天动地》作为正性、负怀情绪的诱发材料。海洋风光片作为中性情绪片段。

2.4.3　句子整理任务

该任务要求在给出的五个词语中选择其中的四个组成句子。五个词语由四个目标词和一个非目标词组成。不同情绪表达方式组只在目标词不同。预实践结果表明，该任务能成功启动三种对应的自动情绪调节方式。

2.4.4　表情图片

表情图片用于测量被试的表情知觉敏感性。这些表情图片选自《面部表情大全》，11 位模特 (6 位男模特、5 名女模特) 的 11 组正面典型表情，分别为平静—快乐、平静—悲伤，共 22 张图片。

使用 Photoshopcs3 将每组快乐或悲伤图片的面部黏贴到标准背景上。调整图片大小为 12cm ×10cm。

使用 Morph 软件制造从平静向快乐或悲伤变化的 100 张表情图片。从每组 100 张图片中选择第 50、60、70、80、90、100 张图片作为表情图片刺激，将第 50 张定为标准刺激，第 60、70、80、90、

图 1　实验程序

100 张图片定为测试刺激 (Donovan， 1997， 2007a，2007b) 。

2.5 实验程序

被试随机分配到表达组、表达抑制组和控制组，同时进行实验。他们被告知实验的目的是进行表情识别，具体流程见图 1。

在表情知觉敏感性实验中，表情图片通过 Eprime 程序在电脑屏幕上呈现，图片后的屏幕背景设定为黑屏。每次实验，首先呈现 500ms 的蓝屏，然后呈现 2000ms 的标准刺激，紧接着 1000ms 的黑屏，最后是 2000ms 的测试刺激，信噪比为 2: 1。被试通过按键反应判断测试刺激与标准刺激是否相同，如果在 2000ms 内被试未作出判断反应，程序会自动进入下一个实验循环。实验开始前，给予被试 8 次练习机会来熟悉实验规则和按键反应，并对被试的反应正确与否进行反馈。正式实验阶段没有反馈。

3 结果分析

剔除句子整理任务正确率低于 . 5 的被试数据 (4 人)，剔除句子整理任务与情绪调节相关的被试数据 (3 人) 被删除，剔除在观看电影片段期间尝试过调节自己情绪的被试数据 (4 人)，最终的有效数据为 92 人，其中表达组 30 人，表达抑制组 31 人，控制组 31 人。虽然有 10 名被试在表情知觉的过程中意识到自己对表情的模拟，但是并没有人尝试抑制这些模拟，排除了有意抑制的影响，所以这 10 名被试的数据依然有效可用。

用 SPSS 对数据进行分析。情绪知觉敏感性用信号检测论中的 d 表示，d 值越大表示敏感度越高。

3.1 正负性情绪状态对正负性情绪知觉敏感性的影响

对控制组数据进行 ANCOVA 协方差分析，结果表明：在负性情绪状态下被试对负性表情更敏感，在正性情绪状态下被试对正、负性表情的敏感性没有显著差异 (见图 2)。具体如下：

单独分析控制组被试数据，来探明不同情绪状
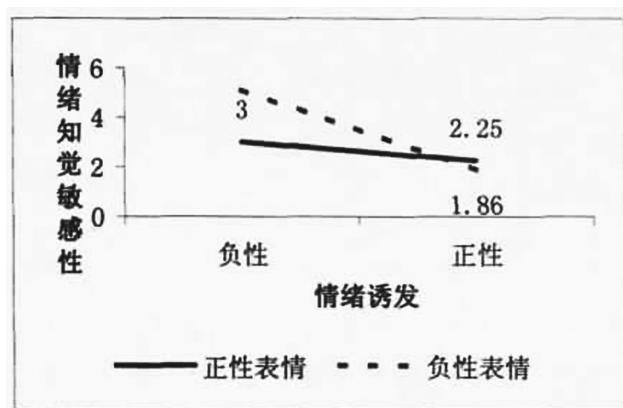
图 2 正负情绪状态对正负情绪知觉敏感性的影响

态对情绪知觉敏感性的影响。将前测的表情知觉敏感性作为协变量，不同情绪状态 (正性、负性) 作为组间变量，不同表情 (正性、负性) 作为组内变量，使用 ANCOVA 协方差分析对被试数据进行分析。回归系数同构型检验结果如下 : 情绪状态 × 表情 × 前测敏感性，$F=0.690$，$p > .05$；情绪状态 × 前测敏感性，$F=0.659$，$p > .05$；表情 × 前测敏感性，$F=0.143$，$p > .05$。结果显示数据符合共变量回归系数同构型假定，能够进行共变量分析，并且前测敏感性数据并不影响正负性情绪状态、正负性表情对表情知觉敏感性的影响。

控制前测敏感性后，情绪状态和正负性表情对表情知觉敏感性的交互作用显著，$F=4.782$，$p < .05$，$\eta^2=0.053$。分析简单效应发现，负性情绪状态下的被试对负性表情更敏感，$F=12.29$，$p < .01$；正性情绪状态下被试对正、负性表情的敏感性差异不显著，$F=0.15$，$p > .05$。

3.2 在不同情绪状态下自动表达抑制对正负性表情知觉敏感性的影响

用 ANCOVA 协方差分析方法分析情绪状态、自动情绪表达方式、正负性表情对表情知觉敏感性的作用，结果显示 : 在负性 (悲伤) 情绪状态下，表达组、控制组的被试对负性表情的敏感性高于正性表情，而抑制组对正负表情的知觉敏感性没有显著差异 (见图 3)。在正性 (快乐) 情绪状态下，三组表达方式的被试对正负表情的知觉敏感性均无显著差异 (见图 4)。
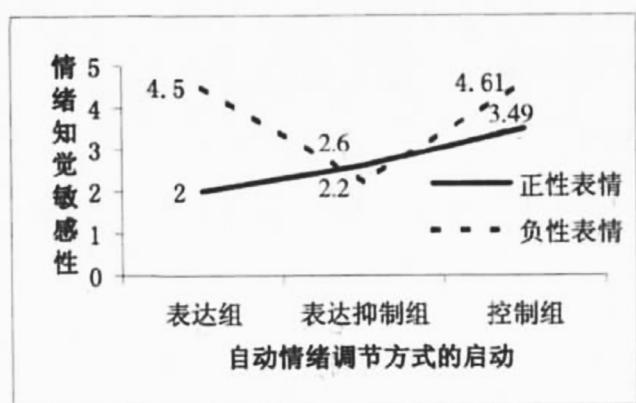
图 3　负性情绪状态下不同启动方式对正负情绪知觉敏感性的影响
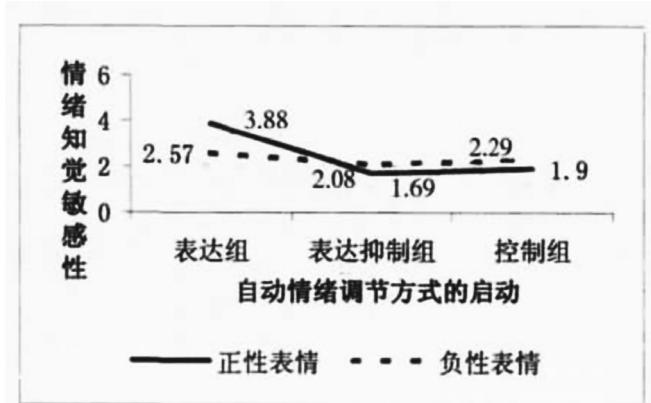
图 4　正性情绪状态下不同启动方式对正负情绪知觉敏感性的影响

将前测的表情知觉敏感性作为协变量，不同情绪状态（正性、负性）、启动（表达、抑制、控制）作为组间变量，不同表情（正性、负性）作为组内变量，使用 ANCOVA 分析对被试数据进行分析。回归系数同构型检验结果如下：情绪状态 × 表情 × 启动 × 前测敏感性，$F=0.520$，$p > .05$；情绪状态 × 启动 × 前测敏感性，$F=0.081$，$p > .05$；情绪状态 × 表情 × 前测敏感性，$F=0.796$，$p > .05$；表情 × 启动 × 前测敏感性，$F=0.691$，$p > .05$。结果显示数据符合共变量回归系数同构型假定，能够进行共变量分析，并且前测敏感性数据并不影响正负性情绪状态、正负性表情对表情知觉敏感性的影响。

控制前测敏感性后，情绪状态、启动、正负性表情对表情知觉敏感性的三向交互作用显著，

$F=4.948$，$p < .05$，$\eta^2=0.095$。

运用 MANOVA 进行进一步的简单交互作用分析。发现在正性（快乐）情绪状态下，自动情绪表达方式的启动并不会影响对不同表情的知觉敏感性，$F=0.16$，$p > .05$；在负性（悲伤）情绪状态下，自动情绪表达方式的启动会对不同表情的知觉敏感性产生影响，$F=6.24$，$p < .05$。分析正性情绪状态下启动和表情变量的简单效应，结果发现：正性情绪状态下，被试对正性表情和负性表情的敏感性没有差异，$F=0.25$，$p > .05$；而与控制组、抑制组被试相比，表达组的被试对表情更敏感，$F=3.45$，$p < .05$，无论是正性表情还是负性表情。

分析负性情绪状态下表情变量在不同启动状态下的简单简单效应，结果发现：表达组、控制组的被试对负性表情更敏感，$F=13.23$，$p < .001$；$F=12.29$，$p < .01$。抑制组被试对不同表情知觉敏感性的差异不显著，$F=0.20$，$p > .05$。

4　讨论

4.1　自动情绪抑制在悲伤和快乐情绪状态下的作用

Mauss 等（2007）研究自动情绪调节对愤怒的调节作用，发现自动抑制组的愤怒情绪显著低于表达组，且这种情绪调节不导致不良心血管反应。该研究未对其它情绪的自动调节进行研究，同时也缺乏正常对照组。本研究发现对悲伤和快乐的自动情绪调节也有类似的结果，看悲伤影片后，抑制组的各种负性情绪体验（悲伤、恐惧、痛苦、紧张）显著低于表达组，控制组仅悲伤情绪体验显著低于表达组，抑制组和控制组平静体验显著高于表达组；看快乐影片后，抑制组的正性情绪体验（快乐、感兴趣）显著低于控制组和表达组，且其平静程度也显著高于表达组和控制组。

4.2 不同情绪状态对不同表情表情知觉敏感性的影响

研究结果发现，情绪状态和正负性表情对表情

知觉敏感性的交互作用显著, 分析简单效应发现, 负性情绪状态下的被试对负性表情更敏感且差异显著, 正性情绪状态下被试对正性表情更敏感但差异不显著。即在负性情绪状态下存在情绪一致性效应, 而在正性情绪下则未发现情绪一致性效应。

根据表情的具身模仿理论, 人们是通过模拟对方的面部表情特征使自己产生适当的情绪体验从而确认对方的情绪状态。由此推测, 当自身情绪状态与对方表情状态相一致时, 个体能够更敏感的识别表情特征、更迅速的觉知对方的情绪状态。Niedenthal 等人 (2001) 的研究发现个体能够更有效的加工与自己所处情绪一致正性或负性的刺激, 但该研究所使用的实验程序不能区分开个体的反应倾向和表情知觉敏感性, 当排除反应倾向后, 正性情绪状态下被试虽然对正性表情更敏感, 但是与对负性表情的知觉相比, 这种差异并不显著。还有一个原因可能是人们对负性表情的敏感性本身就比正性表情高, 这是具有社会适应性意义的。负性表情对人们意味着威胁、危险、坏事的发生等等, 对负性表情更敏感能够帮助人们更快的接收到威胁、危险的信息。造成这种结果的原因可能是负性电影片段诱发情绪的持续时间较长, 而正性电影片段诱发正性情绪的持续时间相对较短 (李芳, 2008) 。

4.3 不同情绪状态下自动表达抑制对表情知觉敏感性的影响

为了更好的证实情绪知觉的模拟说, 本研究加入了自动表达抑制启动变量, 自动表达抑制的启动意味着抑制情绪的表达, 主要是面部表情的表达。因此在正、负性情绪状态下启动自动表达抑制会减少相应的正、负情绪的面部表情表达, 从而会降低与情绪状态相一致的表情知觉敏感性; 而启动情绪表达则相当于鼓励相应的正、负情绪的面部表情表达, 从而使与情绪状态相一致的表情更容易被知觉到。

研究结果部分支持了具身模拟说, 不论是正性情绪状态下还是负性情绪状态下, 表达组被试总比

抑制组被试对表情更敏感。按照模拟说推断, 表达组、控制组还是会出现对与情绪状态相一致的表情较敏感的现象。但是, Jackson 等人 (2000) 的研究发现: 对负性情绪的抑制会导致较少的眨眼和皱眉反应, 而鼓励对负性情绪的表达则会导致较频繁的眨眼和皱眉反应。由此推断, 启动自动表达抑制以后, 面部表情被抑制, 对表情图片的模拟也被抑制, 因此抑制组的被试对正性、负性表情的知觉敏感性都会较低。负性情绪状态下的实验结果支持了这个推断。至于正性情绪状态下为什么没有出现这类结果, 原因可能还是: 一, 正性情绪状态消退得较快, 对知觉敏感性的影响较小; 二, 人们对负性表情本来就比较敏感。

4.4 研究局限与展望

研究采用电影片段诱发的情绪在表情知觉敏感性的程序过程中会消退, 可能会影响研究结果。其次, 本研究对表情图片的要求非常高, 需要同一个人正面的平静和快乐 / 悲伤表情。对此目前尚无成熟图片资料, 研究只能从《面部表情大全》一书选取表情, 对研究结果可能也存在一定影响。

另外, 自动表达抑制的启动还可能受到被试人格等其他因素的影响。将来研究可以对这个影响因素进行控制。

再者, 研究缺少在平静状态下自动表达抑制和表达的启动操作作为控制组。

5 结论

（1）在负性情绪状态下, 表情知觉敏感性存在情绪一致性效应, 即被试对负性表情更敏感; 在正性情绪状态下, 表情知觉敏感性不存在情绪一致性效应, 即被试对正、负性表情的敏感性没有显著差异。

（2）在正性 (快乐) 情绪状态下, 三组表达方式的被试对正负表情的知觉敏感性均无显著差异。在负性 (悲伤) 情绪状态下, 表达组、控制组的被试对负性表情的知觉敏感性高于正性表情, 而抑制

组对正负表情的知觉敏感性没有显著差异。

参考文献

樊召锋, 俞国良. (2009). 自动情绪调节: 基于社会文化与神经科学的考量. *心理科学进展, 17*, 722–729.

黄敏儿, 郭德俊. (2001). 大学生情绪调节方式与抑郁的研究. *中国心理卫生杂志, 15*, 438–441.

李芳, 朱昭红, 白学军. (2008). 高兴和悲伤电影片段诱发情绪的有效性和时间进程. *心理与行为研究, 7*, 32–38.

刘俊升, 桑标. (2009). 情绪调节内隐态度对个体情绪调节的影响. *心理科学, 32*, 571–574.

刘亚, 王振宏, 孔风. (2011). 情绪具身观: 情绪研究的新视角. *心理科学进展, 19* (1), 50–59

Bailey,P. E. ,& Henry,J. D. (2009) . Subconscious facial Bailey,P. E. ,& Henry,J. D. (2009) . Subconscious facial expression mimicry is preserved in older adulthood. *Psychology an Aging, 24 (4)* , 995–1000.

Bargh,J.A.,& Morsella,E. (2008) . The unconscious mind. Perspectives on psychological science,3 (1) ,73–79. Borrill, J. A., Rosen, B. K., & Summerfield, A. B. (1987) . The influence of alcohol on judgment of facial expression of emotion. *British Journal of Medical Psychology, 60,* 71–77.

Corcoran, K. M., Woody, S. R., & Tolin, D. F. (2008). Recognition of facial expressions in obsessive complusive disorder. *Journal of Anxiety Disorders, 22* (1) , 56–66.

Coupland, N. J., Singh, A. J., Sustrik, R. A., Ting, P. , & Blair, R. J. (2003). Effects of diazepam on facial emotion recognition. *Journal of Psychiatry & Neuroscience, 28*(6) , 452–463.

Dimberg, U. (1997). Facial reactions: Rapidly evoked emotional responses. *Journal of Psychophysiology, 11,* 115–123.

Donovan, W., Lewis, A., Leavitt, R. O. & Walsh. (1997). Cognitive set and coping strategy affect mothers' sensitivity to infant cries: A signal detection approach. *Child Development, 68*(5), 760–772.

Donovan, W., Nicole, T., & Lewis, L. (2007a). Maternal self–efficacy, knowledge of infant development, sensory sensitivity,and maternal response during interaction. *Developmental Psychology, 43*(4) ,865–876.

Donovan, W., Lewis, L., Nicole, T., & Jennifer, B. (2007b). Maternal sensory sensitivity, mother–infant 9–month interaction, infant attachment status: Predictors of mother–toddler interaction at 24 months. *Infant Behavior & Development 30,* 336– 352.

Ekman, P. (1973). *Cross cultural studies of emotion.* In P. Ekman(Ed.), Darwin and facial Expression: A century of research in review (pp. 169–222) . New York: Academic Press.

Gallese, V. C. (2005). Embodied simnlation: From neurons to phenomenal experience. *Pehomenology and Cognitive Science, 4,* 23–48.

Gross, J. J., & Thompson, R. (2007) . *Emotion regulation: Conceptual foundations.* In J. J. Gross (Ed.) . Handbook of Emotion R egualtion (pp. 3–24) . New York Guilford Press.

Hatfield, E., Cacioppo, J. L. & Rapson, R. L. (1993). Emotional contagion. *Current Directions in Psychological Sciences, 2,* 96–99.

Hoffman, M. L. (2002). How automatic and representational is empathy, and why. *Behavioral and Brain Sciences, 25,* 38–39.

Jackson, D. C., Malmstadt, J. R., Larson, C. L., & Davidson, R. J. (2000). Suppressioin and enhancement of emotional responses to unpleasant pictures. *Psychophysiology, 37,* 515–522.

Lishner, D. A., Cooter, A. B., & Zald, D. H. (2008). Rapid emotional contagion and expressive congruence under strong test conditions. *Journal of Nonverbal Behavior,*

32, 225–239.

Martin, R. A., Berry, G. E., Dobranski, T., & Horne, M. (1996) . Emotion perception threshold: Indibidual differences in emotional sensitivity. *Journal of Research in Personality, 30* (2), 290–305.

Mauss, I. B., Evers, C., Wilhelm, F.H., & Gross, J. J. (2006) How to bite your tongue without blowing your top: Implicit evaluation of emotion regulation predicts affective responding to anger provocation. *Personality and Social Psychology Bulletin, 32,* 589–602.

Mauss, I. B., Bunge, S. A., & Gross, J.J. (2007a). Automatic emotion regulation. *Social and Personality Psychology Compass, 1,* 146–167.

Mauss, I. B., Cook, C. L., & Gross, J. J. (2007b) . Automatic emotion regulation during anger provocation. *Journal of Experimental Social Psychology, 43,* 698–711.

Mauss, I. B., Bunge, S.A., & Gross, J. J. (2008) . Culture and automatic emotion regulation. In S. Ismer, S. Jung, S. Kronast, C.van Scheve,& M. Vanderkerckhove (Eds.), Regulating Emotions: Culture,Social Necessity and Biological Inheritance (pp. 39–60) . London: Blackwell Publishiing.

Mayer, J. D., & Salovery, P. (1997). *What id emotional intelligence?*In P. Salovey & D. Sluyter (Eds.), Emotional development and emotional intelligence: implications for educators (pp. 3–31). New York: Basic Books.

Montagne, B., Kessels, R. P. C., Dehaan, E. H. F., & Oerrett, D.I. (2007). The emotion recognition task: A paradigm to measure the perception of facial emotional expressions at different intensities. *Perceptual and Motor Skills, 104* (2) , 589–598.

Niedenthal, P. M., Halberstadt, J. B., Margolin, J., & Innes–Ker, A. H. (2000). Emotional state and the detection of change infacial expression of emotion.

European Journal of Social Psychology, 30(2), 211–222.

Niedenthal, P. M., Brauer, M., Halberstadt, J.B., & Innes–Ker, A. S. (2001). When did her smile drop? Facial mimicry and the influences of emotional state on the detection of change in emotional expression.*Cognition and Emotion, 15*(6), 853–864.

Niedenthal, P. M., Mermillod, M., Maringer, M., & Hess, U.(2010) . *Behavioral and Brain Sciences, 33,* 417–433

Orgeta, V., & Phillips, L. H.(2008). Effects of age and emotional intensity on the recognition of facial emotion. *Experimental Aging Research, 34* (1), 63 –97.

Patterson, M. L., & Strockbridge, E. (1998). Effects of cognitive demand and judgment strategy on person perception accuracy. *Journal of Nonverbal Behavior, 22,* 253–263.

Rozin, P., Taylor, C., Ross, L., Bennett, G., & Hejmadi, A. (2005). General and specific abilities to recognize negative emotions,especially disgust,as portrayed in the face and thhe body. *Cognition & Emotion, 19* (3) , 397–412.

Schneider, K.G. (2008). *Expressive control and emotion perception: The impact of expressive suppression and mimicry on sensitivity to facial expressions of emotion.* Department of Psychology and Neuroscience. Duke University.

Strack, F., & Deutsch. R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8,* 220 –247.

Tracy, J. L., & Robins, R. W. (2008). The automaticity of emotion recognition. *Emotion, 8* (1) , 81–95.

Williams, L. E., Bargh, J. A., Nocera, C. C., & Gray, J. .(2009). The unconscious regulation of emotion: Nonconscious reappraisal goals modulate emtional reactivity. *Emotion, 6,* 847–854.

The effects of automatic expressive suppression on the sensitivity of emotion perception in different emotion states

Yang Jing[1,2,3], Tao Rong[1,2], Chen Ruijuan[1,2]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior(CCNU) ， Ministry of Education， Wuhan， 430079)

([2] School of Psychology， Key Laboratory of Human Development and Mental Health of Hubei Province， Central China Normal University，

Wuhan， 430079)

([3] Mental Health Education Center， Wenhua College， Huazhong University of Science and Technology， Wuhan， 430079)

Abstract　Emotion perception is the ability to perceiveor identify an emotion， for example， to identify what do people feel or what sentiments are contained in the art， music and so on． During interpersonal communications， people experience their emotions every now and then． While receiving emotional information from other people， emotions are always in collision． In this process， the accuracy， timeliness and effectiveness of emotion perception， especially of the facial expression， is essential to the smoothness of communication． Emotional state and cognitive load are two important impact factors of emotion perception． In addition， according to emotion contagion theory， mimicry， feedback and contagion are three stages in the process of emotion recognition． Studies have shown that， to stop simulating facial expressions consciously will affect the emotion perception sensitivity． However， suppressing expression mimicry unconsciously does not occupy cognitive resources． The questions that may be asked are: does the cognitive load or the suppression itself consciouslyaffect emotion perception sensitivity? Or is emotion perception not on the level of consciousness， as the embodied simulation theory says? At present， there is no research to explore these questions． This study tries to explore whether emotion perception sensitivity is affected by automatic regulation of different emotions. This study uses signal detection theory to test the emotion perception sensitivity． We select words related to emotion — suppression and emotion — expression to design the sentence unscrambling task， in which the level of automatic emotion regulation is manipulated． Using this task，we try to explore the impact of automatic expressive suppression on the perception sensitivity of different facial emotions under positive or negative emotion． 81 college students participate voluntarily in preliminary experiment and 103 in formal experiment.Results show that:（1）Subjects in automatic expressive suppression group report less emotional experience than those in the emotion expression group． This is consistent with the socio — cultural background of China in that people tend to inhibit their expression of negative emotions rather than positive ones．（2）Emotion perception sensitivity partially shows the effects of emotional coherence． In negative emotional state， people are significantly more sensitive to negative emotional expression． And in positive emotional state， people tend to be more sensitive to positive emotional expression， although the difference is not significant．（3）Automatic expressive

suppression affects the effects of emotional coherence in emotion perception sensitivity. Subjects in automatic expressive suppression groups are sensitive to neitherpositive nor negative facial expressions.

Keywords automatic expressive suppression; emotion perception sensitivity; negative or positive emotion

自伤青少年的冲动性 *

于丽霞 [1,2]，凌　霄 [2,3]，江光荣 [1,2]

([1] 青少年网络心理与行为教育部重点实验室)

([2] 湖北省人的发展与心理健康重点实验室 , 华中师范大学心理学院 , 武汉 430079)

([3] 武汉外国语学校 , 武汉 430022)

摘　要　以自我报告、行为学和脑电为指标 , 检验自伤青少年的冲动性。研究 1, 对 820 名普通中学生和 72 名工读生进行问卷调查 , 探讨自伤行为与情绪调节困难、冲动性的关系。结果表明 , 冲动性能够预测自伤行为 , 且预测效应量大于情绪调节困难。研究 2, 采用 Go/Nogo 范式的 ERPs 实验 , 检验自伤组与对照组冲动控制的行为学与脑电差异。结果表明 , 自伤组 Nogo 正确反应的 N2 波幅显著高于对照组 , N2 潜伏期在部分电极点处高于对照组。脑电地形图显示两者的脑电差异主要体现在前额叶区。结论：自伤青少年的冲动性高于同龄普通青少年。

关键词　非自杀性自伤 ; 自伤 ; 冲动性 ; ERPs; N2

分类号　R395;B844

1　问题提出

自我伤害 (下文简称自伤) 行为指在没有明确自杀意图的情况下 , 个体故意、重复地改变或伤害自己的身体组织 , 如用利器割伤 / 划伤、打火机烧伤、以头撞墙等 ; 这种行为不为社会所认可 , 且不具致死性或致死性较低 (Gratz, 2001)。自伤行为与很多心理问题或精神障碍相关 , 如边缘型人格障碍 (BPD)、摄食障碍、抑郁、焦虑、药物滥用、创伤后应激障碍、分离障碍 (Gratz & Gunderson, 2006;Haw, Hawton, Houston, & Townsend, 2001; Sho etal., 2009; Svirko & Hawton, 2007; Zlotnick, Mattia,& Zimmerman, 1999) 等。

虽然自伤的目的不在自杀 , 但自伤当事人的自杀风险远高于普通人群 (Hawton,Zahl, & Weatherall, 2003; Sinclair, Hawton, & Gray,2010), 因此危险性极大。自伤的相关研究正受到越来越多的研究者和临床工作者的重视 , "为什么要自伤"成为研究者最关切的问题。围绕这个问题 , 目前主要在两方面展开研究 , 一是自伤的功能 , 一是自伤的影响因素。

自伤具有多种功能 , 包括情绪管理、对抗分离感、对抗自杀、人际影响、自我惩罚等 (Klonsky, 2007;Messer & Fremouw, 2008; Suyemoto, 1998), 其中首要功能是情绪管理 (Klonsky, 2007; 郑莺 , 2006)。情绪管理指个体知觉、理解、接受自己的情绪体验以及

灵活地运用策略做出合适的行为，以上任何一种能力的缺失均称为情绪管理障碍 (Gratz, 2007)。情绪管理障碍是自伤者的核心问题，具体表现为情绪感受的脆弱性 (情绪易唤起、强度高、难平复) 和情绪管理能力的缺乏 (Chapman, Gratz, & Brown, 2006;Gratz, 2007; 冯玉 , 2008; 于丽霞 , 江光荣 , 吴才智 ,2011)。研究发现，当事人在自伤前通常会经历强烈的、难以自控的负性情绪体验，比如愤怒、悲伤、受挫、内疚等，自伤能够帮助当事人在短时间内迅速释放情绪和恢复平静 (Klonsky, 2009)。因此自伤是一种适应不良的应对方式，目的是帮助自伤者调节强烈的负性体验。Chapman (2006) 将情绪管理障碍具体分为情绪表达不能、情绪调节困难和高情绪强度三个成分。冯玉 (2008) 以青少年为被试，检验了三个成分对自伤的预测作用，结果情绪表达不能和情绪调节困难进入了回归方程，共解释了自伤行为总变异的 18.5%，其中情绪调节困难的解释力最高，解释了总变异的 12.1%。冯玉的研究同时发现情绪调节困难在早期创伤经验对自伤的影响中起中介作用。另一项以大学生为被试的研究表明，情绪调节困难能将 64% 的自伤者与非自伤者区分开，准确率达 80% (Gratz & Chapman, 2007)。由此可见情绪调节困难是自伤的重要影响因素，本文将选取情绪调节困难作为自伤的预测因素之一。

自伤的影响因素既有来自后天的社会环境因素，如早期创伤性经验 (Gratz & Chapman, 2007;Klonsky & Moyer, 2008; Yates, Carlson, & Egeland,2008)，也有来自先天的人格 或气质特点，如冲动性。不少研究发现，自伤者从考虑到正式实施自伤的时间间隔通常不到 5 分钟 (Nock & Prinstein, 2005; 郑莺 , 2006)，且常同时伴有其他冲动性行为，如暴食、酒精 / 物质滥用、病理性赌博等。并且，根据 DSM-IV 的诊断标准，自伤满足冲动 - 控制障碍的三个核心特点 (Dell' Osso, Altamura, Allen, Marazziti,& Hollander, 2006)：(1) 无法控制伤害自己或他人的冲动性行为；(2) 行为发生之前情绪唤起或紧张感不断加剧；(3) 行为之后有愉悦感或释放感。因此冲动性被认为是影响自伤的核心因素之一 (Herpertz,Sass, & Favazza, 1997)，甚至有研究者

认为自伤应属于一种冲动 - 控制障碍 (转引自 Glenn & Klonsky,2010)。Miller 等 (2003) 对有自杀未遂和自伤经历的 BPD 患者的研究发现，控制 BPD 的效应之后，冲动性仍然可以将自伤者与自杀者区分开，而控制冲动性的效应后，BPD 的区分效应则不显著。这些证据都表明冲动性是自伤的重要风险因素。

冲动性是一种人格特质，有较强的生物学基础，主要表现为个体对于来自内部或外部的刺激产生迅速、无计划的反应倾向，而不考虑对自己或他人造成的负面后果 (Moeller, Barratt, Dougherty,Schmitz, & Swann, 2001)。冲动性与冲动 - 控制障碍、边缘型人格障碍 (Links, Heslegrave, & Reekum,1999)、注意缺陷多动障碍 (Dougherty et al., 2004)、摄食障碍 (Fischer, Smith, & Anderson, 2003)、攻击 (Barratt, 1994; Berkowitz, 1974)、物质滥用 (Jentsch& Taylor, 1999)、自杀 (Baca-Garcia et al., 2005;Swann et al., 2005) 等多种心理病理行为有关联，是研究者给予较多关注的心理病理行为的影响因素。但冲动性的概念和结构多年来仍然没有获得统一的认识。学界都认同冲动性并非单一结构，而是包含有多种独立成分的复合结构 (Evenden, 1999)，但具体包含哪些成分研究者们则莫衷一是 (Whiteside& Lynam, 2001)。由于测量工具依赖于特定的理论框架 (Evenden, 1999)，理论上界定的不统一也导致了冲动性测量工具的多样化 (Gerbing, Ahadi, &Patton, 1987)。当前发展出的冲动性测量方法中，仅自评工具就达数十种，其中最为普遍使用的是第 11 版 Barratt 冲动性量表 (BIS-11, Stanford et al.,2009)，另有多个行为学实验范式 (如行为抑制 / 冲动控制、冒险决策、延迟折扣等)。近年来神经生物学实验 (Carver, Johnson, & Joormann, 2008; Cools,Roberts, & Robbins, 2008) 和脑电实验 (Ramautar,Kok, & Ridderinkhof, 2004; Ruchsow et al., 2008a) 也成为检验冲动性的重要方法。在测量过程中，由于自我报告受个体主观因素的影响，研究者通常建议将自我报告与行为学、神经生物学和 / 或神经心理学等客观指标相结合进行综合评定 (周亮 , 何晓燕 , 肖水源 , 2006)。

虽然冲动性是自伤的重要影响因素之一，但对自

伤者冲动性研究的结论并不一致。首先是不同研究结论之间存在冲突。如，Evan 等 (1996) 对住院自伤当事人冲动性的研究发现，重复自伤者的冲动性高于一次自伤者，后者又高于普通人群。但 Herpertz 等 (1997) 研究发现自伤组与对照组只在冲动性的某些方面 (如计划维度) 存在差异。也有研究发现冲动性与不同自伤者的自伤程度存在相关，但自伤组与对照组的冲动性水平差异不显著 (Simeonet al., 1992)。Hawton 等 (2002) 以 6020 名 15~16 岁的中学生为样本，发现仅女性被试自伤组的冲动性水平高于对照组，而男性被试的两组差异不显著。其次是不同测量方法的结果之间存在冲突。自伤组自我报告的冲动性水平明显高于正常组 (Glenn &Klonsky, 2010; Herpertz et al., 1997; Janis & Nock,2009; Mc Closkey, Look, Chen, Pajoumand, &Berman, 2012)，但行为学实验结果则显示两组在实验任务中的表现均无显著差异 (Glenn & Klonsky,2010; Janis & Nock, 2009; Mc Closkey et al., 2012)。Glenn 和 Klonsky (2010) 采用 UPPS 冲动性量表 (Whiteside & Lynam, 2001) 和停止信号任务 (stopsignaltask) 探讨冲动性与自伤的关系，结果表明自伤组的 UPPS 测量结果显著高于对照组，而两组在停止信号任务中表现一致。Janis 和 Nock (2009) 分别以青少年和成人两个样本为被试，发现两个样本自伤组自我报告的冲动性水平均高于对照组，但两组在不同实验范式 (行为去抑制实验、冒险决策实验、延迟折扣) 下的任务表现无差异。当前对自伤者冲动性研究结论的不一致表明，了解冲动性与自伤的关系还需要更多的研究证据，多个指标比单一指标的测量结果更可靠。

除了传统的自评和行为学指标，事件相关电位 (ERPs) 研究是当前被较多用于检验冲动性的方法。ERPs 实验通常通过反应抑制范式的 Go/Nogo 任务刺激诱发脑部额区 N2 负波来检验冲动性水平 (Falkenstein, Hoormann, & Hohnsbein, 1999)。Go/Nogo 范式给被试呈现两种刺激，要求被试对一种刺激 (Go) 做反应而对另一刺激 (Nogo) 不做反应。每个刺激出现后 200 ~ 350 ms，在脑部额叶区会诱发出一个 ERPs 的主要成分——N2 负波。由于 Go 刺激的呈现频率

高于 Nogo 刺激，Go 反应成为优势反应，因此抑制对 Nogo 刺激的反应需要更多的神经加工，耗费更多的神经能量。Nogo-N2 因此被认为反映了神经加工的早期反应抑制功能 (Folstein &Van Petten, 2008; Jodo & Kayama, 1992; van Boxtel,van der Molen, Jennings, & Brunia, 2001)，而反应抑制功能损伤的一个重要表现就是冲动性 (Newman,1987)。二者的对应关系是，冲动性较高者的冲动抑制能力较差，其 Nogo 刺激下的反应 (包括正确和错误反应) 诱发的 N2 波幅较之正常组更小、潜伏期更短 (Falkenstein et al., 1999); 在 Nogo 刺激正确反应情况下，冲动性高者由于成功抑制一个冲动反应需要耗费更多的神经能量，其 N2 幅值更高、潜伏期更长。

当前已有多个病理性人群冲动性的 ERPs 研究的 N2 证据，但不同人群的研究结论很不一致。Sunohara 等 (1999) 发现 ADHD 儿童的冲动性水平高于对照组，表现为 Nogo 反应的 N2 潜伏期比正常组更短。Kiehl 等 (2000) 对精神分裂症和精神病患者冲动性的研究发现，病患组在 Go/Nogo 刺激中的 N2 效应 (N2d, 即 Go 和 Nogo 刺激反应的 N2 波幅的差值) 显著小于正常组，表明病患组的冲动控制能力受损。Yang 等 (2009) 对海洛因成瘾者的研究发现，成瘾组与控制组在 Go/Nogo 的任务中的 N2 效应差异 (N2d) 显著，但两组的 Nogo 反应的 N2 幅值差异不显著。Ruchsow 等 (2008a) 对 BPD 冲动性的研究发现，BPD 组与正常组的 Nogo 反应 N2 幅值、潜伏期差异均不显著。Munro 等 (2007) 以犯罪人员为样本，发现犯罪组与正常组的任务反应时和准确率的行为学指标差异显著，但 Nogo 反应的 N2 指标无差异。Ruchsow 等 (2008b) 对正常人群高低冲动组的 ERPs 研究发现两组的 Nogo 反应的 N2 指标不存在差异。Luijten 等 (2011) 对吸烟成瘾者的反应抑制能力研究发现，吸烟组比控制组的反应抑制能力更差，表现为吸烟组的 Nogo 反应的 N2 波幅明显降低。对于自伤冲动性的研究当前多采用问卷或行为学实验的方法，还没有来自神经科学方面的证据。

本文将采用 ERPs 实验的 N2 指标结合自我报告和行为学结果共同检验自伤青少年的冲动性。青少年

是自伤的高发群体，国外流行学调查发现自伤在普通青少年中的发生率约为 14%~56% (Bjarehed& Lundh, 2008; Hilt, Cha, & Nolen- Hoeksema,2008)，国内青少年自伤的发生率在 40% 以上 (冯玉 ,2008; 郑莺 , 2006)，因此本文以青少年为研究对象。情绪管理障碍是自伤者的核心问题，但从情绪管理的过程来看，实施自伤行为的直接原因在于对激发的冲动或行为表达的调节，即冲动控制。因此本文将进行两个研究：研究 1 将比较情绪调节困难和冲动性对自伤行为的预测力；在研究 1 结论的基础上，研究 2 将采用 ERPs 的 N2 指标，并结合行为学指标 (反应时、错误率)，比较自伤青少年和普通青少年冲动性的差异。

2 研究 1

2.1 研究目的

研究 1 的目的是检验情绪调节困难和冲动性与自伤行为的关系。研究假设： (1) 情绪调节困难和自我报告的冲动性水平对自伤行为有显著预测作用 ;(2) 自我报告冲动性水平对自伤的预测作用大于情绪调节困难。

2.2 研究方法与程序

2.2.1 被试

采用整群抽样法，向武汉市某普通中学和某工读学校学生集体发放问卷 920 份，有效收回 892 份，回收率为 96.96%。其中，普通中学学生 820 名，男生 430 名，女生 390 名，平均年龄 13.64 岁 (SD=0.99)；工读学校学生 72 名，男生 53 名，女生 19 名，平均年龄 14.28 岁 (SD=0.95)。892 份问卷中有 9 份多题漏答，予以剔除，实际参与统计分析共 883 人。

2.2.2 测量工具

（1）青少年自我伤害行为问卷

该问卷由郑莺 (2006) 编制、冯玉 (2008) 修订，根据自伤史的频次和对身体的平均伤害程度的乘积来综合评估自伤水平 (下文的自伤水平即指频次与对身体平均伤害程度的乘积)。频次和平均伤害程度分别采用 0~3 的四级 (0 次、1 次、2~4 次、5 次以上) 和 0~4 的五级 (无、轻、中、重、极重) 计分。前期研究发现，该问卷内部一致性信度为 0.85, 并具有理想的区分效度、效标效度和聚合效度 (冯玉 , 2008)。

（2）情绪调节困难量表 (Difficulties in Emotion Regulation Scale, DERS)

原问卷由 Gratz 和 Roemer(2004) 编制，本文采用冯玉 (2008) 翻译修订的中文版。DERS 由 5 个维度，共 31 个项目组成，5 个维度分别为：（1）难以意识到自己的情绪反应；（2）不接纳自己的情绪反应；（3）缺乏有效的情绪调节策略；（4）当体验到消极情绪时，难以控制自己的冲动反应；（5）当体验到消极情绪时，难以进行有预定目标的行为。被试根据五级评分直接报告项目描述是否符合自己，5 个维度共解释总变异的 50.76%。总量表和分量表的内部一致性信度系数均在 0.70 以上。本研究采用 DERS 总分作为预测指标。

（3）Barratt 冲动性量表 (Barratt Impulsiveness Scale, BIS–11)

BIS 由 Barratt 于 1959 年初次编制后便不断得到重新修订，经由 Patton 等 (1995) 修订而成的最新版 BIS–11 是当前使用最为广泛的冲动性自评工具 (Stanford et al., 2009)。BIS–11 共 30 个项目，从三个维度评估冲动性人格：注意冲动性 (attentionalimpulsiveness)、运动冲动性 (motor impulsiveness) 和无计划冲动性 (no–planning impulsiveness)。问卷采用 Likert5 级评分 (不是、极少、有时、经常、总是)，总分在 30~150 分之间变化，得分越高，冲动性越强。来自中国大学生样本和普通社区样本的研究表明，中文版 BIS 的结构与原问卷基本一致 (杨会芹等 , 2007), 总量表和各分量表的内部一致性信度系数在 0.56~0.76 之间，重测信度在 0.67~0.85 之间，条目 – 总分相关有统计学显著性意义，适合于中国群体冲动性水平的研究 (周亮，肖水源，何晓燕，厉洁，刘慧铭，2006)。本文采用由北京心理危机研究与干预中心翻译修订的 BIS–11 中文版问卷 (安静 , 2008), 以总分为预测指标。

由于对同一批被试同时采用了三个问卷共89个条目的测量，有可能存在因社会称许性、被试作答偏向和习惯作答等因素导致的共同方法变异(Spector, 2006)。本文在问卷施测过程中采用了匿名作答以尽量减少共同方法变异的产生，同时采用单一方法潜变量法进行检验(Podsakoff, MacKenzie, Lee, & Podsakoff, 2003; 熊红星，张璟，叶宝娟，郑雪，孙配贞，2012; 周浩，龙立荣，2004)。将共同方法作为单独一个潜变量纳入结构方程模型，分析结果(见表1)表明，当加入方法因子后，模型的卡方量变化显著($\triangle \chi^2 = 1508.87$, $\triangle df=97$, $\triangle \chi^2/\triangle df=15.56$)，说明各变量之间存在一定程度的共同方法变异。由于$\triangle \chi^2$同时受到样本量的影响，在比较两个模型时还应参考其他拟合指标的变化情况(温忠麟，侯杰泰，马什赫伯特，2004)。从表1结果可见，纳入方法因子后模型的拟合指数改善不明显，两个模型均与数据有非常好的拟合，这说明各变量之间不存在严重的共同方法变异问题。

2.2.3　分析与统计

采用SPSS 17.0做描述性分析，用Mplus 6.11做相关分析、logistic回归分析和结构方程模型。

2.3　结果

2.3.1　青少年自伤行为的发生率

分别统计普通中学生与工读学校学生的自伤率，两个群体自伤的基本情况见表2。由描述性分析结果可见，从自伤水平看，普通中学生中从未实施自伤行为的比例为74.39%; 单从自伤频次看，从未实施过自伤的比例为64.02%。工读生中，从自伤水平看，得分为0分的比例为11.11%; 单从自伤频次看，从未实施自伤的比例仅为6.94%。

2.3.2　青少年自伤行为与情绪调节困难、冲动性的相关分析

将自伤频次和自伤水平作为分类变量，将被试分别分为四组(0次、1次、2~4次、5次以上; 0分、1分、2~10分、10分以上)，其中自伤水平为人为分组，方差分析表明，四组被试在情绪调节困难、冲动性总分和各维度得分的组别差异显著，表明分组有区分效应。做自伤与情绪调节困难和冲动性的总分与各维度的相关分析。根据表3和表4的结果可见，自伤频次和自伤水平与情绪调节困难、冲动性总分均呈中度正相关。

2.3.3　自伤行为对情绪调节困难、冲动性的logistic回归分析

表1　模型比较 ($n=883$)

模型	χ^2	df	χ^2/df	RMSEA	90%C.I.	CFI	TLI
原模型	8284.69	4553	1.82	0.030	0.029–0.032	0.87	0.87
纳入方法因子模型	6775.82	4456	1.52	0.024	0.023–0.025	0.92	0.92

表2　样本青少年自伤基本情况

自伤水平	普通中学生 (n=820)		工读生 (n=72)		自伤频次	普通中学生 (n=820)		工读生 (n=72)	
	n	%	n	%		n	%	n	%
0分	610	74.39	8	11.11	0次	525	64.02	5	6.94
1分	51	6.22	3	4.17	1次	63	7.68	4	5.56
2~10分	115	14.02	21	29.17	2~4次	92	11.22	8	11.11
10分以上	44	5.37	40	55.56	5次以上	140	17.07	55	76.39

表3　自伤行为与DERS及各维度的相关分析 ($n=883$)

自伤行为	情绪意识困难	情绪接纳困难	缺乏调节策略	冲动控制困难	目标执行困难	DERS总分
自伤频次	0.17**	0.19**	0.37**	0.34**	0.31**	0.39**
自伤水平	0.16**	0.20**	0.36**	0.36**	0.33**	0.38**

表4　自伤行为与BIS及各维度的相关分析 ($n=883$)

自伤行为	注意冲动性	运动冲动性	无计划冲动性	BIS总分
自伤频次	0.31**	0.43**	0.38**	0.45**
自伤水平	0.32**	0.43**	0.36**	0.44**

将自伤频次和自伤水平作为分类变量，以 WLSMV 法评估情绪调节困难、冲动性总分对自伤频次及自伤水平的预测效应。Logistic 回归结果 (见表 5) 显示，冲动性和情绪调节困难共同解释自伤频次的 22.42%，自伤水平的 21.65%。其中，冲动性的解释力较高，能解释自伤频次总变异的 14.60%，自伤水平总变异的 14.24%；情绪调节困难能解释自伤频次总变异的 7.82%，自伤水平总变异的 7.41%。

2.4 讨论

2.4.1 青少年自伤行为的发生率

本结果发现，以自伤频次和自伤水平为指标，分别有 35.98% 和 25.61% 的普通中学生至少实施过一次以上的自伤行为，均低于国内前期调查结果 (冯玉，2008; 郑莺，2006)。郑莺 (2006) 曾采用分层随机抽样方法，从武汉市七个中心城区按初高中生和年级比例共随机抽取了 13 所中学、共 1283 名普通中学生为样本，以自伤频次为指标的调查结果发现，中学生的自伤率高达 57.4%。冯玉 (2008) 曾以三所普通中学的 340 名中学生为样本、以自伤水平为指标，结果发现

有自伤史的比例为 45.6%。不同调查结果的差异可能与样本的代表性有关。本研究样本来自一所重点初中，取样单一，样本可能不具代表性。

本研究发现工读学校学生的自伤率高达 88.89%，这一结论基本与中西方以往调查结果一致。冯玉 (2008) 以 115 名少年犯为样本结果发现自伤的发生率为 83.5%。Nock 等 (2006) 发现自伤在病理性问题青少年中的发生率超过 80%(Nock &Prinstein, 2004)。由此可见，问题青少年是发生自伤行为的高危群体。

2.4.2 青少年自伤行为与情绪调节困难、冲动性的关系

相关分析和 logistic 回归分析结果表明，青少年的自伤频次、自伤水平与情绪调节困难、冲动性总分均有密切关联，这与前期研究结果一致 (Janis& Nock, 2009; 冯玉，2008)。其中，自我报告的冲动性对自伤行为总变异的解释量 (14.24%) 比情绪调节困难的解释量 (7.41%) 更大。在冯玉 (2008) 前期研究中，情绪调节困难能解释自伤行为 12.1% 的总变异，本研究发现情绪调节困难对自伤的解释力低于自我报告的冲

表 5 自伤行为对情绪调节困难、冲动性人格的 logistic 回归分析 (n=883)

自伤行为	预测变量	Estimate	SE	StdYX Estimate	P	R^2
自伤频次	冲动性	0.02	0.003	0.33	<0.001***	0.224
	情绪调节困难	0.01	0.003	0.20	<0.001***	
自伤水平	冲动性	0.02	0.003	0.33	<0.001***	0.216
	情绪调节困难	0.01	0.003	0.19	<0.001***	

动性水平。这一结论的差异可能源于情绪调节困难与冲动性存在高相关。尽管多重共线性检验表明情绪调节困难与冲动性不存在共线性问题 (容忍度 =0.66)，但两者相关达 0.58，这表明两者存在较高的共变关系，情绪调节困难对自伤的效应量一部分被冲动性的效应量所解释。这说明冲动性是自伤的重要危险因子。

3 研究 2

3.1 研究目的

研究 1 结果发现冲动性是影响自伤的重要因素之一，其对自伤的预测作用超过情绪调节困难。研究 2 采用基于 Go/Nogo 范式的 ERPs 实验，比较检验自

伤青少年冲动性的行为学及脑电特征。研究假设为自伤组冲动性水平高于对照组，表现为：(1) 在行为学指标上，自伤组在 Go/Nogo 实验任务中的按键反应时比对照组更短，或 / 和错误率更高 ;(2) 在脑电指标上，自伤被试成功抑制一个冲动行为应需要耗费更多神经能量，表现为 Nogo 刺激成功抑制 (不按键) 的 N2 (下文统称 Nogo-N2, 即 Nogo 刺激正确反应的 N2) 幅值更大、潜伏期延迟。

3.2 研究程序与方法

3.2.1 被试

被试均来自研究 1。比较检验研究 1 中不同自伤水平被试的情绪调节困难和冲动性水平，方差分析结果发现自伤的组别效应显著，得分在 10 分以上 (即

自伤频次 ≥ 5 次、伤害程度中度以上) 自伤者的情绪调节困难和冲动性的总分及分量表得分均显著高于其他三组 (0 分、1 分、2~10 分, p<0.001)。为增加区分效应, 选取自伤得分在 10 分以上和 0 分者分别为研究 2 的自伤组和对照组。

作者向研究 1 中符合条件的学生发出书面及电话邀请, 对所有受邀被试再次进行自伤问卷及访谈筛查。自伤组问卷筛查标准满足 : (1) 过去曾自伤 5 次以上 ; (2) 自伤程度为中度及以上。对照组问卷筛查标准满足自伤频次为 0。访谈内容为问卷填写的真实性。最终确定自伤组 19 人, 其中男生 10 人, 女生 9 人。对照组 15 人, 其中男生 9 人, 女生 6 人。被试的自伤及自我报告冲动性水平见表 6。所有被试均为右利手, 视力或矫正视力正常, 智力正常。通过对陪同前来的家长或教师及被试本人的访谈, 排除其他精神疾病及

表 6　被试的自伤程度与冲动性水平

变量	自伤组 (*n*=19)	对照组 (*n*=15)
年龄 (*M* ± *SD*)	14.11 ± 0.94	13.88 ± 0.62
自伤频次 (*M* ± *SD*)	17.95 ± 9.25	0
自伤频次 × 程度 (*M* ± *SD*)	29.58 ± 25.23	0
冲动性 (*M* ± *SD*)	33.65 ± 2.89	16.23 ± 1.39

严重躯体疾病、24 小时内服用镇静药物或精神活性药物。被试在正式实验之前签署知情同意书, 实验之后给予适当报酬。

3.2.2　实验刺激材料

用 Stim 软件编制、呈现刺激序列。刺激为单个正立和倒立的正三角图形 (边长 6.5 cm), 以黑色背景白色线条呈现于显示器中央 (亮度 60 cd/m2)。刺激序列包括 300 个刺激, 倒立三角形 200 个 (66.67%), 正立三角形 100 个 (33.33%)。刺激随机呈现, 正立三角连续出现次数小于 3 次。视距 120cm, 垂直视角 2.8°, 水平视角 2.9°。刺激呈现时间 100 ms, 刺激间隔 (ITI) 在 650~750 ms 之间随机 (见图 1)。

3.2.3　ERPs 记录

使用 Neuroscan 系统记录脑电, 采样率 500 Hz, 记录带通 0.05 Hz ~ 100 Hz, 双侧乳突参考, 头皮和电极之间阻抗小于 5kΩ。被试佩戴 10–20 系统 64 导电极帽, 双眼外眦记录水平眼电 (HEOG), 左眉上和眶下记录垂直眼电 (VEOG)。同时记录连续 EEG 与行为学数据 (见图 2)。

3.2.4　程序和任务

被试舒适坐位, 双眼平视屏幕中心。被试被告知这是一个按键游戏, 要求他 / 她在倒立三角形呈现时尽快尽准确地按键 (Go 刺激), 对正立三角不反应 (Nogo 刺激), 游戏时长 3 分钟, 他 / 她在游戏中的表现决定最终获得的报酬额度。这一处理用以激发被试的行为动机, 并引起对行为结果的预期。之后刺激快速呈现 (100 ms), 限速反应 600 ms。其中 Go 刺激出现概率为 66.7%, 按键反应成为优势反应 ;Nogo 刺激出现概率为 33.3%, 不按键为抑制性反应。正式实验前, 16 人的预实验 (大学生 10 人, 初中生 6 人 ; 其中自伤初中生 4 人, 对照组 2 人) 表明 Nogo 刺激错误反应的反应时比 Go 刺激的正确反应时短, Nogo 刺激错误率高, 证明实验任务能够成功诱发被试的冲动反应 (Falkenstein, Hohnsbein,Hoormann, & Blanke, 1991)。

正式实验流程为 : 首先是 50 个刺激的练习环节, 确认被试明白作业任务后, 开始正式实验。按键左右手在被试者中交叉平衡。整个实验持续 45 ~60 min。

3.2.5　ERPs 数据分析和统计

采用 Scan 4.3 离线分析数据。自伤组 4 人, 对照组 2 人 (男女各半) 用以预实验参数调整, 不计入正式实验数据分析。正式实验中, 自伤组 3 人, 对照组 1 人的数据因脑电信噪比低 (帽子参考电极出现故障) 被剔除。因此, 最终进入统计分析的有效被试为 24 人, 自伤组 12 人, 对照组 12 人。

具体数据处理过程为 : 合并行为数据, 相关法去除眼电, 分析时程为刺激前 100 ms (用于基线校正) 至刺激后 600 ms。基线校正后波幅超过 ± 100 μV 的在叠加中被剔除。0.1 Hz ~ 16 Hz 无相移数字滤波后对反应正确的 EEG 进行分类叠加, 得到 Go、Nogo 刺激产生的两类 ERPs 成分。组内变量包括 : 任务类型 (Go/Nogo) 和电极位置 (F3/Fz/F4/FC3/FCz/FC4/C3/Cz/C4) 等。组间变量为自伤 / 非自伤。

使用 SPSS 17.0 进行统计分析。(1) 行为学分析为组别 2 (自伤组 / 对照组) × 任务类型 2 (Go 刺激正确反应 /Nogo 刺激错误反应) 重复测量的方差分
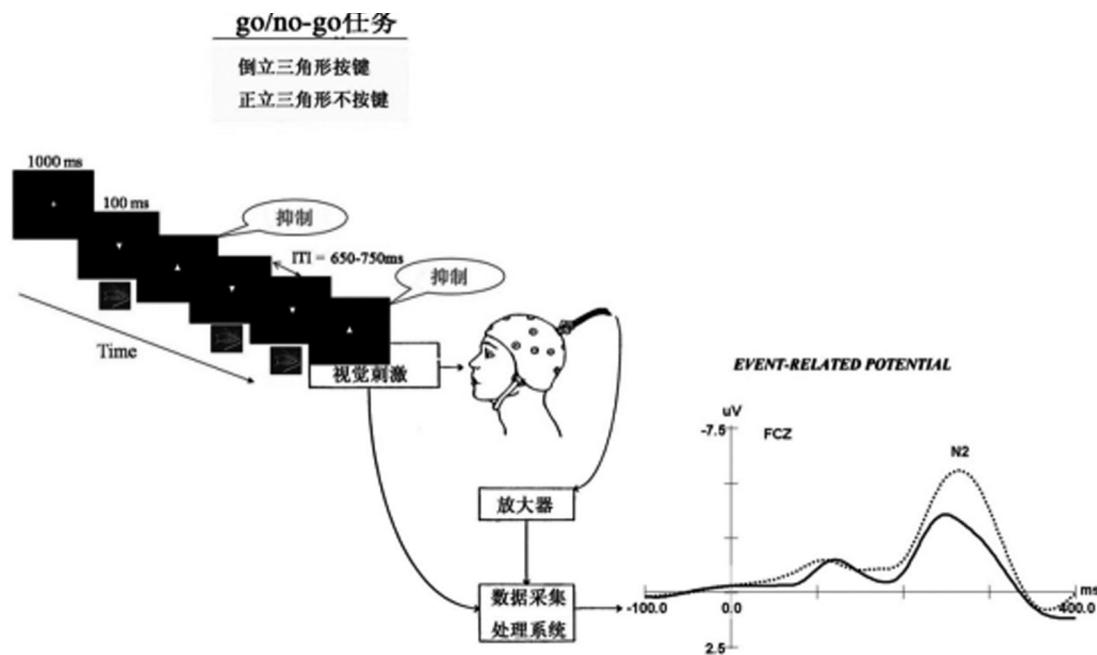
图 1 实验流程及数据采集分析示意图
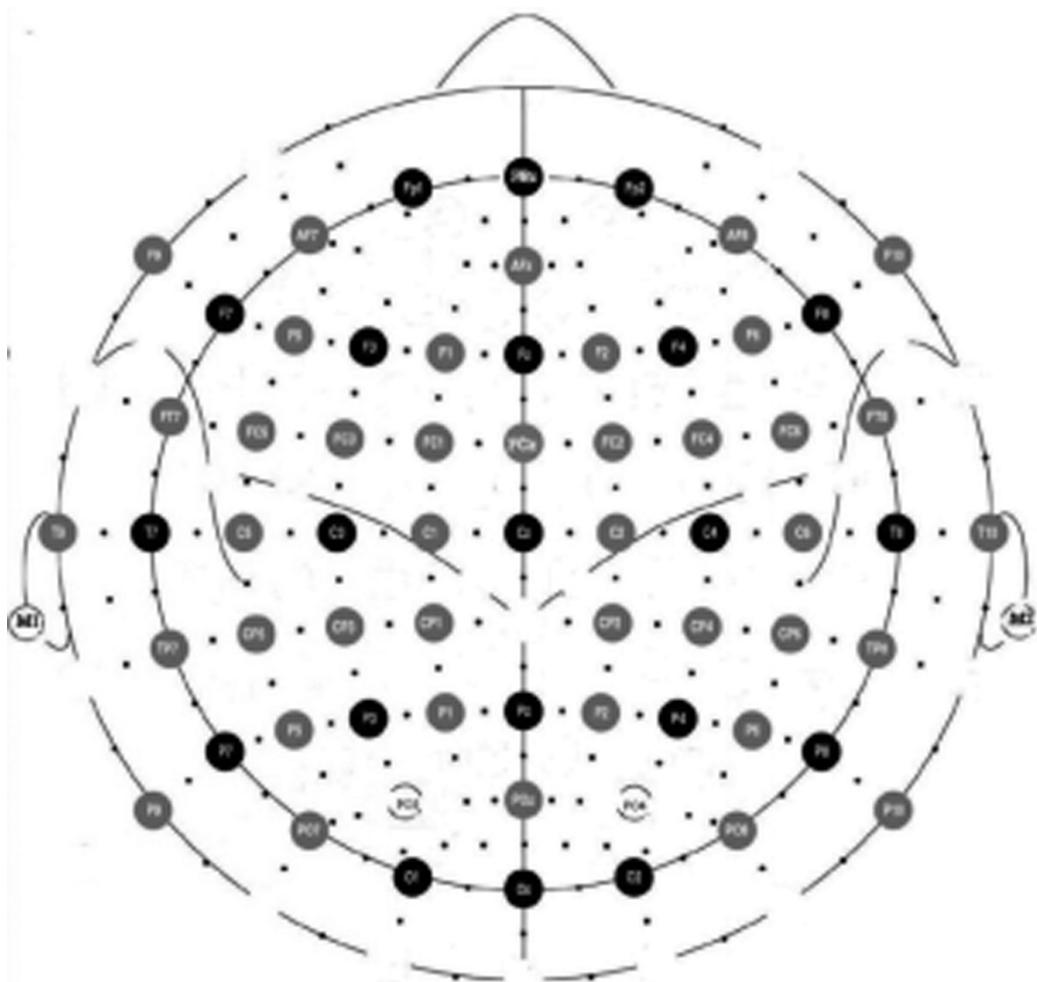
图 2 64 导电极位置分布图

析,因变量为按键反应的反应时与错误率。(2) 脑电分析为组别 2 (自伤组 / 对照组) × 电极位置 9 (9 个电极点) 重复测量的方差分析 , 因变量为 Nogo–N2 幅值和潜伏期。并绘制地形图。当自由度大于 1 时 , 用 Greenhouse–Geisser 法校正 p 值。

3.3 结果

3.3.1 行为学指标—— 反应时与错误率的组间差异比较

分别比较两组被试在不同任务中的反应时与错误率 (见表 7)。对两组按键反应 (Go 刺激正确反应与 Nogo 刺激错误反应) 的反应时进行重复测量的方差分析 , 结果发现任务与组别的交互作用不显著 ,$F(1, 22)$ =0.12, p>0.05。任务的主效应显著 , $F(1, 22)$ =426.61,

p<0.001, 即 Nogo 刺激的错误反应时比 Go 刺激的正确反应时短。组别主效应不显著 , $F(1, 22)$ =0.92, p>0.05, 两组的按键反应时差异不显著。

对两组按键反应的错误率进行重复测量的方差分析 , 任务与组别的交互作用不显著 , $F(1, 22)$ =0.04, p>0.05。组别主效应显著 , $F(1, 22)$ =7.42,p<0.05, 自伤组 Go 刺激与 Nogo 刺激按键的错误率都显著高于对照组。

3.3.2 Nogo–N2 效应的有效性检验

分别对 N2 幅值和潜伏期进行 2 (Go 刺激按键 / Nogo 刺激抑制按键) ×9 (电极位置) ×2 (自伤 / 非自伤) 重复测量的方差分析以检验实验任务是否成功诱发冲动控制反应。结果发现 , 任务对 N2 潜伏期的

表 7　两组按键反应的反应时、错误率 (M ± SD)

组别	Go 刺激正确反应		Nogo 刺激错误反应	
	反应时 (ms)	错误率 (%)	反应时 (ms)	错误率 (%)
自伤组 (n=12)	254.96 ± 18.01	37.50 ± 8.85	212.29 ± 20.36	26.22 ± 11.69
对照组 (n=12)	260.98 ± 15.97	27.83 ± 15.70	219.69 ± 16.60	17.89 ± 7.07

表 8　不同组别 Nogo–N2 幅值、潜伏期差异比较

脑区			潜伏期 (ms)	幅值 (μ V)
额区	F3(左)	自伤组	259.00 ± 31.34	−7.74 ± 4.15
		对照组	254.50 ± 49.46	−3.89 ± 2.29
		$F(1, 22)$	0.07	7.93*
	Fz(中)	自伤组	270.33 ± 19.91	−8.99 ± 4.30
		对照组	258.00 ± 37.67	−4.17 ± 2.00
		$F(1, 22)$	1.01	12.34**
	F4(右)	自伤组	265.17 ± 34.09	−8.37 ± 4.52
		对照组	241.50 ± 39.38	−3.96 ± 2.24
		$F(1, 22)$	2.48	9.18**
额中央区	FC3(左)	自伤组	264.00 ± 22.12	−6.61 ± 3.03
		对照组	247.17 ± 36.95	−2.72 ± 1.64
		$F(1, 22)$	1.83	15.26**
	FCz(中)	自伤组	271.50 ± 9.84	−8.73 ± 2.86
		对照组	254.17 ± 26.81	−3.64 ± 1.37
		$F(1, 22)$	6.75*	30.79***
	FC4(右)	自伤组	269.33 ± 15.76	−7.69 ± 3.75
		对照组	242.50 ± 33.07	−3.22 ± 1.59
		$F(1, 22)$	6.44*	14.47**
中央区	C3(左)	自伤组	269.67 ± 40.80	−3.91 ± 3.62
		对照组	251.83 ± 49.99	−1.36 ± 1.08
		$F(1, 22)$	0.92	5.47*
	Cz(中)	自伤组	269.50 ± 16.25	−1.85 ± 1.68
		对照组	251.67 ± 27.38	−0.43 ± 0.43
		$F(1, 22)$	7.12*	8.04*
	C4(右)	自伤组	281.00 ± 33.56	−3.47 ± 2.34
		对照组	240.50 ± 43.48	−1.57 ± 0.98
		$F(1, 22)$	6.53*	6.75*

注 : *p < 0.05 ; *p < 0.01 ; **p < 0.001 ;

图 3 自伤组与对照组 Nogo-N2 幅值、潜伏期

图 4 自伤组与对照组 Nogo-N2 的地形图比较 (270ms)

主效应边缘显著，F(1, 22) =4.16, p=0.053,Nogo– N2 的潜伏期比 Go–N2 延迟较为显著。任务对 N2 幅值的主效应显著，*F*(1, 22) =23.19, *p*<0.001,Nogo–N2 幅值显著高于 Go–N2。表明实验程序成功诱发了被试的冲动控制反应。

3.3.3　Nogo–N2 幅值、潜伏期的组间差异比较
（1）Nogo–N2 幅值

对两组在不同电极位置的 Nogo–N2 幅值进行 2(组别) × 9 (电极位置) 的重复测量方差分析。组别与电极位置的交互效应显著，F(8, 176) =4.51, *p*<0.01。组别主效应显著，*F*(1, 22) =14.66, *p*=0.001，自伤组的 Nogo–N2 幅值显著高于对照组，*p*<0.001~0.05。电极位置主效应显著，*F*(8, 176)=40.28, *p*<0.001。

（2）Nogo–N2 潜伏期

对两组在不同电极位置的 Nogo–N2 潜伏期进行 2 (组别) × 9 (电极位置) 的重复测量方差分析。电极位置与组别的交互效应不显著，*F*(8, 176) =1.07,*p*>0.05。电极位置主效应不显著，*F*(8, 176) =0.81,*p*>0.05。组别主效应显著，*F*(1, 22) =4.34, *p*<0.05, 自伤组 Nogo–N2 潜伏期明显迟于对照组。进一步检验发现，在 FCz、FC4、Cz、C4 电极位置，自伤组的 Nogo–N2 潜伏期迟于对照组，*p*<0.05, 而其他电极位置没有差异，*p*>0.05 (见表 8 和图 3)。

3.3.3　Nogo–N2 脑电地形图

进一步绘制脑电二维地形图 (刺激开始 270 ms 左右) 比较两组的脑电差异 (见图 4)。红色轴为正电压，标尺正波幅值。蓝色轴为负电压，标尺负波幅值，蓝色调越深，负电压越大。实验任务诱发了 N2 负波，地形图的额区附近均以蓝色调为主。由图可见，成功抑制 Nogo 刺激引起的脑电活动主要集中在额区 (F3/Fz/F4) 附近。在 9 个电极点处，自伤组成功抑制 Nogo 刺激的电压都比对照组更大 (蓝色更深)。与上述方差分析结果一致。

3.4　讨论
3.4.1　行为学指标—— 反应时与错误率
研究 2 在研究 1 的基础上，采用 Go/Nogo 实验范式的 ERPs 方法，检验自伤组与对照组冲动控制的行为学和脑电差异。本研究中 Nogo 刺激错误反应的反应时比 Go 刺激正确反应的反应时短，与以往同类研究一致 (Ciesielski, Harris, & Cofer, 2004;Gehring & Willoughby, 2002; 王振宏 , 2005)。Falkenstein 等 (1991) 研究认为，错误反应的反应时比正确反应的反应时短，表明出错是由快速决策或提前引发反应造成的，是一种冲动反应。本研究中 Nogo 错误反应的反应时比 Go 正确反应的反应时短了 40 多毫秒，且效应显著，这表明实验程序成功诱发了被试的冲动反应。

自伤组与非自伤组按键反应的反应时差异不显著，但自伤组错误率明显高于非自伤组。反应时与错误率是冲动控制的两个行为学指标 (Menon,Adleman, White, Glover, & Reiss, 2001; Ruchsow etal., 2008b; Saunders et al., 2008), 反应时越短或 / 且错误率越高，表明冲动控制能力越差、冲动性越高。Janis 等 (2009) 分别以 64 名自伤青少年和 30 名普通青少年为被试，采用冲动控制实验范式 (Conner'sContinuous Performance Test, CPT) 比较了两组的冲动性水平，结果显示两组被试按键错误率没有显著差异。Glenn 和 Klonsky (2010) 以 82 名大学低年级有过自伤史的学生和 86 名对照组为被试，采用反应控制实验范式 (Stop-Signal Task, SST) 发现两组被试在反应时上差异不显著。两个行为学实验结果均与"自伤组的冲动性更高"的假设不符。本研究中两组在按键错误率上差异显著，证实了研究假设，但反应时与错误率同样作为冲动控制实验范式的行为学指标，在本研究中并未得到一致的结论。根据速度 – 准确率之间的权衡关系可知，当反应速度增加，错误率也会相应增长，表面看来，反应时与错误率的同步变化应该能够同步反映行为特点。但研究 (Prinzmetal, McCool, & Park, 2005; Santee &Egeth, 1982) 表明两者结论的不一致并非偶然，反应时和准确 (或错误) 率并非一定同步表征行为，两者可能来自不同的加工机制。Prinzmetal 等 (2005) 指出，人的注意同时包括自主注意和非自主注意两部分，他们研究发现，在实验室任务中，自主注意同时影响个体的反应时和准确率，而非自主注意仅影响反应时。

由此 Prinzmental 等提出实验室任务中的反应时和准确率结论不一致是正常情况。同时, Santee 和 Egeth (1982) 的研究也表明, 当以速视仪呈现刺激 (即刺激短时呈现) 时, 准确率易受由目标刺激与非目标刺激早期引起的感知觉因素的干扰, 反应时则更易受认知加工后期引起的行为反应因素的干扰。而在刺激长时呈现条件下, 两者均易受知觉后期加工的影响。这说明反应时与准确率有可能是两个独立的成分, 并不能总是同步反映相同的知觉加工, 或者说两个指标有可能分别体现了冲动反应的不同方面。本研究中自伤组的按键错误率高于对照组, 表明自伤青少年较同龄普通人群的冲动反应程度更高。Go/Nogo 任务是检验停止一个正在进行的反应 (或反应类型) 的能力 (趋近行为管理能力), 冲动反应程度高的个体在抑制趋近行为反应时有更高的错误率 (Newman, 1987), 是由于缺乏管理趋近行为的能力 (Lansbergen, 2007)。因此从本研究结论可以推论为, 与普通青少年相比, 自伤青少年缺乏趋近行为管理能力, 在本研究实验任务中可能具体表现为两组在自主注意和早期感知觉特点上存在差异。

3.4.2 脑电指标—— Nogo-N2 幅值、潜伏期与地形图

Go/Nogo 实验范式中, 冲动性高者成功抑制一个优势反应需要耗费更多的神经能量, 表现为抑制 Nogo 刺激比反应 Go 刺激的 N2 幅值更高、潜伏期更长。本研究结果发现, Nogo-N2 的潜伏期比 Go-N2 延迟较为显著, 且 Nogo-N2 幅值显著高于 Go-N2, 表明实验程序成功诱发了被试的冲动控制反应。与前文行为学检验结果一致。

对 Nogo-N2 幅值的重复测量方差分析结果显示, 自伤组的 Nogo-N2 幅值在所有电极点上均显著高于对照组 (p<0.001~0.05), 自伤组的 Nogo-N2 潜伏期在 FCz、FC4、Cz、C4 电极位置明显迟于对照组。前期有部分研究已显示 N2 可能是反映冲动控制水平的电生理指标。Falkenstein 等 (1999) 发现视觉刺激 ERPs 任务中, 错误率高被试的 Nogo 刺激 (同时包括正确和错误) 反应 N2 幅值更小、潜伏期提前; Sunohara 等

(1999) 发现 ADHD 患者的冲动性高于对照组, 表现为 Nogo 反应的 N2 潜伏期提前。本研究中 Nogo-N2 幅值和潜伏期的组间差异显著表明自伤青少年比普通青少年成功抑制优势反应需要更多的神经能量, 也有理由进一步推断这一结果有可能反映了两组冲动控制能力的差异: 自伤组的冲动控制能力更差。该结果同时也为 N2 可能是冲动控制的电生理指标提供了证据。

就两组 Nogo-N2 幅值、潜伏期差异的脑区位置看, 两组差异在额中央区 (FC3/FCz/FC4) 最大; 从左 / 中 / 右脑区分布来看, 中间脑区 (Fz/FCz/Cz) 的差异最大。就脑电地形图来看, 成功抑制 Nogo 刺激的脑电活动主要集中在额区附近, 自伤组成功抑制 Nogo 刺激的额区电压比对照组更大 (蓝色更深)。综合电极位置和脑区地形图的结果可以发现组间差异主要表征在前额皮层 (Prefrontal Cortex, PFC)。前期研究表明 PFC 与行为抑制能力有关。PFC 是认知控制的主要负责脑区, 在情绪加工、识别刺激物的情绪意义及调节情绪反应和行为表现中具有重要作用 (Davidson, Pizzagalli, Nitschke, & Kalin, 2003;Ochsner & Gross, 2005)。反应抑制的 fMRI 研究也表明, 背外侧和腹外侧 PFC 及眶额皮层 (Orbitofrontal Cortex, OFC) 与行为抑制能力密切相关 (Berlin, Rolls, & Iversen, 2005; Horn, Dolan,Elliott, Deakin, & Woodruff, 2003; Liddle, Kiehl, &Smith, 2001)。本研究中发现组间差异主要表征在前额皮层, 进一步表明前额皮层与冲动控制能力的加工有关。Nogo-N2 的幅值增高、潜伏期延迟、分布面积增大也说明, 与普通青少年相比, 自伤青少年抑制一个优势反应需要的神经活动过程可能有所不同, 两者可能存在脑机制的差异。进一步确认该结论还需要更多来自 ERPs、fMRI 研究的佐证。

4 总讨论

本文通过两个研究, 结合自我报告、行为学和脑电指标, 检验了自伤青少年的冲动性。结果发现自我报告的冲动性能够显著预测自伤水平; 自伤组在实

验任务中的按键反应错误率高于对照组；与对照组相比，自伤组在实验任务中成功抑制优势反应诱发的 N2 负波幅值增高、N2 潜伏期延迟、前额区脑电活动增强。自我报告、行为学指标和脑电指标结果一致表明自伤青少年比同龄普通青少年的冲动控制能力更差。

当前为数不多的关于自伤者冲动性水平的研究多来自自我报告和实验室冲动控制实验的行为学的结果。自我报告的结论与研究假设比较一致，自伤组自我报告的冲动性水平显著高于普通组，但自我报告法通常被认为不能准确获得诸如动机、自我控制等需要高级认知参与的行为的信息，从而影响报告结果的可靠性 (Nisbett & Wilson, 1977;Rachlin, 1995)。而多个采用不同研究范式的行为学实验结果均与研究假设和自我报告结果不符，即两组在实验任务中的成绩并无显著差异。行为学指标与自我报告结果的不一致，一方面与冲动性的概念和结构的多样性有关 (Evenden, 1999)，不同测量工具和测量方式有可能测量的是冲动性的不同方面，研究者不应期望自我报告与行为学测量有一致的结果 (Gerbing et al., 1987)；另一方面也表明自伤冲动性的研究需要更多的证据。本研究发现自伤组按键反应的错误率明显高于对照组、相关脑电成分差异显著，为自伤青少年的冲动性提供了新的行为学和脑电证据。

在自伤研究领域，本研究首次采用 ERPs 方法检验自伤群体冲动性，并提供了 N2 指标的证据。在研究冲动控制的脑电实验中，除 N2 之外，P3 是反映冲动控制的另一个脑电成分。P3 是刺激出现之后 300ms 左右出现的脑电正波，研究者检验被试冲动控制能力时通常会同时参考 N2 和 P3 两个指标的结果 (如，Ramautar, Kok, & Ridderinkhof, 2004;Ruchsow et al., 2008a)。N2 和 P3 是否反映了冲动控制加工一直存在很多争议 (Bruin, Wijers, & VanStaveren, 2001; Donkers & Van Boxtel, 2004;Falkenstein et al., 1999; Smith, Johnstone, & Barry,2007)，甚至有人提出 P3 而非 N2，才是冲动控制的指标 (Smith et al., 2007)。但在一些同时考察 N2 和 P3 两个指标的研究中，两者的

结论也不尽一致。比如，Ruchsow 等 (2008a) 对 BPD 患者冲动控制的研究中，BPD 组与对照组在 P3 指标上差异显著，而 N2 指标显示两组不存在差异。而在 Falkenstein (1999) 对反应抑制高低两组脑电成分比较中，两组在 N2 指标上差异显著，但 P3 成分却未出现组间差异。尽管 N2 和 P3 成分的意义表征仍存在争论，但对于自伤群体冲动性水平的后期研究仍然可以同时结合 N2 和 P3 两个指标进行检验。此外，后期研究还可以采用功能性核磁共振 (fMRI) 来提供更多的研究证据。

本研究证实了自伤者冲动控制能力受损，可以帮助临床工作者有针对性的对自伤个案进行冲动控制能力训练 (如 Ainslie, 1975)，以减少自伤行为的发生。但在本研究中冲动性对自伤总变异的解释量不足 15%，与情绪调节困难一起的解释量也不足 30%，这一结果可能与自伤行为本身的复杂性有关。首先是自伤的影响因素众多。除冲动性外，早期创伤经验 (如受虐待、受忽视等)、情绪管理障碍和某些生物学因素等都是自伤的重要危险因子 (见江光荣，于丽霞，郑莺，冯玉，凌霄，2011)，且各个影响因素之间交互作用复杂，这使得单一因素对自伤的解释力非常有限。一个类似的例子是，早期研究曾强调早期创伤经验尤其是性虐待对自伤的影响，但元分析结果 (Klonsky & Moyer, 2008) 却发现性虐待对自伤行为的解释量尚不足 5%。这也提示着在研究和临床干预中，研究者和临床工作者应综合多种因素来理解自伤行为的发生。其次是自伤群体的异质性。从临床案例的观察来看，自伤群体的异质性很高，看似同样的行为实则包含着不同的亚类型。不同自伤亚类背后可能有着非常不同的影响因素。但目前的实际研究均据当前自伤行为的定义 (如 Gratz, 2001) 来选取被试，并未区分不同亚类群体，这使得对某一亚类群体有效的影响因素的解释力被其他亚类所稀释，从而对自伤行为的预测力降低。

5 结论

（1）自伤青少年自我报告的冲动性能显著预测

自伤水平。

（2）自伤青少年的冲动控制能力弱于普通青少年，表现为 Go/Nogo 实验任务中按键反应的错误率更高、成功抑制按键反应产生的 N2 幅值更高、部分电极点处的 N2 潜伏期延迟。

（2）冲动控制能力加工主要表征在前额叶区。

参考文献

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin, 82*(4), 463–496.

An J. (2008). *The study on the relationship of impulsiveness with suicidal ideation.* Unpublished master's thesis, China University of Geosciences.

[安静 . (2008). *冲动性人格特征与自杀意念关系的研究* . 硕士学位论文 , 中国地质大学 .]

Baca-Garcia, E., Diaz-Sastre, C., García Resa, E., Blasco, H., Braquehais Conesa, D., Oquendo, M. A., & De Leon, J. (2005). Suicide attempts and impulsivity. *European Archives of Psychiatry and Clinical Neuroscience, 255*(2), 152–156.

Barratt, E. S. (1994). Impulsiveness and aggression. Violence and Mental Disorder: *Developments in Risk Assessment, 10,* 61–79.

Berkowitz, L. (1974). Some determinants of impulsive aggression: Role of mediated associations with reinforcements for aggression. *Psychological Review, 81*(2), 165–176.

Berlin, H. A., Rolls, E. T., & Iversen, S. D. (2005). Borderline personality disorder, impulsivity, and the orbitofrontal cortex. *American Journal of Psychiatry, 162*(12), 2360–2373.

Bjarehed, J., & Lundh, L. -G. (2008). Deliberate self-harm in 14-year-old adolescents: How frequent is it, and how is it associated with psychopathology, relationship variables, and styles of emotional regulation. *Cognitive Behaviour Therapy, 37*(1), 26–37.

Bruin, K., Wijers, A., & Van Staveren, A. (2001). Response priming in a go/nogo task: Do we have to explain the go/nogo N2 effect in terms of response activation instead of inhibition? *Clinical Neurophysiology, 112*(9), 1660–1671.

Carver, C. S., Johnson, S. L., & Joormann, J. (2008). Serotonergic function, two-mode models of self-regulation, and vulnerability to depression: What depression has in common with impulsive aggression. *Psychological Bulletin, 134*(6), 912–943.

Chapman, A. L., Gratz, K. L., & Brown, M. Z. (2006). Solving the puzzle of deliberate self-harm: The experiential avoidance model. *Behaviour Research and Therapy, 44*(3), 371–394.

Ciesielski, K. T., Harris, R. J., & Cofer, L. F. (2004). Posterior brain ERP patterns related to the go/no-go task in children. *Psychophysiology, 41*(6), 882–892.

Cools, R., Roberts, A. C., & Robbins, T. W. (2008). Serotoninergic regulation of emotional and behavioural control processes. *Trends in Cognitive Sciences, 12*(1), 31–40.

Davidson, R. J., Pizzagalli, D., Nitschke, J. B., & Kalin, N. H. (2003). Parsing the subcomponents of emotion and disorders of emotion: Perspectives from affective neuroscience. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 8–24). New York: Oxford University Press.

Dell'Osso, B., Altamura, A. C., Allen, A., Marazziti, D., & Hollander, E. (2006). Epidemiologic and clinical updates on impulse control disorders: A critical review. *European Archives of Psychiatry and Clinical Neuroscience, 256*(8), 464–475.

Donkers, F. C. L., & Van Boxtel, G. J. M. (2004). The N2 in go/no-go tasks reflects conflict monitoring not response inhibition. *Brain and Cognition, 56*(2), 165–176.

Dougherty, D. M., Mathias, C. W., Marsh, D. M., Papageorgiou, T. D., Swann, A. C., & Moeller, F. G. (2004). Laboratory measured behavioral impulsivity relates to suicide attempt history. *Suicide and Life-Threatening Behavior, 34*(4), 374–385.

Evans, J., Platts, H., & Liebenau, A. (1996). Impulsiveness and deliberate self-harm: A comparison of 'first-timers' and 'repeaters'. *Acta Psychiatrica Scandinavica, 93*(5), 378–380.

Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology, 146*(4), 348–361.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology, 78*(6), 447–455.

Falkenstein, M., Hoormann, J., & Hohnsbein, J. (1999). ERP components in Go/Nogo tasks and their relation to inhibition. *Acta Psychologica, 101*(2–3), 267–291.

Feng Y. (2008). *The relation of adolescents self-harm behaviors, individual emotion characteristic and family environment factors.* Unpublished master's thesis, Central China Normal University.

[冯玉 . (2008). *青少年自我伤害行为与个体情绪因素和家庭环境因素的关系* . 硕士学位论文 , 华中师范大学 .]

Fischer, S., Smith, G. T., & Anderson, K. G. (2003). Clarifying the role of impulsivity in bulimia nervosa. *International Journal of Eating Disorders, 33*(4), 406–411.

Folstein, J. R., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*(1), 152–170.

Gehring, W. J., & Willoughby, A. R. (2002). The medial frontal cortex and the rapid processing of monetary gains and losses. *Science, 295*(5563), 2279–2282.

Gerbing, D. W., Ahadi, S. A., & Patton, J. H. (1987). Toward a conceptualization of impulsivity: Components across the behavioral and self-report domains. *Multivariate Behavioral Research, 22*(3), 357–379.

Glenn, C. R., & Klonsky, E. D. (2010). A multimethod analysis of impulsivity in nonsuicidal self-injury. Personality Disorders: *Theory, Research, and Treatment, 1*(1), 67–75.

Gratz, K. L. (2001). Measurement of deliberate self-harm: Preliminary data on the Deliberate Self-Harm Inventory. *Journal of Psychopathology and Behavioral Assessment, 23*(4), 253–263.

Gratz, K. L. (2007). Targeting emotion dysregulation in the treatment of self-injury. *Journal of Clinical Psychology, 63*(11), 1091–1103.

Gratz, K. L., & Chapman, A. L. (2007). The role of emotional responding and childhood maltreatment in the development and maintenance of deliberate self-harm among male undergraduates. *Psychology of Men & Masculinity, 8*(1), 1–14.

Gratz, K. L., & Gunderson, J. G. (2006). Preliminary data on an acceptance-based emotion regulation group intervention for deliberate self-harm among women with borderline personality disorder. *Behavior Therapy, 37*(1), 25–35.

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment, 26*(1), 41–54.

Haw, C., Hawton, K., Houston, K., & Townsend, E. (2001). Psychiatric and personality disorders in deliberate self-harm patients. *British Journal of Psychiatry, 178*, 48–54.

Hawton, K., Rodham, K., Evans, E., & Weatherall, R. (2002). Deliberate self harm in adolescents: Self report survey in schools in England. *BMJ: British Medical*

Journal, 325(7374), 1207–1211.

Hawton, K., Zahl, D., & Weatherall, R. (2003). Suicide following deliberate self–harm: Long–term follow–up of patients who presented to a general hospital. *British Journal of Psychiatry, 182*(6), 537–542.

Herpertz, S., Sass, H., & Favazza, A. (1997). Impulsivity in self–mutilative behavior: Psychometric and biological findings. *Journal of Psychiatric Research, 31*(4), 451–465.

Hilt, L. M., Cha, C. B., & Nolen–Hoeksema, S. (2008). Nonsuicidal self–injury in young adolescent girls: Moderators of the distress–function relationship. *Journal of Consulting and Clinical Psychology, 76*(1), 63–71.

Hilt, L. M., Nock, M. K., Lloyd–Richardson, E. E., & Prinstein, M. J. (2008). Longitudinal study of nonsuicidal self–injury among young adolescents: Rates, correlates, and preliminary test of an interpersonal model. *The Journal of Early Adolescence, 28*(3), 455–469.

Horn, N., Dolan, M., Elliott, R., Deakin, J. F. M., & Woodruff, P. W. R. (2003). Response inhibition and impulsivity: An fMRI study. *Neuropsychologia, 41*(14), 1959–1966.

Janis, I. B., & Nock, M. K. (2009). Are self–injurers impulsive? Results from two behavioral laboratory studies. *Psychiatry Research, 169*(3), 261–267.

Jentsch, J. D., & Taylor, J. R. (1999). Impulsivity resulting from frontostriatal dysfunction in drug abuse: Implications for the control of behavior by reward–related stimuli. *Psychopharmacology, 146*(4), 373–390.

Jiang, G. R., Yu L. X., Zheng, Y., Feng Y., & Ling X. (2011). The current status, problems and recommendations on non–suicidal self–injury in China. *Advances in Psychological Science, 19*(6), 861–873.

[江光荣 , 于丽霞 , 郑莺 , 冯玉 , 凌霄 . (2011). 自伤行为研究 : 现状、问题与建议 . *心理科学进展，19*(6), 861–873.]

Jodo, E., & Kayama, Y. (1992). Relation of a negative ERP component to response inhibition in a Go/No–go task. *Electroencephalography and Clinical Neurophysiology, 82*(6), 477–482.

Kiehl, K. A., Smith, A. M., Hare, R. D., & Liddle, P. F. (2000). An event–related potential investigation of response inhibition in schizophrenia and psychopathy. *Biological Psychiatry, 48*(3), 210–221.

Klonsky, E. D. (2007). The functions of deliberate self–injury: A review of the evidence. *Clinical Psychology Review, 27*(2), 226–239.

Klonsky, E. D. (2009). The functions of self–injury in young adults who cut themselves: Clarifying the evidence for affect–regulation. *Psychiatry Research, 166*(2), 260–268.

Klonsky, E. D., & Moyer, A. (2008). Childhood sexual abuse and non–suicidal self–injury: Meta–analysis. *British Journal of Psychiatry, 192*(3), 166–170.

Lansbergen, M. (2007). *Impulsivity: A deficiency of inhibitory control?* Unpublished doctorial dissertation, Utrecht University.

Liddle, P. F., Kiehl, K. A., & Smith, A. M. (2001). Event–related fMRI study of response inhibition. *Human Brain Mapping, 12*(2), 100–109.

Links, P. S., Heslegrave, R., & Reekum, R. (1999). Impulsivity: Core aspect of borderline personality disorder. *Journal of Personality Disorders, 13*(1), 1–9.

Luijten, M., Littel, M., & Franken, I. H. A. (2011). Deficits in inhibitory control in smokers during a Go/Nogo task: An investigation using event–related brain potentials. *PLoS ONE, 6*(4), e18898.

Mc Closkey, M. S., Look, A. E., Chen, E. Y., Pajoumand, G., & Berman, M. E. (2012). Nonsuicidal self–injury: Relationship to behavioral and self–rating measures of impulsivity and self–aggression. *Suicide and Life–*

Threatening Behavior, 42(2), 197–209.

Menon, V., Adleman, N., White, C., Glover, G., & Reiss, A. (2001). Error–related brain activation during a Go/Nogo response inhibition task. *Human Brain Mapping, 12*(3), 131–143.

Messer, J., & Fremouw, W. (2008). A critical review of explanatory models for self–mutilating behaviors in adolescents. *Clinical Psychology Review, 28*(1), 162–178.

Miller, J., Flory, K., Lynam, D., & Leukefeld, C. (2003). A test of the four–factor model of impulsivity–related traits. *Personality and Individual Differences, 34*(8), 1403–1418.

Moeller, F. G., Barratt, E. S., Dougherty, D. M., Schmitz, J. M., & Swann, A. C. (2001). Psychiatric aspects of impulsivity. *American Journal of Psychiatry, 158*(11), 1783–1793.

Munro, G. E. S., Dywan, J., Harris, G. T., McKee, S., Unsal, A., & Segalowitz, S. J. (2007). Response inhibition in psychopathy: The frontal N2 and P3. *Neuroscience Letters, 418*(2), 149–153.

Newman, J. P. (1987). Reaction to punishment in extraverts and psychopaths: Implications for the impulsive behavior of disinhibited individuals. *Journal of Research in Personality, 21*(4), 464–480.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84*(3), 231–259.

Nock, M. K., Joiner, T. E., Jr., Gordon, K. H., Lloyd–Richardson, E., & Prinstein, M. J. (2006). Non–suicidal self–injury among adolescents: Diagnostic correlates and relation to suicide attempts. *Psychiatry Research, 144*(1), 65–72.

Nock, M. K., & Prinstein, M. J. (2004). A functional approach to the assessment of self–mutilative behavior. *Journal of Consulting and Clinical Psychology, 72*(5), 885–890.

Nock, M. K., & Prinstein, M. J. (2005). Contextual features and behavioral functions of self–mutilation among adolescents. *Journal of Abnormal Psychology, 114*(1), 140–146.

Ochsner, K. N., & Gross, J. J. (2005). The cognitive control of emotion. *Trends in Cognitive Sciences, 9*(5), 242–249.

Patton, J. H., Stanford, M. S., & Barratt, E. S. (1995). Factor structure of the Barratt impulsiveness scale. *Journal of Clinical Psychology, 51*(6), 768–774.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.

Prinzmetal, W., McCool, C., & Park, S. (2005). Attention: Reaction time and accuracy reveal different mechanisms. *Journal of Experimental Psychology: General, 134*(1), 73–92.

Rachlin, H. (1995). Self–control: Beyond commitment. *Behavioral and Brain Sciences, 18*(1), 109–121.

Ramautar, J., Kok, A., & Ridderinkhof, K. (2004). Effects of stop–signal probability in the stop–signal paradigm: The N2/P3 complex further validated. *Brain and Cognition, 56*(2), 234–252.

Ruchsow, M., Groen, G., Kiefer, M., Buchheim, A., Walter, H., Martius, P., ···Falkenstein, M. (2008a). Response inhibition in borderline personality disorder: Event–related potentials in a Go/Nogo task. *Journal of Neural Transmission, 115*(1), 127–133.

Ruchsow, M., Groen, G., Kiefer, M., Hermle, L., Spitzer, M., & Falkenstein, M. (2008b). Impulsiveness and ERP components in a Go/Nogo task. *Journal of Neural Transmission, 115*(6), 909–915.

Santee, J. L., & Egeth, H. E. (1982). Do reaction time and accuracy measure the same aspects of letter recognition? Journal of Experimental Psychology:

Human Perception and Performance, 8(4), 489–501.

Saunders, B., Farag, N., Vincent, A. S., Collins Jr, F. L., Sorocco, K. H., & Lovallo, W. R. (2008). Impulsive errors on a Go/Nogo reaction time task: Disinhibitory traits in relation to a family history of alcoholism. *Alcoholism: Clinical and Experimental Research, 32*(5), 888–894.

Sho, N., Oiji, A., Konno, C., Toyohara, K., Minami, T., Arai, T., & Seike, Y. (2009). Relationship of intentional self–harm using sharp objects with depressive and dissociative tendencies in pre–adolescence–adolescence. *Psychiatry and Clinical Neurosciences, 63*(3), 410–416.

Simeon, D., Stanley, B., Frances, A., Mann, J. J., Winchel, R., & Stanley, M. (1992). Self–mutilation in personality disorders: Psychological and biological correlates. *American Journal of Psychiatry, 149*(2), 221–226.

Sinclair, J. M. A., Hawton, K., & Gray, A. (2010). Six year follow–up of a clinical sample of self–harm patients. *Journal of Affective Disorders, 121*(3), 247–252.

Smith, J. L., Johnstone, S. J., & Barry, R. J. (2007). Response priming in the Go/Nogo task: The N2 reflects neither inhibition nor conflict. *Clinical Neurophysiology, 118*(2), 343–355.

Spector, P. E. (2006). Method variance in organizational research. *Organizational Research Methods, 9*(2), 221–232.

Stanford, M. S., Mathias, C. W., Dougherty, D. M., Lake, S. L., Anderson, N. E., & Patton, J. H. (2009). Fifty years of the Barratt Impulsiveness Scale: An update and review. *Personality and Individual Differences, 47*(5), 385–395.

Sunohara, G. A., Malone, M. A., Rovet, J., Humphries, T., Roberts, W., & Taylor, M. J. (1999). Effect of methylphenidate on attention in children with attention deficit hyperactivity disorder (ADHD): ERP evidence. *Neuropsychopharmacology, 21*(2), 218–228.

Suyemoto, K. L. (1998). The functions of self–mutilation. *Clinical Psychology Review, 18*(5), 531–554.

Svirko, E., & Hawton, K. (2007). Self–injurious behavior and eating disorders: The extent and nature of the association. *Suicide and Life–Threatening Behavior, 37*(4), 409–421.

Swann, A. C., Dougherty, D. M., Pazzaglia, P. J., Pham, M., Steinberg, J. L., & Moeller, F. G. (2005). Increased impulsivity associated with severity of suicide attempt history in patients with bipolar disorder. *American Journal of Psychiatry, 162*(9), 1680–1687.

van Boxtel, G. J. M., van der Molen, M. W., Jennings, J. R., & Brunia, C. H. M. (2001). A psychophysiological analysis of inhibitory motor control in the stop–signal paradigm. *Biological Psychology, 58*(3), 229–262.

Wang, Z. H. (2005). *Affective style and aggressive behavior in adolescents.* Unpublished doctoral dissertation. Capital Normal University.

[王振宏 . (2005). 青少年情感风格与攻击行为 . 博士论文 , 首都师范大学 .]

Wen, Z. L., Hau, K. T., & Marsh, H. W. (2004). Structural equation model testing: Cutoff criteria for goodness of fit indices and chi square test. *Acta Psychologica Sinica, 36,* 186–194.

[温忠麟 , 侯杰泰 , 马什赫伯特 . (2004). 结构方程模型检验：拟合指数与卡方准则 . *心理学报 , 36,* 186–194.]

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences, 30*(4), 669–689.

Xiong, H. X., Zhang, J., Ye, B. J., Zheng, X., & Sun, P. Z. (2012). Common method variance effects and the models of statistical approaches for controlling it. *Advances in Psychological Science, 20*(5), 757–769.

[熊红星 , 张璟 , 叶宝娟 , 郑雪 , 孙配贞 . (2012). 共同方法变异的影响及其统计控制途径的模型分析 . *心*

理科学进展, *20*(5), 757–769.]

Yang, B., Yang, S. Y., Zhao, L., Yin, L. H., Liu, X., & An, S. S. (2009). Event–related potentials in a Go/Nogo task of abnormal response inhibition in heroin addicts. *Science in China Series C: Life Sciences, 52*(8), 780–788.

Yang, H. Q., Yao, S. Q., Zhu, X. Z., Auerbach, R. P., Abela John, R. Z., & Tong, X. (2007). The Chinese version of the Barratt Impulsiveness Scale 11th version (BIS–11) in college students: Its reliability and validity. *Chinese Mental Health Journal, 21*(4), 223–225.

[杨会芹 , 姚树桥 , 朱熊兆 , Auerbach, R. P., Abela John, R. Z., & Tong, X. (2007). Barratt 冲动量表中文版在 209 名大三学生中的试用 . *中国心理卫生杂志 , 21*(4), 223–225.]

Yates, T. M., Carlson, E. A., & Egeland, B. (2008). A prospective study of child maltreatment and self–injurious behavior in a community sample. *Development and Psychopathology, 20*(2), 651–671.

Yu, L. X., Jiang, G. R., & Wu, C. Z. (2011). Psychological assessment and treatment of self–injury. *Chinese Mental Health Journal, 25*(12), 937–941.

[于丽霞 , 江光荣 , 吴才智 . (2011). 自伤行为的心理学评估与治疗 . *中国心理卫生杂志 , 25*(12), 937–941.]

Zheng, Y. (2006). *Epidemiologic investigation of self–mutilation behavior among adolescents in Wuhan and its functional model.* Unpublished master's thesis, Central China Normal University.

[郑莺 . (2006). *武汉市中学生自我伤害行为流行学调查及其功能模型* . 硕士学位论文 , 华中师范大学 .]

Zhou, H., & Long, L. R. (2004). Statistical remedies for common method biases. *Advances in Psychological Science, 12*(6), 942–950.

[周浩 , 龙立荣 . (2004). 共同方法偏差的统计检验与控制方法木 . *心理科学进展 , 12*(6), 942–950.]

Zhou, L., He, X. Y., & Xiao, S. Y. (2006). Measurement methods of impulsiveness. *Chinese Journal of Clinical Psychology, 14*(5), 455–457.

[周亮 , 何晓燕 , 肖水源 . (2006). 冲动性测量的方法学问题 . *中国临床心理学杂志 , 14*(5), 455–457.]

Zhou, L., Xiao, S. Y., He, X. Y., Li, J., & Liu, H. M. (2006). Reliability and validity of Chinese version of Barratt Impulsiveness Scale–11. *Chinese Journal of Clinical Psychology, 14*(4), 342–344.

[周亮 , 肖水源 , 何晓燕 , 厉洁 , 刘慧铭 . (2006). BIS–11 中文版的信度与效度检验 . *中国临床心理学杂志 , 14*(4), 342–344.]

Zlotnick, C., Mattia, J. I., & Zimmerman, M. (1999). Clinical correlates of self–mutilation in a sample of general psychiatric patients. *Journal of Nervous and Mental Disease, 187*(5), 296–301.

Impulsivity in non-suicidal self-injurious adolescents in China

YU Lixia[1,2]; LING Xiao[2,3]; JIANG Guangrong[1,2]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan 430079, China)

([2] Key Laboratory of Human Development and Mental Health of Hubei Province, School of Psychology,

Central China Normal University, Wuhan 430079, China)

([3] Wuhan Foreign Languages School, Wuhan 430022, China)

Abstract　Impulsivity has been proposed as an important risk factor in Non−Suicidal Self−Injury (NSSI). Yet, research outcomes on the relationship of impulsivity to NSSI have been mixed. The present study clarifies this relationship using event−related potentials (ERPs), along with self−reports and behavioral measures. Study 1 aimed to detect the prediction of emotion dysregulation and impulsivity to NSSI. 820 local common high school students and 72 counterparts with problematic behaviors were surveyed, and then the relation among NSSI, difficulties of emotion regulation (DER) and impulsivity were investigated by self−report measurements. Regression analysis results indicated that both DER and impulsivity could well predict NSSI, and contribution of impulsivity was much bigger than that of DER. In Study 2, a Go/Nogo paradigm was adopted to test the impulsivity of the injurers using behavioural measures and Nogo−N2 of ERPs. Participants were 12 confirmed self−injurious adolescents and 12 typical school middle students chosen from Study 1. The group differences (injurers vs controls) in behavior (response time and false alarm) and ERPs index (N2 amplitude and latency in successful Nogo trials) were analyzed in detail. Results disclosed that the NSSI group's probability of false alarm was higher than the control group's probability of false alarm in both Go and Nogo trials. In ERPs experiment, the NSSI group's N2 amplitudes were significantly higher than the controls in correct Nogo trials, and NSSI group's N2 latencies were clearly more delayed than the controls' in correct Nogo trial. Results from Nogo−N2 amplitudes and latencies combining with the topographic maps showed that impulse processes occurred in prefrontal cortex mainly. According to the results from self−reports, behavioural measures and Nogo−N2 of ERPs, it can be concluded that self−injurious adolescents possessed stronger impulsivity; and they needed much more neural energy to fulfill an impulse inhibition; moreover, they were insensitive to Nogo stimuli. The present study is the first to examin Nogo−N2 in NSSI, and provides further evidence for impaired response inhibition in NSSI.

Keywords　non−suicidal self−injury; NSSI; impulsivity; ERPs; N2

心理咨询中的突然获益：工作同盟及初始症状的影响 *

朱文臻，江光荣 **，朱　旭，陶　嵘

（华中师范大学心理学院 , 湖北省人的发展与心理健康重点实验室 , 武汉 , 430079）

摘　要　对 82 位当事人在咨询中的症状改变过程进行纵向研究，检验突然获益模式（Sudden Gain, SG）的发生及其与咨询效果的联系，并探索初始症状水平的影响。结果表明，24.93% 的当事人经历了 SG，咨询效果显著好于其他当事人；SG 组与对照组在初始症状水平和总体工作同盟水平上差异显著。结论：SG 出现于本土心理咨询中，经历 SG 的当事人取得了更好的咨询效果，工作同盟与初始症状水平与 SG 的发生相关。

关键词　突然获益；咨询效果；工作同盟

突然获益（Sudden Gain, SG）是指当事人在两次相邻会谈的间隔期间出现症状急剧降低的现象，最早由 Tang 和 DeRubeis（1999）在认知行为疗法对抑郁症的治疗过程中发现。经历 SG 的当事人（以下简称"SG 组"）在治疗结束时的抑郁水平显著低于未经历此模式的当事人（以下简称"对照组"），即获得了更好的治疗效果，在后续效果的保持上（治疗结束后第 6 个月，第 18 个月），也表现出了明显的优势。这表明，SG 并不是随机的症状波动，而是有意义且持久的。

SG 是咨询效果改变过程中的一种不连续且有意义的现象，这种不连续的改变表示当事人正在经历重要的改变过程，且重要的因素正在起作用，因此能够提供很有价值的信息，为理解咨询的改变过程和研究改变机制提供很好的视角。大量研究对 SG 现象进行了检验，发现 SG 并非认知行为治疗所特有，在支持表达治疗（Tang, Luborsky, & Andrusyna, 2002）、包括各种流派和心理问题的常规临床机构（Stiles et al., 2003）、非指导的支持性治疗（Gaynor et al., 2003）、人际心理治疗（Kelly, Cyranowski, & Frank, 2007）、药物治疗与安慰剂组（Vittengl, Clark & Jarrett, 2005）以及未接受治疗条件下（Kelly, Roberts, & Bottonari, 2007）中都发现了 SG。SG 在其他心理问题如社交恐惧 (Hofmann et al., 2006)、焦虑障碍（Present et al., 2008）中同样存在，这证明 SG 也是跨治疗方法和人群的普遍现象（e.g. Busch, Kanter, & Landes, 2006; Hardy, Cahill, & Stiles, 2005; Stiles et al., 2003），并使 SG 现象更加受到研究者们的关注。

然而，在这些大量的研究中，并非所有的研究都证明了 SG 与咨询效果的关系（e.g. Kelly, Cyranowski, & Frank, 2007），且各研究得到的 SG 在大小、发生率及复发率上存在一定差异，见表 1。这些差异可能

<center>表 1 SG 的几个主要研究的比较</center>

文献	样本量	当事人问题类型	治疗条件	测量量表	是否出现SG	SG的大小	平均发生时间	发生比例（%）	复发率（%）	预测咨询效果
Tang & DeRubeis, 1999	61	抑郁	CBT	BDI	是	11.2	5	39	17	是
Tang et al., 2002	35	抑郁	SET	BDI	是	10.6	5	43	47	是
Stiles et al., 2003	135	各种类型	常规临床条件	CORE-SF	是	1.11	5	17	43	是
Vittengl et al., 2005	33，32，24	抑郁	CT, PT, PP	BDI	是	10.39	4	30，42，29	19	是
Hardy et al., 2005	76	抑郁	CT	CORE-SF	是	1.13	5	35	30	是
Hofmann et al., 2006	107	社交恐惧	CBGT, EGT	LSAS	是	15.78	4, 5	18.69	10	是
Kelly et al., 2007	60	抑郁	未接受任何治疗	BDI-Ⅱ	是	12.5	–	60	54.3	否
Present et al., 2008	68	焦虑障碍	SET	BAI	是	11.5	8	16.2	36.3	否

注：BDI：贝克抑郁量表；BID-Ⅱ：贝克抑郁量表第二版；CORE-SF：临床效果常规评估–简版；LSAS：列博维茨社交焦虑量表；BAI：贝克焦虑量表 CBT：认知行为治疗；SET：支持表达治疗；常规临床条件：包括各种流派；CT：认知治疗；PT：药物配合临床管理；PP：安慰剂配合临床管理；CBGT：全面认知行为治疗；EGT：暴露治疗。各种类型：在研究中，被试并不经过诊断标准的专门筛选，只要是需要治疗并适合作为门诊病人的，都参与研究。

揭示了不同流派的咨询过程、心理病理发展过程等方面的不同。本研究的目的之一即检验国内的样本中是否也出现 SG 现象，以及 SG 的特点，并与国外的研究结论作比较。

对于与 SG 的影响因素，研究之间尚存在许多争论。一些研究者认为是治疗中的认知改变导致了症状的突然降低（e.g. Tang et al., 2007; Tang & DeRubeis, 1999），但其他研究者更倾向于用共同因素来解释，因为 SG 是跨流派和人群的普遍现象，而"认知改变"这一因素并不是在所有治疗方法中都存在的（e.g. Busch et al., 2006; Hardy et al., 2005）。工作同盟是咨询中重要的共同要素之一，指咨询过程中咨访双方的合作水平，包括目标一致、任务一致和情感联结（Bordin, 1979），大量研究表明工作同盟可以有效地预测咨询效果（Lambert & Ogles, 2004）。Tang 和 DeRubeis（1999）在认知行为治疗中发现 SG 发生后一次会谈的工作同盟水平显著高于前一次会谈（即关键会谈），他们认为症状的减轻会导致工作同盟的改善，然而这一结论是否在其他治疗流派中也成立，还需进一步探索。而针对工作同盟水平是否可以预测 SG，Hardy 等（2005）比较了认知治疗中 SG 组与对照组初次会谈的工作同盟水平，发现差异不显著，由此推得工作同盟不能预测 SG 的产生。然而由于工作同盟在咨询中是动态变化的，仅比较初次会谈的工作同盟水平来说明两者的关系是不严密的。为避免这一问题，我们采用比较 SG 组的初次会谈与关键

会谈的工作同盟水平，来推断是否工作同盟的水平发展到一定的程度时，导致了 SG 的产生。

另外，在当事人特点方面，有研究表明，初始症状更严重的当事人更多地经历了 SG，因为从测量学角度来讲，他们有更大的改变空间（Vittengl, et al., 2005），但也有研究的结论显示 SG 组与对照组的初始症状水平没有显著差异（e.g. Hardy et al., 2005; Tang & DeRubeis, 1999），对于上述结论的不一致还需要进一步的探索。有研究表明，初始症状严重的当事人在咨询中的改变过程主要体现为，症状水平降低和主观幸福感提高的速率较快，而生活功能的改善较慢（Stulz & Lutz, 2007），因此初始症状水平高的当事人可能更容易经历 SG。

鉴于以上论述，本文拟对国内的心理咨询过程中的 SG 现象进行研究，并提出以下假设：1、有一定比例的中国当事人在心理咨询过程中会出现 SG，且咨询效果显著好于对照组；2、初始症状水平高，或在工作同盟水平好的咨询中，当事人更容易产生 SG；3、工作同盟与 SG 存在相互的影响：工作同盟发展到一定程度时导致了 SG 的产生，而 SG 进一步促进工作同盟水平的提高。

1 方法

1.1 样本

在武汉市 5 所大学的心理咨询中心开展纵向研

究，共有 61 位咨询师的 469 位当事人对 1429 次咨询会谈进行了有效评估。挑选符合以下标准的个案进入本研究：（1）已结案；（2）会谈次数大于等于 4；（3）有连续 4 次以上的数据。

符合以上标准的共有 82 名当事人，其中女性 61 名（74.4%），男性 21 名（25.6%）。高中生 2 名（2.4%），大学生 41 名（50%），研究生 37 名（45.1%），2 名未填写。年龄为 18 至 31 岁，$M=22.40$，$SD=3.05$。会谈次数 4 至 18 次不等，$M=8.48$，$SD=3.31$，$Mdn=8$。有 21.15% 的会谈的数据缺失。共 39 位咨询师参与研究，其中男性 8 名（20.5%），女性 31 名（79.5%），年龄 24 至 58 岁，$M=36.33$，$SD=10.42$。咨询经验从 1 年至 22 年，$M=6.06$，$SD=5.38$，$Mdn=5$。咨询师对自己的理论取向进行 5 级评定，平均分由高到低分别为以人为中心（4.18）、认知行为（3.36）及心理动力（2.95）取向。每个咨询师提供 1 至 3 个个案。

1.2 测量工具

咨询效果问卷简版（Outcome Questionnaire-5，OQ-5） OQ 是由 Lambert, Morton 和 Hatfield (2004) 开发的国外广泛使用的咨询效果评估工具，测量当事人的症状困扰、人际关系和社会角色三方面的内容，共 45 题，故也称 OQ-45。OQ-45 采用 Likert 五级记分，得分越高，心理问题越严重。考虑到 OQ-45 中许多项目中国的当事人少有发生，且重复测量当事人负担较重，于是根据以往的研究，在 OQ-45 中选取了 5 道与总分相关最高的项目，组成 OQ-5。这 5 题是原问卷中的第 9、10、15、23、42 题，均为症状困扰维度的项目，且主要测量的是抑郁症状。运用以往的研究数据分析得到，OQ-5 与 OQ-45 总分 r=0.89 至 .90，内 $\alpha=0.82$ 至 .87。（胡姝婧，2008；秦佑凤，胡姝婧，2008; 于丽霞，2009）。将抑郁水平作为咨询效果指标是一种常用的做法，因为抑郁作为一种状态，对改变敏感，适合重复测量时用，因此是衡量改变的最有效的方法（Lambert & Hawkins, 2004），并且由于国外 SG 的研究大多采用抑郁水平的测量，因此本研究同样采用抑郁水平作为效果指标，便于研究

结论之间的比较。

工作同盟问卷（WAQ） WAQ 是朱旭（2010）在对中国当事人进行质的研究基础上，编制的工作同盟问卷，该问卷包括情感联结、目标任务和投入三个维度，共 12 个项目，每个维度 4 个项目，采用 Likert 五级记分（1= 很少，3= 经常，5= 总是）。测量学指标良好 G 各维度 $\alpha=0.70$ 至 .80，与国外广泛使用的工作同盟问卷简版修订版（WAI-SR）（Hatcher&Gillaspy,2006）的总分 r=0.86，说明两者整体上对工作同盟的测量较为相似；验证性因素分析表明 WAQ 具有良好的结构效度（$\chi 2/df<2$，RMSEA<.08，NNFI 和 CFI 均大于 .95）；且 WAQ 有效地预测咨询效果，对咨询效果的解释量与元分析相当（Martinetal.,2000），由于国外大量研究表明工作同盟可以预测咨询效果，这说明 WAQ 具有一定的效标关联效度。

1.3 程序

在各咨询中心由接待员邀请当事人参与研究，同意参与研究的当事人每次咨询会谈结束后填写 WAQ 和 OQ-5，直到结案。当事人被保证所填问卷不会被咨询师看到，不会影响到咨询。当事人会收到一些小礼物作为对参与研究的感谢。

1.4 数据分析

1.4.1 SG 的标准

根据 Tang 和 DeRubeis（1999）最早对 SG 的定义，SG 必须符合以下三个条件：（1）改变量足够大；（2）与改变之前的抑郁水平相关联；（3）与个体的症状波动相关联。由此提出了三个标准：改变量应（1）大等于 7 个 BDI 分值（贝克抑郁量表，在研究中测量抑郁水平）；（2）至少等于改变之前症状水平的25%；（3）改变后 3 次会谈的症状平均值显著小于改变前 3 次会谈。

由于本研究采用的量表是 OQ-5，因此需要重新确定标准一的分值。Tang 和 DeRubeis（1999）在标准一中采用的 7 分是会谈间 BDI 改变分值频数分布点的次峰值。Stiles 等人（2003）认为，7 接近于 BDI 的可靠改变指数（Reliable Change Index, RCI）6.18分，

因此可以根据量表的 RCI 来确定标准一。RCI 是咨询前后临床上显著的改变值，等于咨询总改变量的平均值加 1.96 个标准误。根据此算法，本研究使用另一个样本 [为最大程度地利用数据，在所有 437 名当事人中，挑选出初次会谈和最后一次会谈数据完整的当事人，共 117 名进入 RCI 的计算。M=4.1111，SE=0.4453，得到 RCI=4.98。

由于标准三的计算要求 SG 发生前至少有 3 次会谈，因此无法考察第 1、2 次会谈的 SG，故需要对标准三做修改。对此，Busch 等人（2006）提出对于前两次会谈可将标准三定为：SG 的 50% 至少在后面 2 次会谈得到保持。

由此，本研究对 SG 的标准定义为：（1）SG ≥ 5OQ-5；（2）至少等于改变之前症状水平的 25%；（3）SG 后 3 次会谈的 OQ-5 平均值显著小于前 3 次会谈，若 SG 前（后）只有 2 次会谈，则只使用 2 次会谈；对于第 1、2 次会谈出现的 SG，要求其 50% 至少在后面 2 次会谈得以保持。另外，按照 Tang 等（2007）的建议，对缺失会谈不进行 SG 的分析，但若标准 3 的计算中遇到缺失值，则用缺失之前或之后的 OQ-5 值代替。

2 结果

2.1 SG 的发生

根据 SG 的界定标准，共有 20 位当事人经历了 SG，占总人数的 24.39%，平均发生在第 2.25 次会谈
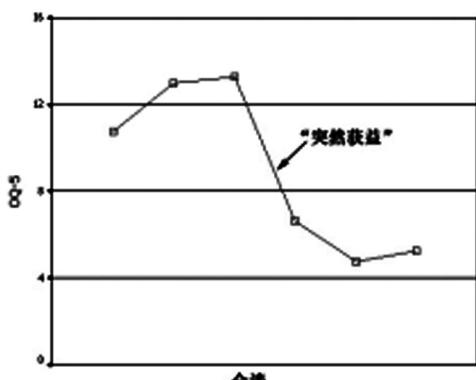
图 1 平均的 SG 以及前三次会谈和后三次会谈 OQ-5 的平均值

后（Mdn=1,SD=1.83），SG 平均为 6.85 分（SD=1.35），占咨询总改变量（M=8.5，SD=4.90）的 80.59%。图 1 显示了这 20 名当事人的平均 SG，以及前 3 次会谈和后 3 次会谈的平均值。

2.2 SG 的复发

参照前人研究的标准（Tang & DeRubeis, 1999），当事人的症状水平增加值超过 SG 的 50%，视为复发。例如一个当事人的 SG 为 6，此时症状水平为 4，若之后的会谈中症状水平增加到 7，则为复发。根据此标准，20 位经历 SG 的当事人中有 3（15%）位经历了复发。

2.3 SG 与效果

为排除初始症状水平和会谈次数对效果造成的影响，以初次会谈的 OQ-5 分值和会谈次数为协变量，SG 是否发生为自变量，对咨询效果（最后一次会谈的 OQ-5 值）做协方差分析（ANCOVA），结果显示，初始症状水平的效应显著，$F_{(1,78)}=24.68$，$p<.0000$，$\eta^2=0.24$；会谈次数效应不显著，$F_{(1,78)}=1.35$，$p=0.249$，$\eta^2=0.02$；SG 的效应显著，$F_{(1,78)}=13.64$，$p<.000$，$\eta^2=0.15$。

根据 OQ-5 的 RCI 值，将康复标准定义为：当事人的总改变量大等于 5 分，且最后一次会谈 OQ-5 的值小等于 4 分。依据此标准，20 位经历 SG 的当事人中有 8 位当事人康复，占 40%，而未经历此模式的当事人（对照组）的康复比例为 16.13%，差异显著（$\chi^2= 5.03$，$p=0.02$）。由此可知，经历 SG 的当事人咨询效果显著好于对照组。

2.4 SG 的影响因素

2.4.1 初始症状水平

SG 组的初始症状水平 (M=12.45，SD=5.05) 与对照组的初始症状水平（M=7.89，SD=3.82）差异显著（$t(80)=-4.28,p<.000$），说明经历 SG 的当事人在求助前比对照组的症状更严重。

2.4.2 工作同盟

在整个咨询过程中，SG 组的平均工作同盟水平显著高于对照组。进一步比较各个维度的平均值，结果发现两组在目标任务、情感联结维度差异不显著，

表 2　SG 组与对照组的工作同盟水平的比较

	SG 组		对照组		t	df	p
	M	SD	M	SD			
工作同盟总分	50.21	6.52	46.54	7.06	−2.05	80	0.043
目标任务维度	16.44	2.29	15.50	2.36	−1.56	80	0.123
情感联结维度	17.42	2.56	16.16	2.60	−1.88	80	0.063
投入维度	16.35	2.22	14.88	2.55	−2.31	80	0.024

而在投入维度上差异显著（表 2）。

对关键会谈与 SG 发生后一次会谈的工作同盟水平进行配对样本 T 检验，结果表明，关键会谈的工作同盟水平（M=51.00,SD=6.47）与后一次会谈（M=49.95,SD=7.50）差异不显著（t(19)=0.89,p=0.387），工作同盟各维度的平均分差异也不显著。比较 SG 组的初次会谈（M=49.60,SD=6.74）与关键会谈（M=51.00,SD=6.47）的工作同盟水平，配对样本 T 检验的结果表明，差异不显著（t(19)=1.03,p=0.317），工作同盟各维度的平均分差异也不显著。

3　讨论

本研究以某市多所高校心理咨询中心的当事人为被试，检验了咨询过程中发生的 SG 现象，并比较 SG 组与对照组在初始症状水平和工作同盟上的差异；另外，为考察工作同盟与 SG 的相互关系，对 SG 发生的前后两次会谈以及初次会谈进行了比较。

结果表明，SG 现象存在于本土心理咨询过程中，且与咨询效果有显著的正向关联，这说明 SG 现象有跨文化的一致性。然而 SG 的平均发生时间（2.25）与复发率（15%）较国外略有差别（表 1），即发生时间更早，而复发率更低，这可能反映了本土的心理咨询与国外的一些不同特点。在咨询的早期阶段，咨询的特异性技术并没有大量使用，主要是共同因素发挥作用。在中国，尤其是高校，咨询师在当事人眼里，同时也是老师、长辈和指导者，因此当事人更容易抱有高期望和信任度，从而产生了类似安慰剂的效应，导致抑郁水平的急剧降低，由此这种期望与信任变得更为牢固，使得之后的咨询更为顺利，因此效果容易保持，复发率较低。另一个可能的解释

是，本研究样本的会谈次数（Mdn = 8）与国外的研究相比普遍较短，若将完整的咨询划分为几个阶段的话，那么会谈总次数小使得各阶段的时间也更短，这可能对 SG 产生的时间有影响。

影响因素方面，本研究的结果显示初始症状水平高的当事人更多地经历了 SG。根据 Howard, Lueger, Maling 和 Martinovich（1993）提出的心理咨询效果三阶段模型，咨询的第一阶段主要是减轻当事人的无望感，提高积极期望。对于问题较严重的当事人来说，减轻其无望感，增强获助希望和对自身问题的掌控感，可能会促使其抑郁水平出现急剧降低。另外，由于本身症状较严重，当事人的求助动机可能会更强烈，改变阻力相对更小，且有较大空间能够出现 SG。

在整个咨询过程中，SG 组的平均工作同盟水平显著高于对照组，即从总体上来看，SG 组的咨询中建立了更好的工作同盟。各维度的具体结果显示，目标任务和情感联结维度差异不显著，而在投入维度上差异显著。这表明经历 SG 的当事人在整个咨询过程中的投入程度更高，在咨询中表现得更为积极主动，认真专注，克服各种困难。一方面投入程度高加快了症状的减轻，而症状减轻促使了当事人更进一步的投入。

微观的比较发现，SG 发生前后两次会谈的工作同盟水平差异不显著，这与 Tang 和 DeRubeis（1999）的结果不一致，Tang 和 DeRubeis 认为在认知行为治疗中 SG 可以激发一个良性循环：认知改变导致症状减轻，促进了工作同盟的改善，而好的工作同盟又使认知改变更加顺利。然而由于本研究中涉及到的治疗方法多样，可能没有这样的循环，因此没有发现 SG 与工作同盟的正向关系。而 SG 组的初次会谈与关键

会谈的工作同盟水平的差异也不显著，这可能说明工作同盟的变化与效果的变化并不同步，因为工作同盟的各个成分是较稳定的，而作为效果指标的抑郁水平则更敏感、易变。

SG 是咨询过程中一种特殊的局部现象，关注这种局部现象能够为我们理解整体的咨询改变过程和研究其中的改变机制提供很有价值的信息。首先，SG 现象表明部分当事人的改变并非线性或逐渐地，而是具有急剧的不连续性，这对于剂量－效果模型（ Howard, Kopta, Krause, & Orlinsky, 1986 ）是一个修正。剂量－效果模型所描述的改变过程为效果与会谈次数的关系是一条负性增长曲线，即咨询效果随着会谈次数而增加，前八次改变最快，之后增长的速度逐渐减小。而 SG 现象却显示出咨询效果的改变是波动的、不连续的。其次，SG 的发生提示我们重要的因素正在起作用，因此关注 SG 前后治疗因素的变化有利于揭示改变机制。研究 SG 对临床实践也有一定的指导作用，比如咨询师要及时注意到症状突然好转的当事人，并使这种好转得以保持。

本研究还存在不足之处：当事人没有明确的诊断信息，这使得我们无法探讨 SG 的发生与当事人本身的问题之间的联系。今后的研究可以在这方面有所改进，对 SG 现象做进一步的检验，并探索其他可能的影响因素，如当事人的期望、生活事件等，还可以结合质的方法，对 SG 出现前后的会谈材料做分析，以此揭示改变机制。

4 结论

在本土心理咨询中有一定比例的当事人经历了 SG；SG 组与对照组在初始症状水平和平均工作同盟水平上差异显著，其中工作同盟的投入维度差异显著。

参考文献

胡姝婧 . (2008). *心理咨询会谈中的当事人行为与咨询效果的关系* . 硕士学位论文 . 华中师范大学 .

秦佑凤 , 胡姝婧 (2008). 心理咨询效果问卷（OQ-45）在中国部分人群的试用报告 . *中国临床心理学杂志* , *16*, 138–140.

于丽霞 . (2009). *心理咨询会谈间体验、工作同盟对咨询效果的影响* . 硕士学位论文 . 华中师范大学 .

朱旭 . (2010). *心理咨询中的工作同盟* . 博士学位论文 . 华中师范大学 .

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research and Practice, 16*, 252–260.

Busch, A. M., Kanter, J. W., & Landes, S. J. (2006). Sudden gains and outcome: A broader temporal analysis of cognitive therapy for depression. *Behavior Therapy, 37*, 61–68.

Gaynor, S. T., Weersing, V. R., Kolko, D. J., et al. (2003). The prevalence and impact of large sudden improvments during adolescent therapy for depression: A comparison across cognitive–behavioral, family, and supportive therapy. *Journal of Consulting and Clinical Psychology, 71*, 386–393.

Hardy, G. E., Cahill, J., & Stiles, W. B. (2005). Sudden gains in cognitive therapy for depression: A replication and extension. *Journal of Consulting and Clinical Psychology, 73*, 59–67.

Hatcher, R. L., & Gillaspy, J. A. (2006). Development and validation of a revised short version of the Working Alliance Inventory. *Psychotherapy Research, 16*, 12–25.

Hofmann, S. G., Schulz, S. M., Meuret, A. E., et al. (2006). Sudden gains during therapy of social phobia. *Journal of Consulting and Clinical Psychology, 74*, 687–697.

Howard, K. I., Kopta, S. M., Krause, M. S., & Orlinsky, D. E. (1986). The dose-effect relationship in psychotherapy. *American Psychologist, 41*, 159–164.

Howard, K. I., Lueger, R.J., Maling, M. S., & Martinovich, Z. (1993). A phase model of psychotherapy outcome:

causal mediation of change. *Journal of Consulting and Clinical Psychology, 61*, 678－85.

Kelly, M. A. R., Cyranowski, J. M., & Frank, E. (2007). Sudden gains in interpersonal psychotherapy for depression. *Behaviour Research and Therapy, 45*, 2563－2572.

Kelly, M. A. R., Roberts, J. E., & Bottonari, K. A. (2007). Non－treatment－related sudden gains in depression: The role of self－evaluation. *Behaviour Resaearch and Therapy, 45*, 737－747.

Lambert, M. J. & Hawkins, E. J. (2004). Measuring outcome in professional practice: Considerations in selecting and using brief outcome instruments. *Professional Psychology: Research and Practice, 35*, 492–499.

Lambert, M. J., Morton, J. J., Hatfield, D. (2004). *Administration and Scoring Manual for the Outcome Questionnaire－45.* Orem, UT: American Professional Credentialing Services.

Lambert, M. J., & Ogles, B. M. (2004). *The efficacy and effectiveness of psychotherapy.* In M. J. Lambert (Ed.), Bergin and Garfield's handbook of psychotherapy and behavior change (5th ed., pp. 139–193). New York: Wiley.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta－analytic review. *Journal of Consulting and Clinical Psychology, 68*, 438－450.

Present, J., Crits－Christoph, P., Gibbons, M. B. C., et al.

(2008). Sudden gains in the treatment of generalized anxiety disorder. *Journal of Clinical Psychology, 64*, 119－126.

Stiles, W. B., Leach, C., Barkham, M., Lucock, M., Iveson, S., Shapiro, et al. (2003). Early sudden gains in psychotherapy under routine clinic conditions: Paractice－based evidence. *Journal of Consulting and Clinical Psychology, 71*, 14－21.

Stulz, N., & Lutz, W. (2007). Multidimensional patterns of change in outpatient psychotherapy: the phase model revisited. *Journal of Clinical Psychology, 63*, 817－833.

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive－behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*, 894－904.

Tang, T. Z., DeRubeis, R. J., Hollon, S. D., et al. (2007). Sudden gains in cognitive therapy of depression and depression relapse/recurrence. *Journal of Consulting and Clinical Psychology, 75*, 404－408.

Tang, T. Z., Luborsky, L., & Andrusyna, T. (2002). Sudden gains in recovering from depression: Are they also found in psychotherapies other than cognitive－behavioral therapy? *Journal of Consulting and Clinical Psychology, 70*, 444－447.

Vittengl, J. R., Clark, L. A., & Jarrett, R. B. (2005). Validity of sudden gains in acute phase treatment of depression. *Journal of Consulting and Clinical Psychology, 73,* 173－182.

Sudden gains in psychotherapy: The influences of working alliance and initial symptom severity

ZHU Wenzhen, JIANG Guangrong, ZHU Xu, TAO Rong

(school of psychology,huazhong normal university ,hubei province key laboratory of human beings development and mental health,wuhan 430079)

Abstract   Sudden gains—sudden and large decreases in depression severity in a single between—session interval—were Z.tang and R.J.Derubeis(1999).This study aimed at investigating the occurrence,resersal rate and the relationship between suddengains and outcomes.Then it wastested wheter the initial symptom severity and working alliance would be associated with sudeen gains. 82 clients were involved in this longitudinal study.The depression level of theclients was measured by 5 items related to the depression in OQ—45 beforeevery session. The working alliance was measured by the WAQ（working alliance questionnaire）,which was developed by zhuxu (2010)based on a qualitativestudy on Chinese clients.To define the sudden gains ,the authors used three criterions as follows: (1) the gain was at least 5 OQ—5 score of the three therapy sessions after the gain.（2） the gain represented at least 25% of the pregain sessions OQ—5score;(3) the mean OQ—5score of three therapies before the gain was significantlyhigher than the mean OQ—5score of three therapy sessions after the gain. The results:There were 20 clients (24.93%)who experienced at least onesudden gain. The average magnitude was 6.85 OQ—5 points,which accounted for 80.59% of the total change .The mean session of occurrence was 2.25.three out of 20(15%) clients experienced a reversal before the end of therapy.The sudden gainers had significantly better outcome compared with the other clients (F(1.78)=13.64,P<.000).In addition, the recovery rate of the clients who experienced the sudden gains was 16.13%,which was significantly lower than the 40% recovery rate of the clients who experienced sudden gains ,(X^2=5.03，p=0.02).Furthermore ,the results showed that the intial symptom severity of the sudden gainers was significantly higher than that of the non—sudden gainers(t=−4.28,p<.000).Also, the average level of working alliance through all of the session of the sudden gainers was significantly higher than the other clients (t=−2.05,p=0.043) and the effect of commiontment (one of the dimensions of the working alliance) was significant (t=−2.31,p=0.024),which indicated that the sudden gainers had a higher commitment level in the therapies.Besides,there was no significant difference in working alliance either between the first sessions and the pregain sessions, or between the pregain sessions and the postgain sessions. These results suggested that(1)t the sudden gains did exist in Chinese clients,with carlier session of occurrence and lower reversal rate than the western countries ; (2) sudden gainers got better outcome than the other clients;(3) working alliance and initial symptom severity had significantly influences on the occurrence of sudden gains. Sofar ,this ,as the first study, investigated the sudden gains in Chinese clients.And this study confirmed the influence of the working alliance and initial symptom severity on sudden gains.

Keywords   sudden gain; outcome; working alliance

心理咨询中工作同盟的发展模式与咨询效果 *

朱　旭，胡　岳，江光荣

(青少年网络心理与行为教育部重点实验室 , 华中师范大学心理学院 , 湖北省人的发展与心理健康重点实验室 , 武汉 430079)

摘　要　为了探索工作同盟在咨询过程中的发展模式 , 采用 Stiles 等 (2004) 提出的 4 个变化特征参数 (shape–of–change parameters) 对 30 个个案的工作同盟在咨询过程中的发展变化进行描述。聚类分析的结果显示有 3 类工作同盟的发展模式 , 分别是线性增长、线性下降和二次增长模式。不过 , 3 种发展模式、4 个变化特征参数及同盟的破裂 – 修复片段对咨询效果均没有影响。比较咨询效果不同的当事人在工作同盟发展模式上的差异 , 发现同一个发展模式对于不同的个案可能有着不同的含义 , 咨询师在早期对同盟关系的处理与调控对咨询效果有着重要的影响。

关键词　工作同盟 ; 发展模式 ; 效果 ; 心理咨询

分类号　R395

1　引言

上世纪中叶 , 新的治疗方法层出不穷 , 让人眼花缭乱。在此背景下 , Bordin (1979) 呼吁研究者们应该关注那些不同疗法中共同起作用的因素 , 只有如此才能使得咨询领域的研究走向整合。他认为工作同盟 (working alliance) 就是这样一个存在于不同疗法中的共同治疗要素 , 工作同盟的强度决定治疗效果。在随后的数十年里 , 工作同盟成为心理治疗领域最受关注的研究变量 (Orlinsky, Ronnestad, & Willutzki, 2004)。虽然理论家们对工作同盟的定义还有一些分歧 , 但多数人会同意将工作同盟定义为咨询中当事人与咨询师之间合作关系的质量与强度 (Horvath & Bedi, 2002)。

目前 , 工作同盟与治疗效果之间的关系已被大量研究所证实。对 200 多个研究的元分析表明 , 在个体咨询中工作同盟和效果之间的相关为 r = 0.275 (Horvath, Del Re, Flückiger, & Symonds, 2011), 且不受各种调节变量的影响 (Flückiger, Del Re, Wampold, Symonds, & Horvath, 2012)。在儿童青少年治疗和夫妻家庭治疗中的元分析也得到了类似的结果 (Friedlander, Escudero, Heatherington, & Diamond, 2011; Shirk, Karver, & Brown, 2011)。

Horvath (2005) 认为工作同盟的研究大致可以分为两个阶段 ; 前一个阶段研究者的主要精力放在探索各种治疗情境中工作同盟与咨询效果的关系上 ; 第二个阶段研究者主要关注的是工作同盟本身在咨询中的作用、发展与管理 , 并提出下一步需要考察工作同盟随时间的变化。实际上 , 虽然工作同盟与咨询效果的关系已被研究所证实 , 但对工作同盟的本质及其如

何促成当事人的改变其实知之甚少 (Stevens, Muran, Safran, Gorman, & Winston, 2007)。例如，近来的研究表明，工作同盟的破裂修复过程与治疗效果密切相关 (Safran, Muran, & Eubanks-Carter, 2011)，探索心理治疗的改变机制需要对工作同盟的发展变化进行研究。

Gelso 和 Carter (1994) 对咨询关系中各成分随咨询进程的变化进行了假设，提出了工作同盟发展的"U"型 (高 – 低 – 高) 变化模式。他们认为特别是在疗程较短的治疗中，由于当事人对于治疗的乐观主义，开始阶段会有较高水平的工作同盟，随后当事人往往会有失望、沮丧等负面情绪，工作同盟的水平也随之下降，但在成功的治疗中，在最后阶段当事人的反应变得更加现实，其工作同盟的水平也会恢复到一个较高的水平。

Kivlighan 和 Shaughnessy (2000) 认 为 Gelso 和 Carter 的论述中包含着两个假设：首先，在工作同盟的发展变化中存在着不同的模式；其次，这些模式中有一些会和咨询效果有着更密切的联系。换句话说，相比于仅仅使用工作同盟的整体水平，工作同盟的发展模式对咨询效果应有更好的解释。他们采用聚类分析的方法探索工作同盟的不同发展模式，结果在两个样本中均发现了稳定的、线性增长的和二次曲线的 (高 – 低 – 高、"U"型) 三种工作同盟的发展模式，并且也证实了 Gelso 和 Carter (1994) 的假设，工作同盟二次曲线的变化模式与其他变化模式相比有更好的咨询效果。

但此后研究者对工作同盟的发展模式进行研究时却常常无法发现 U 型模式 (Kramer, de Roten, Beretta, Michel, & Despland, 2008)，或发现 U 型模式与效果并无相关 (Kramer, de Roten, Beretta, Michel, & Despland, 2009)。以致一些研究者认为在疗程较长的咨询中，不太可能出现整体的 U 型模式 (Stevens et al., 2007; Stiles et al., 2004)。与 U 型模式不同，Safran 等 (Safran, 1993; Safran, Crocker, McMain, & Murray, 1990; Safran & Muran, 1996) 提出了工作同盟的破裂 – 修复 (rupture-repair) 模型，认为在咨询过程中工作同盟是由一系列较短暂的破裂 – 修复片段组成。工作同盟的破裂体现了当事人的某种人际问题，而咨询师对同盟破裂的识别、处理与修复即是一个治疗过程。

Stiles 等 (2004) 采用变化特征参数 (shape-of-change parameters) 来描述工作同盟的发展模式。结果发现了 4 种工作同盟的变化模式，其中两种与 Kivlighan 和 Shaughnessy (2000) 的研究结果一致，但唯独没有发现二次曲线的变化模式，而且 4 种变化模式的咨询效果没有显著差异。Stiles 等进一步检查发现，虽然没有发现整体的"U"型模式，但那些经历了短暂工作同盟破裂 – 修复片段，他们称之为"V"型模式的当事人有着更好的咨询效果。此后，研究者们采用不同方法对工作同盟的破裂 – 修复模式进行了研究，一些研究者发现了同盟的破裂 – 修复模式与治疗效果之间的关系 (Strauss et al., 2006)，一些则没有 (Stevens et al., 2007)。Safran 等对 3 个研究的元分析结果表明，工作同盟的破裂 – 修复片段与治疗效果的相关 r = 0.24 (Safran et al., 2011)。

国内研究者对当事人眼里的工作同盟进行了质的研究 (朱旭，江光荣，2011b)，也有研究发现了工作同盟与咨询效果的关系 (胡姝婧，江光荣，2014)，工作同盟有线性增长的趋势 (牛宏伟，江光荣，郝嘉佳，2013)。目前对于工作同盟发展模式的研究还处于探索阶段，研究结果并不一致，对于不同的理论假设还需要更多的研究证据支持。在研究设计上，以往的研究为了便于比较工作同盟的发展模式，常常采用限时咨询，即事先约定疗程，如 4 次或 8 次会谈。然而，实际的咨询却常常无法事先预知何时结束，而且长短不一，由特定疗程的样本得出的研究结果能否代表一般的咨询情境需要进一步检验。本研究的目的是探索当事人在非限时的自然咨询条件下工作同盟的发展变化模式，同时检验不同发展模式与咨询效果的关系。

2　方法

2.1　研究对象

在 5 所大学的心理咨询中心开展纵向研究，共有 61 位咨询师的 469 位当事人对 1429 次咨询会谈进行

了有效评估。为了研究工作同盟的发展变化，选取已结案的、所有的会谈测量完整且疗程在 4 次或 4 次以上的个案。之所以要求 4 次或 4 次以上的会谈是因为只有如此才能对二次曲线的发展模式及其误差进行估计。符合标准的共有 30 个个案。

当事人 这 30 名当事人来自 4 所高校的心理咨询中心，其中 28 人 (93%) 为在校本科生或研究生，男性 10 人 (33%)，女性 20 人 (67%)。当事人的年龄为 18 至 31 岁，$M=22.6$，$SD=3.4$。疗程从 4 次至 11 次会谈不等，$M=6.53$，$SD=2.03$，$Mdn=6$。

咨询师 有 20 位咨询师参与研究，其中男性 3 人 (15%)，女性 17 人 (85%)，年龄 24 至 58 岁，$M=37.6$，$SD=12.8$。咨询经验从 1 年至 22 年，$M=6.75$，$SD=6.03$，$Mdn=5$。咨询师对自己的咨询取向进行 5 级评定，平均分由高到低分别为以人为中心 (4.21)、认知行为 (3.40) 及心理动力 (3.20) 取向。每个咨询师提供 1 至 3 个个案。

2.2 测量工具

工作同盟问卷 (Working Alliance Questionnaire, WAQ) 该问卷包括情感联结、目标任务和投入三个维度，共 12 个项目，每个维度 4 个项目，采用 Likert 五级记分 (1= 很少，3= 经常，5= 总是)。该问卷有较好的结构效度和预测效度，各维度的内部一致性均在 0.70 以上 (朱旭，江光荣，2011a)。

咨询效果问卷简版 (Outcome Questionnaire-5, OQ-5) OQ 是由 Lambert 等 (2004) 开发的国外广泛使用的咨询效果的评估工具，测量当事人的症状困扰、人际关系和社会角色三方面的内容，共 45 题，故也称 OQ-45。OQ-45 采用 Likert 五级记分，得分越高，心理问题越严重。考虑到 OQ-45 中许多项目中国的当事人少有发生，且重复测量当事人负担较重，于是根据以往的研究，在 OQ-45 中选取了 5 道与总分相关最高的项目，组成 OQ-5。这 5 题是原问卷中的第 9、10、15、23、42 题，均为症状困扰维度的项目，且主要测量的是抑郁症状。在以往的研究中，OQ-5 与 OQ-45 总分的相关为 0.89 至 0.90，内部一致性为 0.82 至 0.87 (秦佑凤，胡姝婧，2008; 于丽霞，郑晓边，夏勉，

Hartmann, & Orlinsky, 2013)。

2.3 研究程序

在各咨询中心由接待员邀请当事人参与研究，同意参与研究的当事人签署知情同意书，每次咨询会谈结束后填写 WAQ 和 OQ-5，直到结案。当事人被保证所填问卷不会被咨询师看到，不会影响到咨询。当事人会收到一些小礼物作为对参与研究的感谢。

3　结果

3.1 变化特征参数

本研究的目的是发现工作同盟是否存在不同的变化模式，常用的纵向数据的处理方法，如多层线性模型等，假定总体的发展模式是同质的，无法区分出各种可能存在的亚模式。Borgen 和 Barnett (1987) 建议将聚类分析的方法引入咨询过程的研究，Kivlighan 和 Shaughnessy (2000) 随后将聚类分析的方法应用于工作同盟发展模式的研究中。不过 Stiles 等 (2004) 认为将每次会谈作为一个聚类分析的变量，在会谈次数较少时是可以的，一旦会谈次数较多，假定每次会谈均代表不同的咨询发展阶段显然不符合实际，而且聚类分析也不适合处理变量太多的情况。另外，增长混合模型 (growth mixture models, GMM) (Muthén & Muthén, 2007) 虽然能对子总体的不同增长趋势进行分析，但和聚类分析一样都要求测量次数相等。而自然条件下的咨询长短不一，因此这两种方法均不符所需。

为了更符合实际咨询情境，本研究尝试采用 Stiles 等 (2004) 提出的 4 个参数来描绘工作同盟的变化特征 (shape-of-change)。这个方法的基本思路是对每个个案都构建一个回归方程，然后用回归方程的 4 个参数来描绘这个个案的工作同盟随时间的变化特征。在本研究中，构建的回归方程如下：

$$WAQ = intercept + slope\ (CS) + curve\ (CS^2) + residual$$

因变量 (WAQ) 为每次会谈工作同盟的得分，自变量 (CS) 为中心化处理的会谈次数。会谈次数中心化处理后，截距 (intercept) 实际上表示的是咨询中点

时的工作同盟水平。斜率 (slope) 代表工作同盟的线性变化趋势。二次项系数 (curve) 表示工作同盟的二次变化趋势。残差 (residual) 是方程的误差，这些误差的均方根误差 (RMSE) 反映了未被方程所解释的变异，也可以理解为变化曲线的标准差，即围绕着方程所确定的曲线的变异程度。

因此，不管个案的疗程长短，当回归方程构建完毕后，都可以用 4 个参数来描绘这个个案工作同盟的变化特征：截距 (intercept) 表示的是工作同盟的水平，斜率 (slope) 和二次曲线 (curve) 代表工作同盟的变化趋势，而均方根误差 (RMSE) 则反映工作同盟的变异程度。

3.2 聚类分析

本研究采用聚类分析的方法来探索可能存在的不同的工作同盟的发展模式。由于关注的是同盟的发展模式，需要排除同盟水平的影响。因此，只对 30 个个案的斜率、二次曲线和均方根误差 3 个变量进行聚类分析，而不包括截距这个变量。根据 Borgen 和 Barnett (1987) 的建议，采用 Ward 离差平方和法进行层次聚类。聚合系数的变化和树状图 (见图 1) 显示可以分为 3 类。考虑到其中一类只有 2 个个案，而且许多学者建议应该采用多种不同的聚类方法以检验所得结果的稳定性 (郭志刚，1999)，于是采用迭代聚类法 (k-means cluster)，指定分类数为 3，重新进行聚类分析，所得分类与层次聚类的结果完全相同。而且，
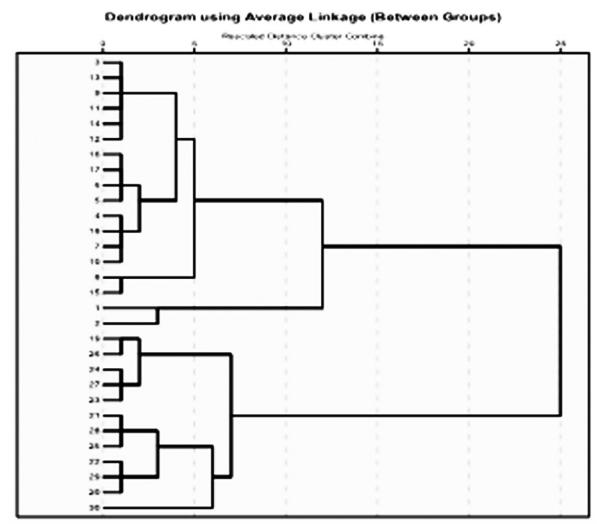
图 1 聚类分析的树状图

将单独聚成一类的两个个案删除，对其余个案分别采用层次聚类及迭代聚类法，个案所属分类均无变化。这说明聚类分析所得结果非常稳定，所以最终决定保留 3 类的分类结果。

各类变化参数的平均值及标准差见表 1，虽然截距没有参与聚类分析，也一同呈现在此。第 1 类的特点是斜率为正，二次曲线明显，变异较大，符合"U"型变化模式的特征，将其命名为二次增长模式，7% 的个案属于此类。第 2 类的特点是斜率为正，几乎没有曲线变化，变异较小，将其命名为线性增长模式，53% 的个案属于此类。第 3 类特征是斜率为负，没有曲线变化，有一定程度的变异，将其命名为线性下降模式，40% 的个案属于此类。当事人的性别、年龄和疗程在 3 类发展模式中均无显著差异。图 2 以 1 个 5 次会谈的咨询为例，用表 1 中的参数构建回归方程，展示 3 种工作同盟的发展模式。由于模型的参数均是以中心化的会谈次数为自变量而构建，因此选择不同的会谈次数会使图形的两端延伸或缩短，但变化趋势不会改变。

表 1 三类变化模式各参数的平均数与标准差

参数	I 类	II 类	III 类
截距 (intercept)			
M	3.49	4.05	3.83
SD	0.74	0.75	0.43
斜率 (slope)			
M	0.30	0.10	- 0.12
SD	0.02	0.08	0.05
二次曲线 (curve)			
M	0.20	0.01	0.01
SD	0.10	0.04	0.11
变异 (RMSE)			
M	0.36	0.16	0.20
SD	0.20	0.09	0.12

注：第 I 类，n=2；第 II 类，n=16；第 III 类，n=12；RMSE= 均方根误差。

3.3 不同发展模式的咨询效果

为了比较 3 种工作同盟发展模式的咨询效果，首先需要计算一个效果指标。以结案时当事人的 OQ-5 水平作为因变量，以初始 OQ-5 水平和疗程作为自变量构建一个回归方程，以回归方程的预测值与结案时当事人实际 OQ-5 水平的差值作为效果指标。这样计
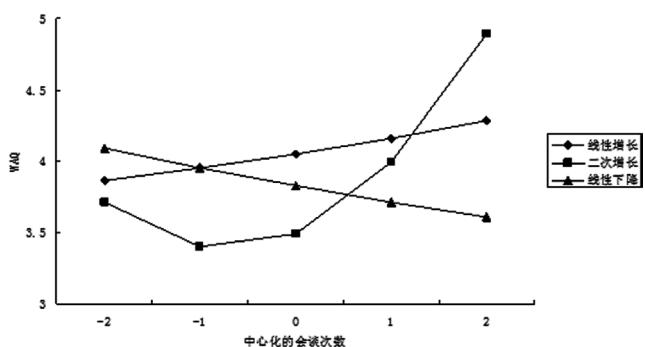
图 2　工作同盟的三种发展模式

算出的差值实际上是排除了初始症状水平及疗程长短对咨询效果的影响。

将 3 种工作同盟的发展模式作为自变量的 3 个水平，以个案所有会谈的工作同盟的平均水平作为协变量，以排除工作同盟水平对咨询效果的影响，以 OQ–5 的残差作为因变量，进行单因素的协方差分析 (ANCOVA)。结果显示，工作同盟平均水平的效应显著，$F(1,26)=5.15, p<0.05, \eta^2=0.17$；而工作同盟发展模式的效应不显著，$F(2,26)=2.09, p>0.10, \eta^2=0.14$。

3.4　变化参数与咨询效果

按照 Stiles 等 (2004) 的说法，直接将发展模式的几个参数与咨询效果求相关也可检验不同的假设。截距 (intercept) 不管在概念上还是在统计上都与工作同盟的平均水平非常相近，因此截距可以用来检验工作同盟的强度与咨询效果的关系。斜率 (slope) 反映的是工作同盟的线性变化趋势，可以用来检验工作同盟的线性增长是否可以改善症状水平。二次曲线 (curve) 代表工作同盟的二次变化趋势，可以用来检验 Gelso 和 Carter (1994) 提出的 "U" 型变化模式是否有更好的咨询效果。而均方根误差 (RMSE) 则是工作同盟的变异程度，可以反映工作同盟的稳定性是否与咨询效果相关。因此，将这 4 个参数与作为咨询效果指标的 OQ–5 的残差求相关以检验以上假设，结果相关均不显著，$-0.29 < r < -0.02, p>0.10$。

3.5　破裂 – 修复模式与咨询效果

Stiles 等 (2004) 认为工作同盟破裂 – 修复的过程可能很短暂，而且不同的当事人可能发生在不同时间，采用对变化参数进行聚类分析等方法可能无法发

现这些片段。于是，Stiles 等依据每次会谈工作同盟的得分以及变化特征的 4 个参数重新制定标准以甄别咨询过程中出现的破裂 – 修复片段。这些标准包括工作同盟实际得分低于变化参数所确定方程的预测值的 2 个 RMSE 以上，方程的斜率为非负，工作同盟的得分低于 6 (7 点计分) 等。在 Stiles 等的研究中有 22% 的当事人经历了同盟关系的破裂 – 修复。然而，依此标准 (因为本研究使用的 WAQ 是 5 级计分，所以不包括得分低于 6 这条标准)，在本研究中只有 1 个当事人出现过同盟关系的破裂 – 修复片段。

Stevens 等 (2007) 根据每次会谈工作同盟得分的变化来规定破裂 – 修复片段。标准包括破裂出现前工作同盟是稳定的，某次会谈工作同盟的得分下降 1 分以上，在此后的 3 到 5 次会谈内又恢复到不低于此前 0.25 分的水平 (7 点计分)。依此标准，Stevens 等的样本中有 50% 的当事人经历过同盟关系的破裂 – 修复片段。本研究中有 3 个 (10%) 当事人符合此标准。为比较有无破裂 – 修复片段当事人的咨询效果，同样将 WAQ 的均分作为协变量，以 OQ–5 的残差作为因变量，进行单因素的协方差分析 (ANCOVA)。结果显示，工作同盟平均水平的效应显著，$F(1,27)=5.79, p<0.05, \eta^2=0.18$；而破裂 – 修复片段的效应接近显著，$F(1,27)=4.10, p=0.053, \eta^2=0.13$。

在以往研究者提出的标准的基础上 (Strauss et al., 2006)，本研究还尝试以当事人工作同盟得分的标准差为单位来定义破裂 – 修复片段。不过，依据各种标准所区分的是否经历同盟关系破裂 – 修复片段的当事人在咨询效果上均没有显著差异。

3.6　不同咨询效果的同盟发展模式

到目前为止，本研究还没有发现工作同盟发展模式与咨询效果之间的关系。国外也有许多研究得到同样的结果，但却没有研究者给出研究证据来解释这一现象。为了探索这一现象的原因，本研究试着反过来看看不同咨询效果的同盟发展模式又是如何。因此，以 OQ–5 的残差作为效果指标，分别选取咨询效果最好和最差的三个个案，将这些个案每次会谈的 WAQ 得分依据其各自的平均数和标准差转化为标准分后

作图，以了解其工作同盟的变化趋势。

图 3 显示的是咨询效果最好的 3 个个案的同盟发展模式。第 I 个案在聚类分析中被归为线性增长模式，第 II 个个案被归为二次增长模式，第 III 个个案被归为线性下降模式。其中，第 III 个个案还符合 Stevens 等 (2007) 的破裂－修复片段的标准。从图上看，3 个个案最明显的一个共同点就是在第 2 次会谈时出现了某种程度上的同盟"破裂"，不过所有个案在随后的 1~2 次会谈中，同盟关系均被"修复"。而在后续咨询中，有的个案同盟关系继续加强，如个案 II；有的则保持稳定，如个案 I；还有的出现了下降，如个案 III。不过，这些个案最终都取得了好的咨询效果。
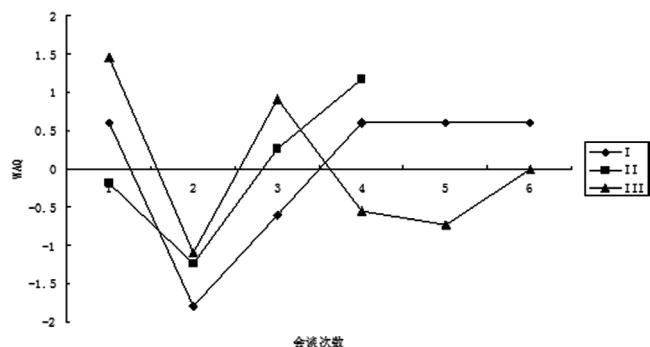
图 3 咨询效果最好的三个个案的同盟发展模式

图 4 显示的是咨询效果最差的三个个案的同盟发展模式。其中个案 I 和 III 在聚类分析中被归为线性增长模式，而个案 II 则被归为线性下降模式。由于个案 III 的会谈次数较多，同盟关系的变异也较大，在形式上有点类似于同盟关系的破裂－修复片段，不过这种波动并未达到同盟破裂－修复的各种标准。
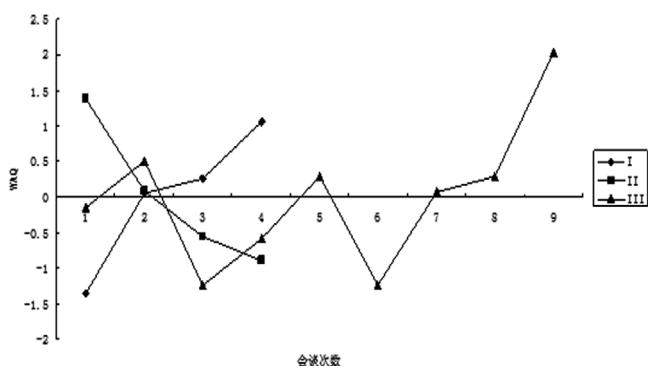
图 4 咨询效果最差的三个个案的同盟发展模式

4 讨论

本研究以自然情境中的咨询个案为研究对象，采用 Stiles 等 (2004) 提出的变化特征参数来描绘工作同盟的发展模式，通过聚类分析最终得到 3 类工作同盟的发展模式，分别是线性增长、线性下降和二次增长模式。

工作同盟的线性增长和线性下降模式在以往的研究中均有发现 (de Roten et al., 2004; Kivlighan & Shaughnessy, 2000; Kramer et al., 2008; Stevens et al., 2007; Stiles et al., 2004)，而二次增长模式虽然在理论中被提及，但在研究中并不多见。事实上，Stevens 等 (2007) 认为在实际的咨询过程中一个整体的工作同盟发展的高－低－高模式根本不存在。Stiles 等 (2004) 也认为难以想象一个成功的咨询大部时间工作同盟的水平都处于谷底。这些研究者认为咨询过程中更为短暂的同盟破裂－修复片段才是更为常见的现象。Kivlighan 和 Shaughnessy (2000) 所发现的二次增长模式很可能与其研究的咨询过程太短有关 (4 次会谈)，因为在这么短的时间里实际上无法区分一个整体的高－低－高模式和一个特定的破裂－修复片段。虽然使用多种聚类方法，所得结果都非常一致，但本研究中的二次增长模式只有 2 个个案，而且恰好这 2 个个案的疗程都比较短 (1 个 4 次，1 个 5 次会谈)，使得对二次增长模式的解释要谨慎。虽然本研究没有发现其他研究经常得到的稳定的工作同盟的发展模式，不过由图 2 可以看出，其实 3 种发展模式的变化幅度并不大，主要集中在 WAQ 的一个计分单位内。

本研究发现，在控制了工作同盟的水平后，工作同盟的发展模式、变化特征参数及破裂－修复片断均对咨询效果没有影响，反而是工作同盟的水平与咨询效果之间的关系一再被证实。国外研究也得到了类似的结果 (Kramer et al., 2008; Stevens et al., 2007; Stiles et al., 2004)。对此结果学者们有一些猜测，但并无研究证据的支持。为此，本研究反过来检查了不同咨询效果个案的同盟发展情况，希望得到一些启发。

首先，没有发现同盟的发展模式与咨询效果之间关联的一个可能原因是同一发展模式对于不同的当事人可能有着不同的含义。例如，一个稳定的同盟对一些当事人来说可能意味着咨询中的良好互动，但也有可能是咨询没有挑战当事人的不良行为模式，或是当事人对咨询师的反应不现实（只有积极移情而无同盟）(Horvath & Luborsky, 1993, p. 568)。Bordin (1994) 认为应该对同盟建立后所出现的破裂与同盟还没有充分建立两种情况进行区分。Bordin 认为在当事人认可改变的目标及正在进行的治疗工作后，同盟的破裂才能更清楚地显示当事人的问题。本研究对不同咨询效果个案的同盟发展模式的检查结果支持这一假设，从图3和图4可以发现，同一发展模式会同时出现在咨询效果好和不好的个案中。

其次，没有发现同盟发展模式与咨询效果的关系也可能与所采用的同盟的定义和测量工具有关。Stevens 等 (2007) 认为采用一般的工作同盟测量工具来甄别破裂－修复片段并不理想，因为通常工作同盟测量工具的维度，如信任、目标任务和投入等都较为稳定，对短暂的关系波动不够敏感，而且以症状改善作为衡量关系破裂－修复的效果指标也缺乏针对性。从图2可以看出，几种同盟发展模式的变化幅度的确不大。在整体同盟水平已经较高的情况下，这些细微的波动当事人可能难以察觉。因此，今后的研究可能需要采用对同盟发展变化更加敏感的同盟的概念与测量工具，如 Safran 等根据咨访双方的相互协商来定义工作同盟 (Safran, 1993; Safran et al., 1990; Safran & Muran, 2000; Safran, Muran, Samstag, & Stevens, 2001)。

此外，还有一些因素可能会影响同盟的发展模式与效果之间的关系。例如，Piper 等 (Piper, Boroto, Joyce, McCallum, & Azim, 1995) 发现对于客体关系质量较差的当事人，增长的同盟模式有更好的效果，同盟的改变模式比同盟的平均水平更重要；而对于客体关系质量较好的当事人，则是同盟的水平，而非同盟的变化模式与效果相关。此外，当事人的人格问题、抑郁经历都会调节工作同盟与咨询效果之间的关

系 (Falkenstöm, Granstöm, & Holmqvist, 2013; Lorenzo-Luaces, DeRubeis, & Webb, 2014)。由于样本所限，本研究无法对咨询类型、当事人及咨询师变量对工作同盟与咨询效果的关系的影响做出进一步检验。疗程长短也是需要考虑的一个因素。从图4可以看出，疗程较长的个案更有可能出现同盟的波动。本研究多数个案疗程较短，这也可能是本研究中出现同盟破裂－修复片断的比例低于国外研究的一个原因。目前，对工作同盟发展模式的研究还受到研究方法的限制。在同盟发展过程中出现的这些变化、波动和破裂－修复片段往往发生时间不确定，而且相对短暂，从而对研究者提出了挑战。研究者需要找到更好的方法来反映同盟关系随时间的波动。

对于不同咨询效果个案的同盟发展变化的检查有点类似于个案研究，其结果也带给我们一些临床上的启示。由图3可以看出，咨询效果最好的3个个案不约而同都在第2次会谈时出现了某种程度的"破裂"，不过所有个案在随后的1~2次会谈中同盟关系均被"修复"。Stiles 等 (2004) 发现大部分的破裂－修复片段发生在咨询的早期。Horvath 和 Luborsky (1993) 认为前5次会谈是同盟建立的关键期，其中第3次会谈又特别关键。有趣的是3个个案的破裂－修复片段恰好都发生在此阶段，证明此阶段对同盟关系的处理的确会对咨询效果产生重要的影响。Gelso 和 Carter (1994) 也认为由于时间有限，同盟的早期形成在短程治疗中更为重要。同时，令人意外的是同盟的增长模式也出现在了咨询效果不好的个案中。这也提醒咨询师，不是越来越好的同盟关系就一定会导致好的咨询效果，有可能是治疗没有触及核心问题或有移情的干扰 (Horvath & Luborsky, 1993, p. 568)。

对于工作同盟发展变化的研究有助于我们更好理解咨询的改变机制，但寻找与理论相符的工作同盟发展模式并不容易。如果再考虑文化的因素，则更加复杂。目前此领域的研究还处于探索阶段，各种方法在不断尝试之中，没有充分的信息可以下过多的结论，每一个研究都在为此领域提供着新的证据。

参考文献

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. *Psychotherapy: Theory, Research & Practice, 16,* 252–260.

Bordin, E. S. (1994). Theory and research on the therapeutic working alliance: New directions. In A. O. Horvath & L. S. Greenberg (Eds.), The working alliance: *Theory, research, and practice* (pp. 13–37). New York: John Wiley & Sons.

Borgen, F. H., & Barnett, D. C. (1987). Applying cluster analysis in counseling psychology research. *Journal of Counseling Psychology, 34,* 456–468.

de Roten, Y., Fischer, M., Drapeau, M., Beretta, V., Kramer, U., Favre, N., & Despland, J. N. (2004). Is one assessment enough? Patterns of helping alliance development and outcome. *Clinical Psychology & Psychotherapy, 11,* 324–331.

Falkenstr–m, F., Granstr–m, F., & Holmqvist, R. (2013). Therapeutic alliance predicts symptomatic improvement session by session. *Journal of Counseling Psychology, 60,* 317–328.

Fl ü ckiger, C., Del Re, A. C., Wampold, B. E., Symonds, D., & Horvath, A. O. (2012). How central is the alliance in psychotherapy? A multilevel longitudinal meta–analysis. J*ournal of Counseling Psychology, 59,* 10–17.

Friedlander, M. L., Escudero, V., Heatherington, L., & Diamond, G. M. (2011). Alliance in couple and family therapy. *Psychotherapy, 48,* 25–33.

Gelso, C. J., & Carter, J. A. (1994). Components of the psychotherapy relationship: Their interaction and unfolding during treatment. *Journal of Counseling Psychology, 41,* 296–306.

Guo, Z. G. (Ed.). (1999). *Social statistics analysis method: SPSS softeware application.* Beijing, China: Renmin University of China Press.

郭志刚 (编). (1999). 社会统计分析方法——SPSS *软件应用* . 北京 : 中国人民大学出版社 .

Horvath, A. O. (2005). The therapeutic relationship: Research and theory. *Psychotherapy Research, 15,* 3–7.

Horvath, A. O., & Bedi, R. P. (2002). The alliance. In J. C. Norcross (Ed.), Psychotherapy relationships that work: *Therapist contributions and responsiveness to patients* (pp. 37–69). New York: Oxford University Press.

Horvath, A. O., Del Re, A. C., Fl ü ckiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy, 48,* 9–16.

Horvath, A. O., & Luborsky, L. (1993). The role of the therapeutic alliance in psychotherapy. *Journal of Consulting and Clinical Psychology, 61,* 561–573.

Hu, S. J., & Jiang, G. R. (2014). Relationship between working alliance, insight and outcome. *Journal of Psychological Science, 37,* 1491–1496.

[胡姝婧 , 江光荣 . (2014). 工作同盟和领悟对咨询效果的影响 . *心理科学 , 37,* 1491–1496.]

Kivlighan, D. M., Jr., & Shaughnessy, P. (2000). Patterns of working alliance development: A typology of client's working alliance ratings. *Journal of Counseling Psychology, 47,* 362–371.

Kramer, U., de Roten, Y., Beretta, V., Michel, L., & Despland, J.–N. (2008). Patient's and therapist's views of early alliance building in dynamic psychotherapy: Patterns and relation to outcome. *Journal of Counseling Psychology, 55,* 89–95.

Kramer, U., de Roten, Y., Beretta, V., Michel, L., & Despland, J.–N. (2009). Alliance patterns over the course of short–term dynamic psychotherapy: The shape of productive relationships. *Psychotherapy Research, 19,* 699–706.

Lambert, M. J., Morton, J. J., Hatfield, D., Harmon, C., Hamilton, S., Reid, R. C., ··· Burlingame, G. M. (2004). *Administration and scoring manual for the outcome questionnaire-45*. Orem, UT: American Professional Credentialing Services.

Lorenzo-Luaces, L., DeRubeis, R. J., & Webb, C. A. (2014). Client characteristics as moderators of the relation between the therapeutic alliance and outcome in cognitive therapy for depression. J*ournal of Consulting and Clinical Psychology, 82*, 368–373.

Muthén, L. K., & Muthén, B. O. (2007). *Mplus user´s guide* (5th ed.). Los Angeles, CA: Muthén & Muthén.

Niu, H. W., Jiang, G. R., & Hao, J. J. (2013). Development pattern of working alliance and impact of client attachment in psychological counseling. *Chinese Mental Health Journal, 27*, 513–517.

[牛宏伟 , 江光荣 , 郝嘉佳 . (2013). 心理咨询中当事人工作同盟的发展及与依恋的关系 . *中国心理卫生杂志* , 27, 513–517.]

Orlinsky, D. E., Ronnestad, M. H., & Willutzki, U. (2004). Fifty years of psychotherapy process-outcome research: Continuity and change. In M. J. Lambert (Ed.), *Bergin and Garfield´s handbook of psychotherapy and behavior change* (5th ed., pp. 307–389). New York: Wiley.

Piper, W. E., Boroto, D. R., Joyce, A. S., McCallum, M., & Azim, H. F. A. (1995). Pattern of alliance and outcome in short-term individual psychotherapy. *Psychotherapy: Theory, Research, Practice, Training, 32*, 639–647.

Qin, Y. F., & Hu, S. J. (2008). Usability report of Outcome Questionnaire-45 in part of Chinese sample. *Chinese Journal of Clinical Psychology*, 16, 138–140.

[秦佑凤 , 胡姝婧 . (2008). 心理咨询效果问卷 (OQ–45) 在中国部分人群的试用报告 . *中国临床心理学杂志* , 16, 138–140.]

Safran, J. D. (1993). Breaches in the therapeutic alliance: An arena for negotiating authentic relatedness. *Psychotherapy: Theory, Research, Practice, Training, 30*, 11–24.

Safran, J. D., Crocker, P., McMain, S., & Murray, P. (1990). Therapeutic alliance rupture as a therapy event for empirical investigation. Psychotherapy: *Theory, Research, Practice, Training, 27*, 154–165.

Safran, J. D., & Muran, J. C. (1996). The resolution of ruptures in the therapeutic alliance. *Journal of Consulting and Clinical Psychology, 64,* 447–458.

Safran, J. D., & Muran, J. C. (2000). Resolving therapeutic alliance ruptures: Diversity and integration. *Journal of Clinical Psychology, 56*, 233–243.

Safran, J. D., Muran, J. C., & Eubanks-Carter, C. (2011). Repairing alliance ruptures. *Psychotherapy, 48*, 80–87.

Safran, J. D., Muran, J. C., Samstag, L. W., & Stevens, C. (2001). Repairing alliance ruptures. Psychotherapy: Theory, Research, *Practice, Training, 38*, 406–412.

Shirk, S. R., Karver, M. S., & Brown, R. (2011). The alliance in child and adolescent psychotherapy. *Psychotherapy, 48*, 17–24.

Stevens, C. L., Muran, J. C., Safran, J. D., Gorman, B. S., & Winston, A. (2007). Levels and patterns of the therapeutic alliance in brief psychotherapy. *American Journal of Psychotherapy, 61,* 109–129.

Stiles, W. B., Glick, M. J., Osatuke, K., Hardy, G. E., Shapiro, D. A., Agnew-Davies, R., ··· Barkham, M. (2004). Patterns of alliance development and the rupture-repair hypothesis: Are productive relationships U-Shaped or V-Shaped? *Journal of Counseling Psychology, 51*, 81–92.

Strauss, J. L., Hayes, A. M., Johnson, S. L., Newman, C. F., Brown, G. K., Barber, J. P., ··· Beck, A. T. (2006). Early alliance, alliance ruptures, and symptom change in a nonrandomized trial of cognitive therapy for avoidant and obsessive-compulsive personality disorders.

Journal of Consulting and Clinical Psychology, 74, 337–345.

Yu, L. X., Zheng, X. B., Xia, M., Hartmann, A., & Orlinsky, D. E. (2013). Intersession experiences and treatment outcome in shortterm psychotherapy: Evidence from longitudinal data. *Journal of Psychological Science, 36,* 1216–1222.

[于丽霞 , 郑晓边 , 夏勉 , Hartmann, A., & Orlinsky, D. E. (2013). 短程心理咨询的会谈间体验对咨询效果的影响——来自纵向测量的证据 . *心理科学, 36,* 1216–1222.]

Zhu, X., & Jiang, G. R. (2011a). Development of the working alliance questionnaire. *Chinese Journal of Clinical Psychology, 19,* 449–454.

[朱旭 , 江光荣 . (2011a). 工作同盟问卷的编制 . *中国临床心理学杂志 , 19,* 449–454.]

Zhu, X., & Jiang, G. R. (2011b). The working alliance in clients' eyes: A qualitative analysis. *Acta Psychologica Sinica, 43,* 420–431.

[朱旭 , 江光荣 . (2011b). 当事人眼里的工作同盟 : 质的分析 . *心理学报 , 43,* 420–431.]

The developmental patterns of working alliance in counseling: Relationships to therapeutic outcomes

ZHU Xu; HU Yue; JIANG Guangrong

(Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education; School of Psychology, Central China Normal University; Key Laboratory of Human Development and Mental Health of Hubei Province, Wuhan 430079, China)

Abstract　Working alliance in counseling has been a highly attended research area for several decades. Although a positive relationship between working alliance and counseling outcome has been well established, little do we know how working alliance functions in counseling process. Studying the development of working alliance would contribute to this understanding. Some theories have focused on describing developmental patterns of working alliance based on existing research findings, such as the U−shaped model and the rupture−repair episodes model. However, empirical effort has failed to show significant effect of these hypothesized developmental patterns on counseling outcomes. This naturalistic study aimed at exploring the developmental patterns of working alliance as emerged in counseling sessions and examining the relationship between these patterns and therapeutic outcomes. The participants included 30 clients from 4 university counseling centers, 10 males and 20 females, with the number of sessions ranging from 4 to11, M=6.53, SD=2.03, Mdn=6. By the time of data collection, all of the clients had finished their therapies. There were 20 therapists in the study, 3 males and 17 females, with professional experience from 1 to 22 years. They rated their therapeutic orientations using a 5−point Likert scale, which resulted in the strength of endorsement, from high to low, on person−centered, cognitive−behavioral and psychodynamic approaches. Each therapist offered 1 to 3 cases respectively. Clients filled out measurements of working alliance and depression symptoms after each session. Four shape−of−change parameters (Stiles et al., 2004) were used to describe the developmental patterns of working alliance in the counseling process of these 30 cases. Results from the cluster analysis revealed three developmental patterns, labeled as linear increase, linear decrease, and quadratic increase. However, none of these developmental patterns had significant effects on the counseling outcomes. Then, correlations between the four shape−of−change parameters (i.e., indicators of developmental patterns) and outcomes were tested, but none of the correlations was significant either. It was also found that the rupture−repair episodes defined in terms of various criteria could not differentiate good outcomes from poor ones. Notably, the levels of working alliance were still found to predict outcomes. In order to explore the reasons why the developmental patterns had no relationship with outcomes, we conducted case studies comparing the working alliance developmental patterns in cases with good or poor outcomes. Results showed that the same developmental patterns emerged in both types of cases, but these pattern may have different meanings for different clients. It appeared that therapists' regulations of working alliance in early sessions may have great influence on therapeutic outcomes.

Keywords　working alliance; developmental patterns; outcomes; counseling

当事人眼里的工作同盟：质的分析

朱　旭，江光荣

（华中师范大学心理学院，湖北省人的发展与心理健康重点实验室，武汉 430079）

摘　要　为了解中国当事人如何看待咨询中的工作同盟，对工作同盟做初步的本土概念化，采用协商一致的质的研究方法对来自 1 所大学心理咨询中心的 20 名当事人的访谈结果进行了分析。结果发现，与工作同盟相关的域有 6 个，分别是情感联结、任务、投入、合作模式、发展变化、影响因素。前三个域可以看作是工作同盟的构成要素，而后三个域则是对其外部特征的描述。对每个域的含义及其与现有理论和研究的关系进行了讨论。

关键词　咨询关系；工作同盟；质的分析

分类号　R395

1　前言

工作同盟成为近三十年来治疗研究中最受关注的变量 (Orlinsky, Ronnestad, & Willutzki, 2004)。原因可能有三，一是工作同盟的概念虽然源于精神分析，但现在已被看作是各种疗法所共有的治疗要素 (Bordin, 1979; Gelso & Carter, 1985; Horvath & Bedi, 2002)。二是工作同盟与治疗效果之间的联系被大量研究所证实，效果量虽然不大（r = 0.21 至0.26），但却非常稳定 (Horvath & Bedi, 2002; Horvath & Symonds, 1991; Martin, Garske, & Davis, 2000)。最后，相对于以往咨询理论对咨询师或当事人单方面因素的关注 (e.g., Rogers, 1957; Strong, 1968)，工作同盟强调咨询师与当事人相互合作的概念为理解咨询关系提供了一个更为全面的视角。

早期，Freud 认为积极移情中有一部分是有意识的、可以接受的（unobjectionable positive transference），可以利用这部分移情帮助分析治疗取得成功 (1912/1958, p. 105)。后来，Freud 提到分析师和当事人相对正常的自我结盟而对抗其自体中不受控制部分的必要性 (1937/1964, p. 235)。Sterba (1934) 提出了"自我同盟（ego alliance）"的概念，认为分析师与当事人自我中的理性部分的同盟可以使当事人能够自我观察，从而从分析师的解释中获益。Zetzel (1956) 在对移情进行讨论时使用"治疗同盟（therapeutic alliance）"的概念来指当事人自我中健康的部分与分析师联合而完成治疗任务的能力。Greenson (1965, 1967) 首先使用了"工作同盟（working alliance）"一词，并使其成为精神分析中广泛使用的概念。Greenson 将工作同盟定义为当事人与分析师之间相对非神经症的、理性的、中立的关系，这种关系使得当事人有可能与分析师合作，并在分析情境中有目的地工作。工作同盟既包括当事人对治疗师的情绪感受，也包括当事人的动机与工作能力。

Bordin (1979) 整合不同的理论观点，认为工作同

盟是咨询师与当事人之间一种相互协商、相互合作的治疗关系，包括三个成分，即咨询师与当事人对咨询目标的一致看法（goal agreement），对如何实现目标所涉及的一系列任务达成共识（task agreement）及相互之间的情感联结（bond）。Bordin 的工作同盟概念强调咨询师与当事人之间的相互性与合作，而非单方面的作用。而且，Bordin 明确指出工作同盟是存在于所有疗法之中的共同要素，工作同盟的强度决定了治疗效果。Bordin 的工作同盟理论一经提出就受到了广泛关注，激起了大量有关工作同盟的研究。

虽然工作同盟的研究已经持续了几十年，理论家们对工作同盟的概念仍存在着诸多分歧。工作同盟与其他关系成分的区分，工作同盟与治疗技术的关系等都是争论的焦点。虽然理论家们对工作同盟的定义各有侧重，不过多数人认为咨询师与当事人之间有目的、有意识的合作是工作同盟的核心 (Hatcher & Barends, 1996; Horvath & Bedi, 2002)。至于这种合作具体包括哪些成分，或者说是工作同盟的结构如何，则不同的学者有不同的看法。其中，影响最大的是前面介绍的 Bordin 关于工作同盟三成分的论述。

从工作同盟概念的发展过程可以看出，目前关于工作同盟的知识主要来自理论家，代表的是专业人士或咨询师对合作过程的看法，而对于合作的另一方——当事人的观点知之甚少。在研究中，实际上是假定当事人与咨询师的看法相同，用同样的维度去评价咨询中的合作过程。例如，根据 Bordin 的理论假设编制的《工作同盟问卷》（Working Alliance Inventory，WAI）(Horvath & Greenberg, 1989) 的当事人版本与咨询师版本具有相同的结构，只是项目表述上有所差异。但是，这一假定受到了挑战。元分析的结果表明咨询师与当事人对工作同盟评价的相关并不高（$r = 0.36$）(Tryon, Blackwell, & Hammel, 2007)，说明当事人与咨询师对工作同盟的看法很可能存在差异。Bedi (2006) 认为研究者是在以专家的概念框架来理解当事人，而许多在当事人看来重要的同盟要素可能没有被触及。

少数几个研究采用质的方法探索了当事人对工作同盟的看法。Bachelor (1995) 对 34 位当事人在咨询的三个阶段提供的 66 个对工作同盟的描述进行了现象学分析。结果发现有三类不同的同盟，一类是抚育型（nurturant）（46%），强调的是咨询师的助长态度；一类是领悟型（insight-oriented）（39%），看重增进自我理解；一类是合作型（collaborative）（15%），体现为当事人的积极参与。另外，Bedi 等 (Bedi, 2006; Bedi, Davis, & Williams, 2005) 分别使用关键事件技术（critical incident technique, CIT）和多变量概念绘图（multivariate concept-mapping, MVCM）的方法研究了当事人对咨询早期有助于建立同盟的外显言行活动的知觉。结果发现，咨询环境、身体语言、非言语姿势、情感支持、坦诚、确认、指导与挑战、教育、推荐相关材料、当事人的个人责任和会谈管理等均被当事人认为有助于同盟的建立。上述研究者认为所得结果与已有工作同盟的文献与理论并不吻合。例如，当事人提到的一些变量，如咨询师的友好、幽默、建议、当事人的自我理解等，现有文献并未充分说明；而被研究者看重的概念，如合作、相互性，在当事人那里则显得不是那么重要。

导致质的研究结果与理论不符的部分原因可能来自研究的目的与方法。Bedi 等关注的是有助于建立同盟的具体的、可观察的言行活动，所得结果自然会包括咨询环境、非言语反应等这些通常不被认为是同盟构成要素的内容，而合作与相互性这些较为抽象的元素则可能被分解为具体的事件。另外，上述研究者为了使当事人明白研究的内容，均使用"治疗关系"或"工作关系"来指代工作同盟。将工作同盟与咨询关系等同看待与通常对工作同盟的理解并不一致，这也可能导致了研究结果与理论的差异。

目前研究关注的都是当事人眼里良好的同盟关系，而对不良的同盟关系没有涉及。Bedi (2006) 认为好的同盟与不好的同盟并不一定是一个维度，即两者的区别可能不仅只是在一些因素上存在量的差别，而还可能有结构上的不同。显然，此假设对工作同盟的测量有着重要的意义，但要对此假设进行验证必须同时对积极和消极的工作同盟进行探索性的研究。

此外，当我们对源于西方的概念进行研究时，还应注意概念的对等性（construct equivalence），即这些概念在不同的文化中是否具有相同的涵义，如概念的结构与维度是否一致 (梁觉，周帆，2010)。虽然工作同盟在国外有较长的研究历史，但在国内却少有研究者涉及，采用质的方法对其进行本土的概念化是后续研究的基础。

鉴于以上原因，本研究将紧扣工作同盟的定义，即咨询中当事人与咨询师之间的合作关系，对中国当事人眼里的工作同盟进行质的研究，如合作关系的特点、发展变化和影响因素等。为此，我们选择协商一致的质的研究方法（Consensual Qualitative Research, CQR）(Hill et al., 2005; Hill, Thompson, & Williams, 1997) 来开展研究。在咨询的过程研究领域，CQR 是一种被广泛使用的质的研究方法 (Orlinsky, et al., 2004, p. 309)，其最大的特点是数据的分析是由一个小组来完成，所有的决定均需小组成员协商达成一致。这样既保证了对复杂现象进行研究时视角的多样性，又减少了单一研究者容易产生偏差的影响。

2 方法

2.1 参与者

2.1.1 当事人 来自一所大学心理咨询中心的 20 位当事人参与了研究。3 人为在校研究生，其余为本科生。年龄在 19 到 29 岁之间（M=22.05，SD=2.44）。3 人为男性。7 人已经结束咨询，结束的时间均在半年之内，其余当事人的咨询仍在继续。在参与研究时，咨询的会谈次数为 8 至 27 次（M=11.10，Mdn=9，SD=4.53）。结束咨询和未结束咨询的当事人在年龄和会谈次数上没有显著差异。其中 4 人的主诉问题为抑郁，其他均为发展性问题，主要包括人际关系、自我认识等。

2.1.2 研究小组 所有的访谈工作由 1 位心理咨询方向的博士生（本文的第一作者，男，30 岁）完成。此外，研究小组还包括 5 位心理咨询方向的硕士生（1 男 4 女，年龄均为 20 多岁）。部分成员参加了 Clara

Hill 举办的介绍 CQR 方法的工作坊（2009 年 3 月，武汉）。所有人被要求阅读有关介绍 CQR 方法的两篇文献及数篇使用 CQR 方法的研究文献。然后大家一起对 CQR 的研究方法和程序进行讨论（6 小时），特别强调了鼓励表达和建立平等的交流氛围的重要性。

在研究开始之前，结合访谈提纲，所有组员一起讨论了各自对咨询中的工作同盟的体验与看法，包括个人的偏见（bias）与期望（expectation）。大家均认为，与西方相比，将咨询关系看作是合作显得较为陌生，中国的当事人可能更倾向于将这种合作看作是对咨询师的配合。大家也一致同意情感因素在良好合作关系中的重要性。大家也承认，对工作同盟或合作关系的理解会受到相关理论的影响，如 Bordin 的工作同盟理论。在此提供这些信息，作为读者理解研究结果的背景。组员表达各自的看法后，被提醒尽可能搁置这些期望，而忠实数据本身。在此后的数据分析过程中，组员们继续交流各自对工作同盟的看法。

2.2 访谈提纲

访谈提纲的第一部分询问当事人的一些基本信息，如年龄、咨询是否结束、疗程及主诉问题等。第二部分询问当事人对咨询过程中合作或合作关系的看法，比如好的合作关系是个什么样子，哪些因素比较重要等，要求尽可能详细地说明和举例。如果当事人认为与咨询师的合作关系并不好，则让其对自己的经历进行描述。第三部分让当事人描述合作关系的变化过程。最后，让当事人对其与咨询师之间的合作关系做一个 5 级的评价（1 = 非常不好，3 = 一般，5 = 非常好）。访谈提纲为半结构式，对每个当事人访谈的内容基本一致，但具体问法则根据当事人的回答进行调整。在访谈提纲的开发过程中，对 3 位有过当事人经历的硕士研究生及 3 位当事人进行了预访谈，根据反馈，对访谈提纲进行了反复修改。

2.3 数据收集

2.3.1 招募当事人 通过两种方式招募当事人。第一种方式是在一所大学咨询中心的接待室放置广告，对研究进行介绍，有兴趣参与的当事人留下联系方

式，随后研究者与其联系。结果发现有兴趣参与的当事人大多刚开始咨询，预访谈发现他们能够提供的信息有限。为了找到对咨询有更多了解的当事人，考虑采用第二种招募方式。在征得咨询中心负责人的批准后，通过该咨询中心的当事人信息数据库获取本学年咨询会谈在 8 次及 8 次以上当事人的名单及联系方式（包括结束咨询和未结束咨询的）。先向当事人发送短信简单介绍研究，然后在随后的一周里再一一进行电话确认。愿意接受访谈的当事人付给象征性的报酬（10 元）。事实上，许多参与研究的当事人并未领取报酬。选取当事人时特别排除了那些有可能与研究者认识的当事人，如同院系的学生。

2.3.2 访谈 所有的访谈由一名研究者完成，采用电话访谈的形式。先对研究目的及保密、录音等事项进行介绍，若当事人同意，则开始正式访谈。访谈大约持续半小时左右。访谈结束后对当事人表示感谢，邀请其对访谈过程提供一些反馈，如有哪些地方不明白或可以改进。询问当事人是否愿意留下邮箱地址，以便可以接收访谈逐字稿和研究结果。最后访谈者记下此次访谈的备忘，如访谈的感觉等，作为后续分析材料的背景信息。

2.3.3 转录 访谈录音由几位组员转录为逐字稿。其后访谈者又对照录音对所有逐字稿进行了核对，以保证转录的准确性，并将所有身份信息移除。每个个案使用一个代号，并将访谈备忘附在每个个案的后面，作为背景信息。访谈逐字稿约为 10 万字。

2.4 数据分析

采用协商一致的质的研究方法（Consensual Qualitative Research, CQR）(Hill, et al., 2005; Hill, et al., 1997) 对数据进行分析。按照 Hill 等 (2005) 的说法，CQR 主要是建构主义取向，在其中加入了一些后实证主义的元素。其分析方法大致可以分为以下三步：首先，将文本中所有与合作相关的信息划分为几个域（domain，即话题范围）；然后，将每个个案中各个域的信息精练成核心观点（core idea，相当于写摘要，尽量忠实原意）；最后，将所有个案中同一个域的核心观点放在一起做交叉分析（cross analysis），

找出其中的共同主题，聚成不同的类别（category），形成研究结果。所有的数据分析由一个研究小组来完成，整个分析过程均需小组成员共同协商，达成一致，并且邀请研究小组之外的人来做审核员（auditor），以避免整个小组出现重大偏差。在本研究中，每次讨论至少有 4 名研究小组成员参加，整个数据分析过程花费 160 余小时。

2.4.1 域编码 按访谈提纲和逐字稿的内容将所有与合作相关的内容划分为几个域（即话题范围）。小组成员先各自将每个个案中的意义单元划分到一个或多个域中，然后大家一起讨论，直到所有材料的划分达成一致。在分析的过程中，由于新信息的出现或对材料理解的加深，可能需要对原来的域进行修改，如合并或增加新的域。最后，按照最终确定的域对所有数据重新划分一遍，使得前后编码一致。

2.4.2 提炼核心观点 每个组员先各自对每个个案每个域中的内容进行概括、提炼，写成核心观点。核心观点尽可能用当事人的原话，保持原意，其形式往往是形成几个不同含义的句子。然后所有组员对每个核心观点的内容及文字表述进行讨论，直到达成一致。

2.4.3 交叉分析 首先对 18 个个案的数据进行交叉分析（留下两个个案做稳定性检查）。每个组员各自对所有个案同一个域中的核心观点进行检查，提出分类。然后所有组员一起对分类及其名称，还有每个核心观点的归属进行讨论，直到全部达成一致。

2.4.4 稳定性检查 剩余的两个个案被加入，检查初步交叉分析所得的类别是否也适用于新的个案，各个类别的代表性是否会发生变化。结果发现，新增个案并没有改变原有结果，因此所得的结果被认为是稳定的。

2.4.5 审核 审核分为内部审核和外部审核。在得出初步结果后，组员们又一起对照原始的逐字稿重新对每个个案进行回顾，以防止一些背景信息被忽略而影响对抽取出来的信息的理解（因为一些类别的划分是建立在对整个个案的理解之上的），同时对所得结果进行检查。当组员达成一致后，这一步的

结果被发给 4 位心理咨询方向的博士生，邀请其对结果进行审核。审核的内容包括是否还有其他划分框架，域和类别的名称是否合适，核心观点的归属是否准确等。研究小组对所有的反馈意见再次进行讨论，以决定是否修改结果。

2.4.6 当事人反馈 将每个当事人的访谈逐字稿及研究结果（域、类别、核心观点及相应说明）通过电子邮件发给当事人，让其对身份的保密性进行确认，并对研究结果提供一些反馈，为研究提供证明效度（testimonial validity）(Stiles, 1993)。在本研究中没有当事人给予回复。

2.4.7 类别的代表性 按照 Hill 等 (2005) 的标准，可以对结果中各个类别的代表性进行划分。若此类结果能代表所有个案或只有一个个案例外，在本研究中即为 19 或 20 个个案，则该类别被视作是普遍的（general）；如果能代表一半以上的个案，在本研究中为 11 至 18 个个案，则为典型的（typical）；若能代表 4 至 10 个个案则为变异的（variant）；代表 2 至 3 个个案为少有的（rare）。所有只包括一个个案的结果，放在"其他"类中，不予报告。

3 结果

在 20 名当事人中，对合作关系持积极评价的有 12 人（当事人自己的评分为 4、4.5 或 5 分），认为合作关系一般的有 6 人（评分为 3 或 3.5 分），认为合作关系不好的有 2 人（评分为 2 分）。结束咨询和未结束咨询的当事人在对合作关系的评价上没有差异。为了检验积极与消极的同盟关系是否在结构上存在不同，可以看不同性质的同盟关系能否使用同一个框架进行描述。结果发现，积极的和消极的同盟关系都可以用一些共同的类别来描述，两者正好从正反两面提供证据。和同盟相关的内容被分为情感联结、任务、投入、合作模式、发展变化和影响因素 6 个域。所有域、类及其下子类的结果呈现在表 1 中，表中还包括每个类别的代表性及核心观点举例。文中则按域来呈现结果，对每个域的含义及其中的类别（以

典型类别为主）进行介绍。在每个域中，先呈现积极的同盟关系，再呈现消极的同盟关系的结果。

3.1 情感联结

情感联结是咨询师与当事人之间的情绪感受。其中，当事人对咨询师的信任是一个典型的结果。如果当事人对咨询师很信任则会"在咨询师面前表露最真实的自己"，而"如果信任打破了对当事人则会有毁灭性的影响"。同盟关系出现问题的当事人也往往是由于没有对咨询师产生充分的信任，如"当事人很少信任别人，因为不想让咨询师太为难，才有什么话尽可能与咨询师说，但过几天回想会觉得不舒服，时间长了就觉得怯场，有种怕了咨询师的感觉"。其他被当事人认为良好的工作同盟所包含的情感还有坦率真诚、关注关心、轻松自然、接纳、依赖和理解。

3.2 任务

任务是合作过程中咨询师与当事人各自或双方要做的一些事情和活动，包括围绕着问题解决的倾诉 – 了解、探索 – 分析和指导 – 执行，还有对整个咨询的过程调控。在刚开始咨询时，当事人往往"很主动地说自己的情况、经历和思想状态"，认为"信息提供得越充分，感受表达得越真实，咨询师就越能在了解的基础上提出有帮助的建议"。此后，咨询师"具体分析问题，让当事人明白了一些以前没有意识到的问题，恍然大悟"，或是"在当事人说出一些困惑的东西时，咨询师及时给出有针对性的提示，能够说出当事人隐隐约约的感觉，让当事人意识到是哪方面出了问题"。而在有问题的同盟里，"都是当事人一个人在讲，咨询师没有针对当事人的问题做分析，提出一些剖析性的见解或建议，当事人不知道该怎么做，当事人要知道自己怎么了，然后才能自救"。最后，咨询师还会"给当事人建议、方法或布置一些作业"，而当事人则会"努力去做"、"尽力完成"。工作同盟不佳的当事人要么觉得"咨询师讲当事人知道的大道理，但不给具体的方法指引，这让当事人不知道如何做"，要么"刚开始时更听从咨询师的建议，到后来咨询师布置的任务有

表 1　当事人眼里的工作同盟：CQR 的结果

域、类及子类	代表性	核心观点举例
情感联结		
当事人的信任	T	
咨询师安全可靠	T	当事人信任咨询师，说问题时不会有保留。
咨询师的能力	R	当事人相信咨询师能帮助自己解决问题。
坦率真诚	V	当事人很坦诚，完全暴露自己的问题；咨询师与当事人的交流很坦率。
咨询师的关注关心	V	咨询师对当事人表示关注，发自内心地想去帮助当事人，而不是像完成任务一样。
轻松自然	V	当事人和咨询师像好朋友，很自然很放松。
咨询师的接纳	V	咨询师不会对当事人讲的内容进行批评，觉得当事人不对。
当事人的依赖	V	咨询师请假，当事人体验到了对咨询的依赖，觉得咨询是重要的心理支柱和情感出口。
咨询师的理解	V	当事人有被理解的感觉，咨询师非常容易听懂当事人说什么，能身临其境地感受当事人的感受。
任务		
倾诉 – 了解	T	
信息收集	T	当事人提供一些有用的信息给咨询师。
情感宣泄	V	早期当事人急于倾诉，咨询师起到安抚作用。
探索 – 分析	T	
咨询师的分析与启发	T	咨询师对当事人的问题进行分析、总结，让当事人挺受启发，能看清自己。
当事人的反思与反馈	V	当事人会去思考、反省咨询师指出来的一些问题，并慢慢接受。
指导 – 执行	T	
咨询师指导	T	
具体建议	T	咨询师给当事人一些指导，如让当事人练习自信地讲话。
一般方法	V	咨询师教当事人一些做事情的方法，例如做选择之前分析各个选择的后果，让当事人感觉很贴心。
当事人执行	V	当事人按照咨询师提出的建议去做，例如去调节自己的心情，看有没有效果。
过程调控	T	
结构化	V	咨询师应该在咨询前做些介绍，增加当事人对咨询活动的认识，减少当事人表述自己时的顾虑。
目标计划	V	咨询师与当事人建立共同目标，然后讨论具体问题。
反馈调整	V	咨询师始终提醒当事人把自己当作咨询师这样一个角色来理解，不能依赖自己，后来当事人就反复提醒自己不能依赖咨询师。
投入		
积极主动	V	当事人愿意说，而且有意识地往点子上说，如果做不到，至少要勤开口，不要有太多的疑虑和隐瞒。
克服困难	V	当事人不想与外界联系，但当事人每周能坚持咨询，很努力地让自己从那个情绪当中走出来。
认真专注	V	当事人和咨询师要有认真的心态，当事人要将咨询当作重要的事情来做。
合作模式		
咨询师主导	V	好的合作就是咨询师对当事人进行指导，当事人去做，而不会有质疑。
相互合作	V	合作就是咨询师与当事人建立共同目标，然后共同努力。
当事人主导	R	咨询中当事人是主导，咨询师是辅助作用，咨询师帮助当事人探索，当事人要主动探索，并做一些努力和调整。
发展变化		
稳定	T	合作关系没什么变化。
上升	V	合作关系一开始就比较好，是一个上升的过程，但上升的幅度不大。
倒 U 型	R	当事人觉得开始时比较生疏，不愿意讲所有的事情；中间时最亲密、关系最好，什么都愿意讲；再到后来反而没什么可讲的，关系又拉远了。
影响因素		
咨询师的专业能力和经验	V	咨询师要有专业素养，有很多方法和技巧，有很硬的专业功底，有丰富的经验。
当事人的个人特点	V	当事人是一个感性的人，做事情凭感觉，如果对咨询师感觉好，会更愿意配合。
当事人的改变动机	V	当事人一直希望能够得到帮助，所以一直很配合。
当事人对咨询的期望	V	当事人对咨询不要抱太高期望，咨询师不能帮当事人解决所有问题。
当事人的咨询经历	V	当事人曾经做过心理咨询，所以知道咨询师不会给倾向性的内容，也知道不用怕咨询师，不用顾忌得罪咨询师，怎么想就怎么说。
效果	V	当事人对咨询师很满意，感觉咨询师对自己有帮助，当事人也很愿意配合咨询师，配合出来的结果对当事人也比较有帮助。
当事人对咨询的信念	R	当事人坚信心理咨询有效，相信咨询师的实力能帮助自己，即积极的心理作用。
当事人对咨询的接纳	R	当事人对咨询活动的认识比较客观，不认为自己是有病，就是对现实不太满意，做一个积极的调整，所以没有很大的戒备心理。
性别	R	咨询师与当事人都是女性，当事人能自如、随意地说。
匹配	R	咨询师与当事人感觉比较对，就可以讲很多。

注：N=20；T=typical，11–18 个个案；V=variant，4–10 个个案；R=rare，2–3 个个案。

时会拖过去不做"。当然，以上三个阶段的划分是相对的，事实上三者往往是交叉或循环进行的，如"当事人回去实践咨询师布置的任务，然后跟咨询师分享，咨询师再帮当事人分析"。

与此同时，咨询师还需对咨询的过程进行调控，以使各项任务活动能够顺利进行。比如，咨询师"在刚开始时引导当事人说要建立信任的关系，把咨询比喻成走迷宫的过程"，"反复强调良好关系的重要性，让当事人对咨询引起重视，帮当事人树立积极的态度，让当事人投入到咨询过程中"，与当事人"建立共同目标，然后讨论具体问题"。而身处不良同盟关系中的当事人则抱怨"咨询师没有给当事人任何明示或暗示，告诉当事人如何配合，所以当事人不知如何合作，如不知道在咨询中说什么"，或是"咨询师跟当事人说什么都要说，每次基本上都是当事人在说，当事人不明白咨询的主要问题，感觉糊涂"，而且"当事人跟咨询师提出一些自己的问题和想法，表达自己的不满，咨询师却没有做出相应的改变"。

3.3 投入

投入是合作关系中咨询师与当事人的付出程度，具体包括积极主动、克服困难和认真专注三个方面。双方投入才能保证合作最终产生成效，实现目标。一些当事人在咨询中会"特别积极主动地探索解决方式，多想多思考，表述自己特别真实的想法，形象地描述自己的感觉"，在咨询之外也会"刻意寻求别人的帮助，改变自己不好的生活习惯，充实自己的生活，搜集相关书籍和资料，来调节自己的情绪"。而且，"当事人刚开始的时候觉得进展不大，但当事人挺沉得住气，给自己积极的暗示，反复提醒自己要相信咨询师，应该继续去做没有做到的事情"。当然，"当事人和咨询师的主动和投入都很重要"，咨询师的投入主要体现为"咨询师认真负责"。不过，也有一些当事人"一直很被动，咨询的时候常有大段的沉默"，"平时很忙，练习咨询师教给当事人的方法次数不多"。

3.4 合作模式

合作模式是根据当事人对咨询师的依赖程度和在咨询过程中的参与程度来划分的，可以分为咨询

师主导、相互合作和当事人主导三种类型。一些当事人认为"咨询更多依靠咨询师，当事人就想着把自己的问题说出来让咨询师解决"，或者是"比较好的合作就是咨询师给很好的方法指导当事人，给建议让当事人去做"，当事人"希望咨询师帮自己解决问题，没有思考过怎样把咨询做得更好，很少想自己要在里面做什么，只会跟着咨询师的思路走"。而其他一些当事人则认为"合作就是咨询师与当事人建立共同目标，然后共同努力"，"双方都需要积极主动"。少数当事人认为"咨询中当事人是主导，咨询师起辅助作用，咨询师帮助当事人探索，当事人要主动探索，并做一些努力和调整"。当然，这三种合作模式还会随着咨询过程而有所变化，如"刚开始当事人向咨询师讲自己的问题，时间由当事人来掌握，合作关系的重点在咨询师，获取当事人的信赖；后来一起互动，时间由两人掌握，合作关系重点落到当事人身上，要当事人行动与思考"。

3.5 发展变化

同盟关系在整个咨询过程中典型的发展变化是保持稳定，如"合作关系没什么变化"；也有上升的，如"关系越来越熟悉，越来越亲切，越来越好，对咨询师的信任感逐步增加"；少数出现倒 U 型，如"当事人觉得开始时比较生疏，不愿意讲所有的事情；中间时最亲密、关系最好，什么都愿意讲；再到后来反而没什么可讲的，关系又拉远了"。

3.6 影响因素

对同盟关系有影响的因素包括咨询师的专业能力和经验，如"咨询师要有专业素养，有很多方法和技巧，有很硬的专业功底，有丰富的经验"；当事人的个人特点，如"当事人的性格，例如防备心和表达水平，会影响当事人在咨询中表述的充分程度"；当事人的改变动机，如"当事人将咨询作为最后的希望，需要得到咨询师的救助，非常想通过咨询帮助自己成长，愿意全力配合"；当事人对咨询的期望，如"当事人对咨询不要抱太高的期望，咨询师不能帮当事人解决所有问题"；当事人的咨询经历，如"当事人之前做过咨询，让当事人对咨询有足够的信任，

是后面求助咨询的基础"；效果，如"当事人对咨询师很满意，感觉咨询师对自己有帮助，当事人也很愿意配合咨询师，配合出来的结果对当事人也比较有帮助"。少数当事人还提到当事人对咨询的信念、当事人对咨询的接纳、性别和当事人与咨询师的匹配对同盟关系也会产生影响。

4　讨论

本研究通过质的方法对当事人眼里的同盟关系进行分析，得出了与工作同盟相关的 6 个域，分别是情感联结、任务、投入、合作模式、发展变化和影响因素。若将咨询看作是一个咨询师与当事人相互合作而达成目标的过程 (Hatcher & Barends, 2006)，则前三个域可以看作是这个合作过程的构成要素，而后三个域则是对其外部特性的描述。

4.1　情感联结、任务和投入

假如将合作作为工作同盟的核心内涵，则情感联结可以看作是合作的情感基础，即愿不愿意合作。在情感基础上，任务则是合作的内容，即如何合作。而为了保证合作能达到预期目标，双方的投入不可缺少。投入是合作的动力，即花多大力气合作。这三者构成了良好合作的三个成分。Hougaard (1994) 认为同盟关系包含两个方面的内容，一个是与社会情感相关的个人关系；一个是和任务相关的合作关系。显然，情感联结属于个人关系，而任务与投入则属于合作关系。

情感联结域与 Bordin 工作同盟理论中的情感联结因素基本对应。Bordin 指出各种治疗关系都需要基本水平的信任，而若要探索更加隐秘的内在体验时，则需要更深的信任关系 (1979, p. 254)。同样，情感联结域中最突出的也是当事人对咨询师的信任。情感联结域中的坦率真诚、关注关心、接纳和理解与 Rogers (1957) 强调的助长条件一致。这些助长条件既促使了当事人的改变，也促进了当事人与咨询师的合作，成为良好合作的情感基础。

咨询师与当事人在合作过程中需要完成的任务

大部分指向问题解决，也包括了对咨询过程进行调控的要求。许多当事人提到的咨询师的活动通常可以被归为治疗技术，如指导、结构化、建议、解释、倾听等。其中，当事人最为看重的是咨询师的指导，如建议、教育、布置任务等，反映了中国当事人对咨询师指导性及咨询过程结构化的要求。不过，在国外有关工作同盟的质的研究中，国外当事人也同样看重这些带有较强咨询师指导性的活动，如效果聚焦的技术和策略 (Bedi, et al., 2005)、教育 (Bedi, 2006)、咨询师指导当事人的能力、促进当事人理解的干预、建议 (Bachelor, 1995) 等。这说明工作同盟中的合作任务可能具有一定程度的文化普遍性。另外，Bordin (1979, 1994) 工作同盟理论中的任务类似于各个流派的治疗策略与技术，如自由联想、空椅子技术等，而本研究结果中的任务则不局限于某种特定治疗取向。

投入本来应该包括咨询师与当事人双方，但考虑到咨询师往往较为"认真负责"，所以理论家们大多关注的是当事人的投入。本研究投入域中的内容也大都是针对当事人而言。Frieswyk 等主张将同盟定义为当事人在治疗过程中的合作 (Frieswyk, et al., 1986; Frieswyk, Colson, & Allen, 1984)，实际上就是当事人积极投入到治疗任务中的程度，如当事人在会谈中积极主动，谈论重要的主题，坦诚地提供信息和表达感受，好好利用治疗师的帮助，将治疗收获应用于生活，积极改变等 (Allen, Newsom, Gabbard, & Coyne, 1984)。Gaston (1990, 1991) 认为应该对治疗同盟与工作同盟的概念进行区分，前者指当事人对咨询师的情感依附，后者侧重当事人在同盟中的工作能力。这些能力包括当事人自我揭示重要的信息，自我观察，探索自己的问题，体验情绪，对咨询师的意见反应积极，加深对重要主题的探索，为解决问题有目的地工作等。从中可以看出，虽然 Gaston 称之为当事人的"工作能力"，但其内容与这里所说的当事人的投入其实非常相似。另外，Horvath 和 Bedi (2002) 认为工作同盟不仅指咨询师与当事人有共同的目标，还包括双方承担各自的责任，共同为完成各项目标任务而努力。这其中的"承担责任"与"努力"也与本研究所说

的投入的含义相吻合。

Bachelor (1995) 发现在当事人的描述中有三类不同的同盟，一类是抚育型（nurturant），强调的是咨询师的助长态度；一类是领悟型（insight-oriented），看重增进自我理解；一类是合作型（collaborative），体现为当事人的积极参与。这三类同盟的特点恰好与本研究中情感联结、任务和投入三个合作成分一一对应。事实上，抚育型的同盟除了看重咨询师的态度，同样也包括咨询师的指导和当事人充分的自我揭示；领悟型的同盟除了看重咨询师的澄清，也少不了信任和与咨询师的合作；合作型的同盟除了看重当事人的参与，也强调咨询关系与咨询师的反应技术。这说明 Bachelor (1995) 发现的三类同盟中其实均包含着与本研究所得结果类似的几个共同成分，只不过当事人对这些成分的重要性及在同盟中的作用理解不一，所以形成了不同的同盟类型。同样的现象两个研究采用了不同的分析方法，一个是类型学，一个是因素论，所得结果却可以相互补充。

4.2 合作模式、发展变化和影响因素

根据当事人对咨询师的依赖程度和在咨询过程中的参与程度将工作同盟分为咨询师主导、相互合作和当事人主导三种合作模式。大部分的当事人属于前两种合作模式。这说明当事人在同盟关系中仍较多依赖咨询师，或至少认为"合作是一方对另一方的响应，合作关系需要咨询师引导——提供一个大的框架，然后在具体问题上纠正和帮助——这样当事人才会更积极地配合"。国外的研究也一致发现同盟关系的质量更多取决于咨询师，当事人更多将责任放到咨询师身上，在同盟理论中被强调的合作并没有被当事人过多提及 (Bachelor, 1995; Bedi, 2006; Bedi, et al., 2005)。在 Bachelor (1995) 的研究中，合作型的同盟仅占 15%（本研究中与之含义相近的当事人主导的合作模式占 10%），而且当事人更多认为合作是咨询师与当事人各出各的力，而不是同盟理论中通常提到的双方的协商与一致。显然，当事人理解的合作仍然少不了咨询师的积极参与和指导。

在本研究中，工作同盟有三种发展模式，稳定、上升与倒 U 型。前两种变化趋势在研究中均有发现 (de Roten et al., 2004; Kivlighan & Shaughnessy, 2000; Kramer, Roten, Beretta, Michel, & Despland, 2008; Stevens, Muran, Safran, Gorman, & Winston, 2007; Stiles et al., 2004)，而倒 U 型的发展模式则与已有的理论假设相反。Gelso 和 Carter (1994) 认为工作同盟的发展是"U"型（高－低－高）变化模式。他们认为特别是在疗程较短的治疗中，由于当事人对于治疗的乐观主义，开始阶段会有较高水平的工作同盟。随后，当事人往往会经历失望、沮丧等负面情绪，工作同盟的水平也随之下降。但在成功的治疗中，在最后阶段当事人的反应变得更加现实，工作同盟也会恢复到一个较高的水平。倒 U 型的发展模式则正好与之相反。开始当事人与咨询师的同盟关系不断加深，中间双方密切合作完成咨询的各项任务，待各项任务告一段落，合作关系又逐渐疏远。可能的情况是，若咨询目标达到，咨询随即结束，则最后逐渐疏远的阶段不会持续太长。若咨询还有下一步的任务，则这段倒 U 型的关系就只是整个同盟关系发展过程中的一个片段，疏远阶段既可能代表暂时的调整，如商讨下一步的目标，也可能是关系出现了破裂，需要进一步的处理。此外，也不排除咨询师对咨询进程把握不当的可能。不过，此发展模式的代表性并不高（为 rare）。

本研究呈现了 10 个对同盟质量有影响的因素。其中一些在其他质的研究里也被当事人提及，如效果 (Bachelor, 1995)、咨询师的个人特点以及当事人之前的咨询经历等 (Bedi, et al., 2005)。另一些则出现在检验同盟关系影响因素的研究里，如当事人与咨询师的依恋风格、咨询师的经验、咨询师与当事人的匹配 (Horvath, 2001)、当事人的期望 (Patterson, Uhlin, & Anderson, 2008)、动机 (Rumpold et al., 2005)、性别 (Wintersteen, Mensinger, & Diamond, 2005) 等。咨询师的各类反应技术也是常被研究者关注的同盟关系的影响因素 (e.g., Ackerman & Hilsenroth, 2001, 2003)。不过在本研究中，咨询师的反应技术被视作合作过程中的任务活动，被放在了任务域中。

4.3 局限

采用质的方法对中国当事人眼里的工作同盟进行描述，在得到一些有意义的结果的同时，也要认识到本研究的局限。首先，本研究对结果的解释有赖于对工作同盟概念的理解。例如，投入到底是一个当事人变量，或是同盟的构成要素，还是良好同盟的结果 (Horvath & Bedi, 2002)？对咨询的信心是同盟本身，还是同盟的影响因素 (Hatcher & Barends, 1996)？工作同盟是咨询关系的一种，还是一个看待咨询过程的视角 (Hatcher & Barends, 2006)？对此类问题不同研究者可能有不同答案，以致对研究的结果也可能有着不同的解释。

在谈到 CQR 的取样问题时，Hill 等 (1997, 2005) 建议研究者应该从同质的总体中随机选取那些对所研究的问题有着丰富知识的被试，而且最好是近来有过相关的经历，即基于标准的取样。据此，本研究将取样标准定为咨询会谈达到 8 次及以上的当事人，而且咨询经历发生在半年内。但事后检讨，发现在努力满足这个标准的同时，却对另一问题考虑不周：会谈次数达到 8 次及以上的当事人中，消极工作同盟的比例很可能比较小，进而使得研究结果对这类同盟关系的代表性不足。另外，研究中所有当事人都来自大学咨询中心，均为在校学生，且绝大数为女性，这虽然符合 Hill 提出的同质性要求，但也同样可能因此而减少了样本的代表性。不过，代表性问题是质的研究方法的固有特性，其根本解决办法可能需要采取量的研究范式。

此外，不清楚当事人确切的诊断信息，而当事人的症状与问题类型显然会影响其建立同盟的能力 (Gaston, 1991)。比如，边缘型人格障碍的当事人建立同盟关系会特别困难。因此，对这些当事人而言，工作同盟的建立本身就是一个关键的效果变量 (Frieswyk, et al., 1986)。本研究也没有收集有关咨询师的信息，所得结果可能与某些特定的咨询师变量有着密切的关系，比如咨询师的治疗取向、经验，甚至性别等。这些限制使得本研究的结果只能被看作是对工作同盟本土概念化的一个初步探索。

4.4 对临床的启示

当事人来寻求帮助时，大都指望咨询师帮其解决问题，对咨询师较为依赖。当事人的这种想法是可以理解的，原因可能有二。一是当事人找到咨询师时，大多已经非常脆弱，无力或不愿再承担更多的责任。当事人"将咨询作为最后的希望，需要得到咨询师的救助"，而国外的当事人同样觉得"不能处理这么多的责任了"(Bedi, 2006, p. 32)。另外一个原因是当事人对咨询过程不了解，不知道该做些什么，如何与咨询师合作。于是，当事人根据自己的日常经验，如到医院看病，开始向咨询师寻求问题的解决之道。因此，在所有研究里（包括本研究）当事人都强调了咨询师指导的重要性。

事实上，咨询师可以利用各种同盟管理或调控策略来调动当事人，以建立更好的同盟关系。在本研究中，咨询过程的调控也同样是任务域中的重要内容。调控的内容既包括咨询结构上的设置，如结构化、目标任务等，也包括咨询关系的反馈与调整。也就是说，咨询师一方面要告诉当事人在咨询中做什么，如何做，对咨询过程保持一定的结构化，同时，又要鼓励当事人积极参与，主动反馈，保持一定的开放性与灵活性。Bachelor (1995) 发现当当事人可以对咨询师的行为或咨询效果进行讨论或评论时，可以增强同盟关系。也许在此之前，当事人并不知道自己是可以参与到咨询过程之中，而且是被咨询师所鼓励和看重的。同样，当事人也希望更多地参与咨询过程，如"整个咨询都是咨询师说，当事人回答，当事人充当一个小孩子，听咨询师说应该怎么办，感觉不是特别好，希望更积极地加入咨询过程，且能提出一些自己的想法"。咨询师若不够开放或缺乏灵活性与敏感，则常常会抑制当事人的积极性，如"当事人跟咨询师提出一些自己的问题和想法，表达自己的不满，但咨询师没有做出相应的改变"。咨询师能够察觉同盟关系中的问题并与当事人进行协商，既能促进同盟的质量，也会对当事人有治疗性的帮助 (Hill & Knox, 2009)。

一些理论家认为当事人会随着咨询的进程越来

越投入，如 Luborsky (1976) 的二阶段 / 类型同盟理论认为当事人与咨询师建立了信任关系之后则会与咨询师共同努力，共度难关。但 Bachelor (1995) 发现合作型的同盟比例并没有随着咨询的进程而提高，而 Bedi 等 (2005, p. 317) 更是认为即使使用合适的角色导入策略，当事人还是会有意或无意地将建立或增强同盟的责任放在咨询师身上。看来，无论何时咨询师身上的担子都轻不了。

4.5 对研究的启示

本研究为工作同盟这一治疗过程领域的重要变量的本土概念化提供了初步证据。结果表明，中国当事人对工作同盟的看法与西方工作同盟的理论与研究基本吻合。这在一定程度上证明了已有工作同盟理论和结构的跨文化适用性，为后续开展本土的工作同盟研究奠定了基础。例如，可以在此基础上修订与编制相应的工作同盟的测量工具。另外，Bedi (2006) 认为积极与消极的同盟关系在结构上可能不同。为检验此假设，本研究同时对积极的、一般的和消极的工作同盟进行了探索性的分析。结果发现，积极的和消极的同盟关系都可以用一些共同的类别来描述，意味两者有共同的结构。这个结果显然对工作同盟的测量有着重要的意义。这使得目前的测量工具不仅能告诉我们工作同盟有多好，而且还可以让我们知道工作同盟有多差。研究结果表明，合作过程中咨询师的指导与对咨询过程的调控对工作同盟有着重要的意义，而这些变量目前并没有得到充分地研究。这些变量可以成为下一步的研究方向。

致谢： 作为研究小组的成员，梁佳、王曼、朱文臻、赵丽和赵金在数据分析的过程中付出了大量精力，在此表示衷心的感谢。

参考文献

梁觉，周帆 . (2010). 跨文化研究方法的回顾及展望 . 心理学报 , 42, 41–47.

Ackerman, S. J., & Hilsenroth, M. J. (2001). A review of therapist characteristics and techniques negatively impacting the therapeutic alliance. *Psychotherapy: Theory, Research, Practice, Training, 38*, 171–185.

Ackerman, S. J., & Hilsenroth, M. J. (2003). A review of therapist charcteristics and techniques positively impacting the therapeutic alliance. *Clinical Psychology Review, 23*, 1–33.

Allen, J. G., Newsom, G. E., Gabbard, G. O., & Coyne, L. (1984). Scales to assess the therapeutic alliance from a psychoanalytic perspective. *Bulletin of the Menninger Clinic, 48*, 383–400.

Bachelor, A. (1995). Clients' perception of the therapeutic alliance: A qualitative analysis. *Journal of Counseling Psychology, 42*, 323–337.

Bedi, R. P. (2006). Concept mapping the client's perspective on counseling alliance formation. *Journal of Counseling Psychology, 53*, 26–35.

Bedi, R. P., Davis, M. D., & Williams, M. (2005). Critical Incidents in the Formation of the Therapeutic Alliance from the Client's Perspective. *Psychotherapy: Theory, Research, Practice, Training, 42*, 311–323.

Bordin, E. S. (1979). The generalizability of the psychoanalytic concept of the working alliance. Psychotherapy: Theory, *Research & Practice, 16*, 252–260.

Bordin, E. S. (1994). Theory and research on the therapeutic working alliance: New directions. In A. O. Horvath & L. S. Greenberg (Eds.), The working alliance: *Theory, research, and practice* (pp. 13–37). New York: John Wiley & Sons.

de Roten, Y., Fischer, M., Drapeau, M., Beretta, V., Kramer, U., Favre, N., et al. (2004). Is one assessment enough? Patterns of helping alliance development and outcome. C*linical Psychology & Psychotherapy, 11*, 324–331.

Freud, S. (1958). The dynamics of transference. In J. Strachey (Ed. & Trans.), *The standard edition of the*

complete psychological works of Sigmund Freud (Vol. 12, pp. 99–108). London: Hogarth Press. (Original work published 1912)

Freud, S. (1964). Analysis terminable and interminable. In J. Strachey (Ed. & Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 23, pp. 209–253). London: Hogarth Press. (Original work published 1937)

Frieswyk, S. H., Allen, J. G., Colson, D. B., Coyne, L., Gabbard, G. O., Horwitz, L., et al. (1986). Therapeutic alliance: Its place as a process and outcome variable in dynamic psychotherapy research. *Journal of Consulting and Clinical Psychology, 54*, 32–38.

Frieswyk, S. H., Colson, D. B., & Allen, J. G. (1984). Conceptualizing the therapeutic alliance from a psychoanalytic perspective. *Psychotherapy*: *Theory, Research, Practice, Training, 21*, 460–464.

Gaston, L. (1990). The concept of the alliance and its role in psychotherapy: Theoretical and empirical considerations. *Psychotherapy: Theory, Research, Practice, Training, 27,* 143–153.

Gaston, L. (1991). Reliability and criterion–related validity of the California Psychotherapy Alliance Scales–Patient version. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3*, 68–74.

Gelso, C. J., & Carter, J. A. (1985). The relationship in counseling and psychotherapy: Components, consequences, and theoretical antecedents. *The Counseling Psychologist, 13,* 155–243.

Gelso, C. J., & Carter, J. A. (1994). Components of the psychotherapy relationship: Their interaction and unfolding during treatment. *Journal of Counseling Psychology, 41*, 296–306.

Greenson, R. R. (1965). The working alliance and the transference neurosis. *The Psychoanalytic Quarterly, 34,* 155–179.

Greenson, R. R. (1967). *The technique and practice of psychoanalysis*. New York: International Universities Press.

Hatcher, R. L., & Barends, A. W. (1996). Patients' view of the alliance in psychotherapy: Exploratory factor analysis of three alliance measures. *Journal of Consulting and Clinical Psychology, 64*, 1326–1336.

Hatcher, R. L., & Barends, A. W. (2006). How a return to theory could help alliance research. Psychotherapy: Theory, Research, Practice, Training, 43, 292–299.

Hill, C. E., & Knox, S. (2009). *Processing the therapeutic relationship. Psychotherapy Research, 19,* 13–29.

Hill, C. E., Knox, S., Thompson, B. J., Williams, E. N., Hess, S. A., & Ladany, N. (2005). Consensual Qualitative Research: An Update. *Journal of Counseling Psychology, 52*, 196–205.

Hill, C. E., Thompson, B. J., & Williams, E. N. (1997). A guide to conducting consensual qualitative research. *The Counseling Psychologist, 25*, 517–572.

Horvath, A. O. (2001). The alliance. *Psychotherapy: Theory, Research, Practice, Training, 38,* 365–372.

Horvath, A. O., & Bedi, R. P. (2002). The alliance. In J. C. Norcross (Ed.), *Psychotherapy relationships that work: Therapist contributions and responsiveness to patients* (pp. 37–69). New York: Oxford University Press.

Horvath, A. O., & Greenberg, L. S. (1989). Development and validation of the Working Alliance Inventory. *Journal of Counseling Psychology, 36,* 223–233.

Horvath, A. O., & Symonds, B. D. (1991). Relation between working alliance and outcome in psychotherapy: A meta–analysis. *Journal of Counseling Psychology, 38,* 139–149.

Hougaard, E. (1994). The therapeutic alliance: A conceptual analysis. *Scandinavian Journal of*

Psychology, 35, 67–85.

Kivlighan, D. M., Jr., & Shaughnessy, P. (2000). Patterns of working alliance development: A typology of client's working alliance ratings. *Journal of Counseling Psychology, 47*, 362–371.

Kramer, U., Roten, Y. d., Beretta, V., Michel, L., & Despland, J.–N. (2008). Patient's and therapist's views of early alliance building in dynamic psychotherapy: Patterns and relation to outcome. *Journal of Counseling Psychology, 55*, 89–95.

Leung, K., & Zhou, F. (2010). Cross–cultural research methods: Review and prospect. *Acta Psychologica Sinica, 42*, 41–47.

Luborsky, L. (1976). Helping alliances in psychotherapy. In J. L. Claghorn (Ed.), *Successful psychotherapy* (pp. 92–116). New York: Brunner/Mazel.

Martin, D. J., Garske, J. P., & Davis, M. K. (2000). Relation of the therapeutic alliance with outcome and other variables: A meta–analytic review. *Journal of Consulting and Clinical Psychology, 68*, 438–450.

Orlinsky, D. E., Ronnestad, M. H., & Willutzki, U. (2004). Fifty years of psychotherapy process–outcome research: Continuity and change. In M. J. Lambert (Ed.), *Bergin and Garfield´s handbook of psychotherapy and behavior change* (5th ed., pp. 307–389). New York: Wiley.

Patterson, C. L., Uhlin, B., & Anderson, T. (2008). Clients' pretreatment counseling expectations as predictors of the working alliance. *Journal of Counseling Psychology, 55*, 528–534.

Rogers, C. R. (1957). The necessary and sufficient conditions of therapeutic personality change. *Journal of Consulting Psychology, 21*, 95–103.

Rumpold, G., Doering, S., Smrekar, U., Schubert, C., Koza, R., Schatz, D. S., et al. (2005). Changes in motivation and the therapeutic alliance during a pretherapy diagnostic and motivation–enhancing phase among psychotherapy outpatients. *Psychotherapy Research. Special Issue: The Therapeutic Relationship, 15*, 117–127.

Sterba, R. (1934). The fate of the ego in analytic therapy. *The International Journal of Psychoanalysis, 15*, 117–126.

Stevens, C. L., Muran, J. C., Safran, J. D., Gorman, B. S., & Winston, A. (2007). Levels and patterns of the therapeutic alliance in brief psychotherapy. *American Journal of Psychotherapy, 61*, 109–129.

Stiles, W. B. (1993). Quality control in qualitative research. *Clinical Psychology Review, 13*, 593–618.

Stiles, W. B., Glick, M. J., Osatuke, K., Hardy, G. E., Shapiro, D. A., Agnew–Davies, R., et al. (2004). Patterns of Alliance Development and the Rupture–Repair Hypothesis: Are Productive Relationships U–Shaped or V–Shaped? *Journal of Counseling Psychology, 51*, 81–92.

Strong, S. R. (1968). Counseling: An Interpersonal Influence Process. *Journal of Counseling Psychology, 15,* 215–224.

Tryon, G. S., Blackwell, S. C., & Hammel, E. F. (2007). A meta–analytic examination of client–therapist perspectives of the working alliance. *Psychotherapy Research, 17,* 629–642.

Wintersteen, M. B., Mensinger, J. L., & Diamond, G. S. (2005). Do Gender and Racial Differences Between Patient and Therapist Affect Therapeutic Alliance and Treatment Retention in Adolescents? *Professional Psychology: Research and Practice, 36,* 400–408.

Zetzel, E. R. (1956). Current concepts of transference. *The International Journal of Psychoanalysis, 37,* 369–375.

The working alliance in clients' eyes: A qualitative analysis

ZHU Xu; JIANG Guangrong

(School of Psychology, Huazhong Normal University, Wuhan 430079, China)

Abstract　The working alliance has been a focus of psychotherapy research for several decades in the western literature. Although theorists differ somewhat in their conceptualizations of the alliance, most of them agree that the core of this construct is the collaboration between therapist and client that emphasizes the contributions of both participants. Most of our knowledge in this field derives from theorists' hypotheses, and clients' perspective is often neglected. In fact, it was often assumed that clients use the same conceptual dimensions as therapists do to rate the collaborative processes in psychotherapy. This assumption, however, has been challenged by reported low correlation between client–therapist alliance ratings and by results of qualitative studies that show divergence between clients' and therapists' perspectives. Moreover, the literature regarding clients' actual experiences of the alliance is sparse. In an attempt to examine clients' perspectives in viewing working alliance and the cultural adaptability of western alliance theories to Chinese culture, the present study sought to delineate the alliance from Chinese clients' perspective via a qualitative approach and addressed the concordance between clients' perceptions and theoretician–derived views of the alliance. Twenty clients at a university counseling center consented to participate in the study. Clients were interviewed by phone about their experiences of being collaborative with their therapists and their opinions about what is important to this collaborative process in counseling. Data were analyzed by a research team using Consensual Qualitative Research (CQR) (Hill et al., 2005; Hill, Thompson, & Williams, 1997). CQR requires that research team members reach consensus about the classification and the meaning of data through three steps, namely, domain coding, core ideas, and cross subject analysis. Six domains with regard to the collaborative process in counseling emerged from the data: bond, task, engagement, collaboration pattern, development, and influencing factors. Bond is the emotional and personal aspect of the client–therapist relationship, which serves as the emotional base for collaborative work. Task refers to the activities and things client and therapist need to do during the collaborative process, which is the content of the collaborative work. Engagement implies the degree to which client and therapist devote themselves to the process, lending an impetus to the collaboration. These three domains can be viewed as components of working alliance and the other three as descriptions of major features of the alliance.

The study results suggest that the working alliance is perceived by Chinese clients to include bond, task, and engagement. This finding is comparable to current theories of alliance in the western literature and advances conceptualizations of the alliance especially when culture differences are concerned.

Keywords　therapeutic relationship; working alliance; qulitative analysis

管理心理研究所团队成员

马红宇，博士，教授，管理心理研究所所长，主要研究领域为工作生活平衡与组织文化建设、新生代职业心理与管理、胜任特征模型构建与应用、网络心理与应急管理等。

洪建中，特聘教授，主要研究领域为心理与知识管理，包括文化－历史活动理论在知识管理中的应用、知识管理中心理与文化机制、合作性学习与知识创新、心理与知识管理的跨文化研究、网络环境下知识管理的心理研究等。

徐富明，博士，教授，主要研究领域为经济心理与行为决策、教师职业心理学、组织行为和人力资源管理等。

王忠军，博士，副教授，管理心理研究所副所长，主要研究领域为员工发展与职业健康，包括新生代职业心理、工作动机与激励、上下级关系、情绪劳动与情感管理、职场压力与健康、职业生涯成功等。

刘亚，博士，副教授，主要研究领域为组织行为学与人力资源管理、中国文化与中国式管理、大学生职业决策等。

谢员，博士，讲师，主要研究领域为组织行为学与人力资源管理、生涯心理辅导与咨询、员工心理援助计划 EAP，包括压力管理、员工使命感、野心家型员工、生涯适应、生涯决策风格、教师职业发展与心理健康等主题。

唐汉瑛，博士，讲师，主要研究领域为组织行为与人力资源管理，包括反生产行为、工作－生活平衡、谦卑领导、胜任特征建模与应用等

老年人网络使用与心理健康

洪建中，黄　凤，皮忠玲

（华中师范大学 心理学院，青少年网络心理与行为教育部重点实验室，湖北 武汉 430079）

摘　要　基于对 54 名老年人的访谈数据，本研究考察了老年人的网络使用状况以及网络使用对老年人心理健康的影响，结果发现：（1）大多数使用网络的老年人几乎每天都上网，而且上网目的多样化，不仅通过网络与熟人联系、获取信息，还使用网络进行娱乐、交易活动。（2）与不使用网络的老年人相比，使用网络的老年人可以感受到更高的社会联结，而且能够更好的理解他人情绪和错误信念。使用和不使用网络的老年人在孤独感水平上不存在差异。（3）老年人是否使用网络与其感受到的社会联结和理解他人情绪的能力呈显著正相关。这说明网络使用对老年人心理健康具有积极的促进作用。

关键词　老年人；网络使用；社会联结；孤独感；情绪理解；错误信念理解

1　引言

据新华网 2014 年 2 月发布的报道，我国 60 岁以上人口已超过 2 亿，占全国总人口的 14.9%。到 2020 年，我国 60 岁以上的老年人口将达到 2.48 亿，占全国总人口的 17% 以上。随着老年人口数量的逐步增多，其心理健康问题引起了学者们的关注。在心理健康社会情绪方面，闫志民的研究发现，从 1995 年到 2011 年，中国老年人的孤独感水平呈现出上升的趋势。在心理健康社会认知方面，李孝明的研究发现，与青年组相比，两组教育水平分别为中学和大学的老年被试在认知加工任务上的表现都较差；其中，教育水平为中学的老年组被试理解他人情绪和错误信念的能力最差。这说明老年人由于退休、朋友或配偶死亡、身体疾病等原因导致其社交网络缩小，这一方面，使得他们容易产生孤独、抑郁等消极情绪；

另一方面容易引起他们社会认知退化，理论退化是社会认知退化的重要表现之一，心理理论对老年人的正常社会交往起着重要作用。心理理论是指个体对他人或自己的心理状态，如情绪、意图和信念等的认知，以及据此对行为进行因果性解释和预测。心理理论退化的典型表现就是不能准确理解他人的情绪、意图和信念等，以及不能准确解释和预测他人行为，从而在与他人交往时表现出较大的困难。

有关网络使用在促进老年人心理健康中的作用，已被西方学者的研究所证实。如西方学者研究发现，网络使用不仅可以帮助老年人与家人、朋友以及同事保持联系，维持和拓展其社交网络，缓解老年人的孤独感、抑郁等消极情绪反应，还有利于减缓老年人认知退化的速度。

随着信息技术的发展，我国使用网络的老年人也呈显著增长的趋势。第 34 次中国互联网发展报告

指出，截至 2014 年 6 月，国内 60 岁以上的互联网使用者已经达到 1327 万人，相比较于 2013 年底，增长了 153 万人。对于老年人来说，互联网是一个重要的资源，既可以为老年人的人际交往提供便利，减缓其由于退休、行动不便等原因引起的社交困难，也可以直接影响老年人的社会认知功能，对其心理健康具有一定的保护和促进作用。国内学者徐旭对互联网与老年人继续社会化关系的研究发现，使用互联网改善了老年人与子女的联系情况，降低了其孤独感。此外，耿协鑫、周宗奎等学者通过整合关于老年人玩网络视频游戏对其心理功能影响的研究发现，玩网络视频游戏可以有效缓解老年人的认知退化。

从已有文献来看，虽然许多研究表明网络使用对老年人心理健康有重要影响，但他们主要关注的是孤独感、抑郁等消极社会情绪，对积极社会情绪如社会联结关注不足。社会联结反映了老年人对周围世界人际亲密度的自我觉察，它是评价老年人心理健康的一个重要组成部分。已有研究一直强调网络使用有助于老年人保持与他人的联系，那这种联系是否有助于提升老年人感受到的社会联结？这一问题并没有得到足够的重视。目前仅有一项研究考察了网络使用与老年人感受到的社会联结之间的关系。Mellor, Firth 和 Moore 的研究发现尽管被试声称网络使用很有益，测量结果却并没有证实网络使用起到改善老年人感受到的社会联结作用。因此，需要进一步考察网络使用与社会联结之间的关系。此外，已有研究对网络使用与老年人社会认知之间的关系关注也不够。心理理论是评价个体社会认知的重要指标，心理理论水平的高低直接影响着个体在人际交往中能否准确理解他人的情绪、观念和想法，在个体的社会交往中扮演着重要的角色。综上所述，老年人网络使用与其心理健康的社会情绪（社会联结和孤独感）和社会认知（心理理论）层面之间的关系需要进一步澄清。因此，本研究以高校退休教师为对象，探讨老年人网络使用与孤独感、社会联结、心理理论之间的关系，以期能够掌握老年人网络使用和心理健康的基本情况，为帮助老年人使用网络

进而提升其心理健康水平提供科学的依据和指导。

二、 研究方法

鉴于老年人自身的发展特点，如视力退化、阅读困难等，本研究采用结构式访谈的方法搜集数据。2014 年 5 月至 12 月期间，研究者通过两种途径：方便取样和滚雪球取样访谈湖北省武汉市两所高校的退休员工 54 人。方便取样指依靠学校离退体工作处、老年活动中心介绍或在老年人聚集的广场、公园等约谈被试。滚雪球取样是指依靠被访谈的老年人介绍另一些被访人。

54 名被试中，31 名是网络使用者，23 名是非网络使用者。其中 38 名男性，16 名女性。被试的年龄在 60–89 之间，平均年龄为 76.17 ± 7.55。被试中，43 人（79.6%）受教育水平为大学及以上，5 人（9.3%）为大专毕业，3 人（5.6%）为高中毕业，其余 3 人（5.6%）是初中毕业。其居住状况为：43 人（79.6%）与配偶居住在一起，4 人（7.4%）与子女居住在一起，6 人（11.1%）独居，1 人（1.9%）与保姆居住在一起。其收入水平为：46 人（85.2%）收入水平为 4000 元以上，其余 8 人（14.8%）收入在 2000–4000 元之间。96.3% 的被访谈者（52 人）身体健康，行动自如。网络使用者的网络使用时长范围是 2—20 年，大多数被试（23 人）报告已经使用网络 5 年以上。研究采取个别访谈的方式进行，每位被访者访谈时间在 30–45 分钟之间。

本研究依据与心理健康相关的社会情绪和社会认知问题以及网络使用等相关主题构建了访谈提纲，主要涉及三个方面的问题：（1）老年人的网络使用情况；（2）心理健康的社会情绪层面——老年人的孤独感和社会联结水平；（3）心理健康的社会认知层面——老年人心理理论水平。为了确保研究信效度，访谈提纲中所用的问题均来自先前文献中使用的量表，并根据老年人的特定情况进行了适当修改，例如，考虑到问题过多、过长可能会使老年人疲劳，从而引起其反感情绪，本研究选取了最具有代表性的题目考察老年人的网络使用及心理健康水平。具

体如下所述：

2.1 网络使用情况

本研究主要从以下 8 方面考察了老年人网络使用情况：是否使用网络、使用网络的频率、使用网络的目的、使用网络与谁联系、联系的频率以及使用何种工具与他人联系，并且询问了老年人对网络的态度以及为什么不使用网络。其中老年人对网络的态度和为什么不使用网络这两项为开方式问题，其余的问题均提供一定的选项供被试选择。被试对是否使用网络这一条目的回答作为区分上网和不上网被试的依据。

2.2 认知功能量表

为了控制个体的认知功能对社会认知的影响，本研究将认知功能作为控制变量纳入考察范围。采用生活质量综合评定问卷中测量认知功能的 5 个条目测量老年人的认知功能，这 5 个条目主要测量了被试的思维能力、注意力、记忆力、做决定的能力以及对这四方面能力的满意程度。该量表采用 3 点计分，得分越高说明被试认知能力越好，对自己认知能力也越满意。

2.3 孤独感量表

选用 UCLA 孤独感量表的中文版，该量表反应了个体对社交的渴望与实际情况之间存在差距而产生的孤独感。本研究选用该量表的简版共 5 个条目测量个体的孤独感水平，每个条目 3 点计分（1= 从不，3= 一直）。得分越高说明被试感受到的孤独感越高。

2.4 社会联结量表

选用 Lee 和 Robbins 于 1995 年编制的社会联结感量表中的 3 个条目测试被试的社会联结。每个条目同样为 3 点计分，得分越高说明被试的社会联结越强，体验到与社会失去联系的感觉越少。

2.5 心理理论故事

本研究采用的心理理论任务有：（1）情绪理解任务，（2）错误信念理解任务。其中情绪理解任务采用 1 个故事考察被试在人际交往情景中理解他人情绪的能力，故事结束后有 6 个问题，这 6 个问题主要要求被试判断故事的主人翁是否说错了话并说明为什么不该那样说，该任务得分在 0–6 分之间。错误信念理解任务同样采用 1 个故事考察被试在生活情景中理解他人错误信念的能力，故事结束后有 2 个问题，这 2 个问题主要考察被试是否觉察到故事的主人翁对于客观事实有一个错误的假定或想法，该任务得分在 0–3 分之间。具体操作方法是分别给被试讲故事，故事结束后要求被试根据故事内容回答问题。

3 研究结果

3.1 老年人网络使用的特点

3.1.1 老年人的网络使用情况

通过访谈发现，大多数使用网络的老年人几乎每天都上网，说明老年人上网的频率还是很高的。他们使用网络的目的主要是获取信息和交流沟通。此外，使用网络与他人交流沟通的老年人中，大部分每周会通过网络与家人、朋友联系 2–3 次，没有老年人使用网络结识陌生人。最后，老年人最喜欢使用电子信箱和 QQ 与熟人联系（详见表 1）。

3.1.2 老年人使用和不使用网络的心理动因

访谈结果表明，使用网络的老年人对网络使用持积极态度，认为网络使用给其带来了积极影响。由表 2 可知，老年人使用网络的首要原因是利用网络学习新知识，开阔眼界，被访谈者提到比较多的包括"使用网络开阔眼界"、"网络信息丰富"、"方

表 1 老年人的网络使用基本情况（31 人）

使用网络频率	人次	使用网络的目的（多选）	人次	使用网络与他人联系的频率	人次	与谁联系（多选）	人次	使用何种网络工具（多选）	人次
每月 1 次	0	获取信息	30	几乎不使用	5	家人	23	电子邮箱	23
每月 2–3 次	1	交流沟通	27	每月 2–3 次	2	朋友	25	QQ	20
每周 1 次	3	网络娱乐	16	每周 1 次	5	陌生人	0	微信	12
每周 2–3 次	2	商务交易	6	每周 2–3 次	3	——	—	微博	6
几乎每天	25	——	—	几乎每天	16	——	—	其他	2

表 2　老年人使用和不使用网络的心理动因

使用网络的心理动因	人次	不使用网络的心理动因	人次
学习新知识	15	身体不便	15
方便生活	10	不会使用	12
获得快乐	4	没有电脑	3

便请教别人问题";其次,老年人认为网络给生活带来了便利,例如有被访谈者提到,使用网络"方便与儿子、孙子联系"、"使用网络预订服务,出游更方便";第三,使用网络的老年人认为网络给他们的生活带来了快乐,认为网络是老年人的朋友,有被访谈者指出使用网络可以"传递正能量"、"对身心健康有帮助"、"获得快乐和健康"。

不使用网络的老年人由于受到客观条件的限制而没有使用网络。由表 2 可知,老年人不使用网络的首要原因是身体不便。当问及被试为什么不使用网络时,老年人回答最多的是"眼睛不好"、"动作迟缓"、"思维跟不上"。其次,不会使用电脑、不会拼音等是阻碍老年人使用网络的第二大原因,被访谈者提到"尝试学习使用电脑,后来因为学不会放弃了"、"电脑太复杂"、"搞不清楚怎么用"、"不会打字"。第三,也有一小部分老年人提到没有使用网络是因为没有电脑。

3.2　老年人心理健康状况

3.2.1　使用与不使用网络老年人心理健康状况比较

为了便于认识老年人心理健康的基本特点,表 3 列出了社会联结、孤独感、认知能力、情绪理解能力和错误信念觉察理解能力的平均数和标准差。由表 3 的结果可知,使用网络和不使用网络的老年人在社会联结、情绪和错误信念理解能力上存在显著差异,不使用网络的老年人的社会联结、情绪理解能力和错误信念觉察能力都低于使用网络的老年人。而使用网络和不使用网络的老年人在认知功能上并不存在显著差异,这说明使用网络确实有助于老年人降

低孤独感、提升理解他人情绪和觉察他人错误信念的能力,对老年人的心理健康产生了积极影响。

3.2.2　老年人网络使用与心理健康状况的关系分析

通过相关分析考察老年人网络使用与心理健康之间的关系,结果表明(见表 4),老年人是否使用网络与其社会联结、情绪理解能力存在显著的正相关关系。这说明,老年人越是使用网络,其感受到的社会联结就会越强,理解他人情绪的能力也越高。

4　讨论

本研究深入分析了老年人网络使用与心理健康

表 4　老年人网络使用与心理健康之间的相关性

	社会情绪		认知功能	心理理论	
	社会联结感	孤独感		情绪理解	错误信念理解
是否使用网络	0.29*	−0.06	−0.09	0.31*	0.24+

的基本状况以及二者之间的关系问题,这有助于认识老年人网络使用和心理健康的特点以及如何指导老年人使用网络、促进其心理健康发展等问题具有启发意义。

4.1　老年人网络使用的现状与发展

本研究发现,绝大多数使用网络的老年人几乎每天都上网,使用网络已然成了老年人日常生活的一个组成部分。老年人的网络使用主要有三个特点:第一,老年人使用网络的目的与年轻人并不存在差异。交流沟通和获取信息是老年人使用网络的两大主要目的。在访谈中,被试提到,他们使用网络主要是为了与子女、亲戚、老同学、老朋友联系,这表明使用网络对老年人的社会交往有着重要的促进作用。网络也是老年人信息的重要来源,通过网络老年人对社会上发生的事了解的更清楚。网络还是老年人的重要娱乐方式。

表 3　使用与不使用网络老年人心理健康状况的比较

	社会情绪		认知功能	心理理论	
	社会联结	孤独感		情绪理解	错误信念理解
使用网络	8.67+0.94	5.97+1.14	11.22+2.91	4.16+1.00	1.42+1.11
不使用网络	8.00+1.31	6.13+1.54	11.73+2.51	3.26+1.79	0.96+0.64
t	2.21*	0.44	−0.68	2.17*	1.92+

注:+ 表示 $p<0.1$,* 表示 $p<0.05$;以下均同。

通过网络看电视、玩游戏也受到老年人的青睐，甚至还有老年人使用网络进行商务交易，包括在网上购物、预定车票、酒店、旅行社等。第二，网络是老年人维系其已有的社交关系的重要途径，表现为老年人使用网络主要是与熟人联系。本研究中，使用网络与他人沟通的老年人没有一人报告通过网络结识了新朋友。第三，老年人偏爱比较熟悉的、易用的网络通讯工具，但新兴的网络工具也能引起老年人的兴趣。本研究发现，老年人最多的是使用电子邮箱交流心得、互换信息。QQ 作为及时通讯工具也颇受老年人喜欢。也有老人使用微信、微博等网络通讯工具与他人联系。

4.2 老年人网络使用与心理健康状况

本研究发现，与不使用网络的老年人相比，使用网络的老年人报告的社会联结水平显著较高，而且老年人的网络使用能显著预测老年人的社会联结水平。已有研究指出，网络使用可以改善老年人与他人联系的状况，但与其感受到的社会联结水平无关。如 Mellor 等通过对 8 名被试进行一年的追踪研究发现网络使用与老年人感受到的社会联结之间没有关系。这一研究结果与本研究结果并不一致，可能由以下两个原因引起的：（1）Mellor 的研究中被试过少，本研究中的被试量远远多于他们的研究；（2）Mellor 研究中的被试接触网络最长的只有一年，本研究中被试的网龄最短有 2 年，最长有 20 年。因此，可以认为本研究为网络使用与社会联结之间的关系提供了更直接、更有力的证据，证实了网络使用在老年人社会联结发展中的重要性。

本研究发现，使用网络的老年人的社会认知能力显著高于不使用网络的老年人，具体表现为使用网络的老年人能更好的理解他人情绪和错误信念。已有研究大多关注网络使用对老年人认知功能和社会适应功能的影响，如 Kueider 等综述已有研究发现，玩网络游戏对老年人的注意力、认知加工速度以及其他认知功能的退化具有保护作用；耿协鑫等的研究发现玩网络视频游戏可以改善老年人的社会适应情况，本研究进一步考察了网络使用改善老年人社会适应情况的社会认知机制。这一研究结果证实了网络使用对老年

人社会认知发展的重要性。使用网络有助于老年人克服由退休、身体不便等带来的社会交往障碍，使得老年人的社会交往不再局限在物理空间上可以到达的地方，极大地拓展了老年人的社交网络，对提升老年人理解他人情绪和错误信念能力有积极的作用。

尽管本研究证实，网络使用对老年人社会联结、情绪和错误信念理解能力都有显著的影响，但是，本研究并未发现，使用网络与不使用网络的老年人在孤独感水平上存在差异。关于网络使用对老年人孤独感水平的影响，已有多项研究证实。如 Cotton 和其同事的研究发现，老年人网络使用的频率与孤独感存在显著的负相关关系，使用网络频率越高的个体，孤独感水平越低。然而，本项研究发现，老年人是否使用网络、使用网络的频率以及使用网络与他人联系的频率均不能预测其孤独感。这与已有研究的结论存在差异，主要可能是由以下原因引起的：本研究中的被试均来自高校退休员工，其收入水平和受教育水平比较高。高收入水平以及高受教育水平对老年人心理健康的保护作用已被证实。如彭义升、李娟和戴必兵的研究发现，北京市高校离退休老人的整体心理健康水平显著高于全国平均水平。本研究样本来源比较单一，未来研究可考虑在其他老年群体（如社区老人、福利院老人等）中展开研究。与高校退休员工相比较，这些老年群体受教育水平以及收入水平相对较低，其心理健康状况与高校退休员工存在一定的差异，对其进行研究可以提高本研究结果的普适性。

4.3 启示

本研究结果发现使用网络对老年人心理健康的社会情绪和社会认知层面均有积极的保护和促进作用。网络使用对老年人心理健康的影响主要是通过信息获取和交流沟通这两个功能实现的，信息获取有助于老年人了解其周围的世界从而起到提升其社会联结的作用，交流沟通有助于老年人维持其社交网络，并在交往过程中保持情绪和错误信念理解的能力。本研究结果对帮助老年人使用网络，提升老年人心理健康的水平方面主要有以下几点启示：

第一，在网络使用方面，要继续加强老年大学有关计算机和网络使用相关课程建设以及相关的培训，保证畅通的网络学习渠道，以降低老年人由于客观条件（如，不会用、没有电脑）的限制而不能接触到网络的可能。同时，应特别关注收入水平、受教育水平较低的老年人，尽可能为他们提供平等的接触和使用网络的机会。

第二，在社会联结方面，通过互联网平台给老年人提供获取外界信息的渠道，保证老年人能够了解其周围正在发生事情，让他们感觉到自己与周围的世界是有联系的，以降低老年人，尤其是不使用网络的老年人由于退休、身体不便等带来的社会脱节感，促进其心理健康水平的提升。

第三，在社会认知方面，积极通过互联网拓展老年人的社交网络和视野，增加老年人与他人联系、交往的机会，从而减缓其情绪以及错误信念理解能力的退化，提升其与他人交往的质量，最终达到提升其心理健康水平的作用。

参考文献

耿协鑫，周宗奎，魏华 等 . (2014). 视频游戏对成功老龄化的影响，*心理科学进展, 22.*

李董平，张卫，李霓霓 等 . (2008). 未来时间洞察力，目标定向，社会联结与大学生学习适应，*心理发展与教育, 1.*

李孝明，汪凯，陶倩 . (2010). 老年人心理理论与认知加工关系研究 . *心理科学, 2.*

林佳燕，傅根跃，刘文庆 . (2010). 成人心理理论的研究回顾与展望 . *心理科学进展, 3.*

穆光宗，张团 . (2011). 我国人口老龄化的发展趋势及其战略应对 . *华中师范大学学报 (人文社会科学版), 5.*

彭义升，李娟，戴必兵 . (2009). 北京市高校离退休老人心理健康状况及其相关因素 . *中国老年学杂志, 29.*

汪向东，王希林，马弘 等 . (1999). *心理卫生量表评定手册* ，北京：中国心理卫生杂志社 , 127–131.

王异芳，苏彦捷 . (2005). 成年个体的心理理论与执行功能 . *心理与行为研究, 3.*

新华网：《我国 60 岁以上老年人数量突破 2 亿》，北京：新华社，http://news.xinhuanet.com/politics/2014-02/19/c_119412552.htm，2014 年 2 月 19 日。

徐旭 . (2013). *互联网对老年人继续社会化的影响* . 东北财经大学硕士学位论文 .

闫志民，李丹，赵宇晗 等 . (2014). 日益孤独的中国老年人：一项横断历史研究 . 心理科学进展, (7).

中国互联网信息中心：《中国互联网发展状况统计报告》，北京：中国互联网信息中心， http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg/201407/P020140721507223212132.pdf，2014 年 7 月 21 日

Cotten, R., Shelia, A. A., & William, M. M. (2013). Brandi. "Impact of internet use on loneliness and contact with others among older adults: Cross-sectional analysis." *Journal of Medical Internet Research, 15*(2), e39. doi:10.2196/jmir.2306.

Cotten, R., Shelia, Ford, G., Ford, S., et al. (2012). "Internet use and depression among older adults." *Computers in Human Behavior, 28*(2), 496–499.

Johnson, G. M. (2008). "Cognitive processing differences between frequent and infrequent Internet users." *Computers in Human Behavior, 24*(5), 2094–2106.

Kueider, M. A., Parisi, M. J., Gross, L. A., et al. (2012). "Computerized cognitive training with older adults: a systematic review." *PloS ONE, 7*(7), e40588. doi:10.1371/journal.pone.0040588

Lee, M. R., & Robbins, B. S. (1995). "Measuring belongingness: The Social Connectedness and the Social Assurance scales." *Journal of Counseling Psychology, 42*(2), 232–241.

Mellor, D., Firth, L., & Moore, K. (2008). "Can the Internet Improve the Well-being of the Elderly?" *Ageing international, 32*(1), 25–42.

Internet use and psychological well-being among old adults

Hong Jianzhong , Huang Feng , Pi Zhongling

(Key Laboratory of Adolescent Cyberpsychology and Behavior, Minstry of Education, School of Psychology, Central China Normal University, Wuhan, 430079)

Abstract　Based on the interview of 54 old adults, this study investigated the Internet use among old adults and examined the influences of Internet use on the psychological well–being of old adults. The results demonstrated that: (1) Most of the old adults used Internet on daily basis. They used Internet for communication, seeking information, entertainment and commercial purposes. (2) Compared with the one who did not use Internet, old adults who used Internet reported higher level of social connectedness, emotion understanding ability and false belief understanding ability. Both groups using and not using Internet did not have any significant difference in loneliness. (3) There was a significant positive correlation between Internet use and social connectedness and emotion understanding ability. The results imply that Internet use is not only good for the development of social emotion but also beneficial to that of social cognition for old adults.

Keywords　old adults; Internet use; social connectedness; loneliness; emotion understanding; understanding of others' false believes

职业决策理论的线索与趋势

刘 亚，龙立荣

摘 要 本文从（1）强调结果的个人－环境匹配理论；（2）关注决策过程的职业决策研究；（3）关注职业决策困难（职业未决）的诊断工具研究；（4）关注决策因素的职业价值观研究共四个方面，重新梳理了职业决策理论的脉络，分析了西方职业决策理论本身的理论和应用问题，并结合我国区域社会经济发展不平衡的实际及市场经济条件下大学生需求的特点，初步展望了大学决策研究中关注社会生态因子的趋势。

关键词 职业决策；职业价值观；社会生态模型

国外关于职业决策的研究始于 20 世纪初 Parsons(1908) 的开创性研究。从 20 世纪 50 年代至今，已经形成了许多较为成熟的理论。但令人困扰的是这一领域至今都没有形成相对统一的概念框架和理论范式。同时，已有的研究对于解释中国大学生的职业决策行为和过程既缺乏有效的解释力，也难以指导职业辅导的实践。为便于理解，本文将已有的理论归结为以下 4 类：（1）强调结果的个人－职业匹配理论；（2）关注决策过程的职业决策研究；（3）关注职业决策困难（职业未决）的诊断工具研究；（4）关注决策因素的职业价值观研究。基于对已有研究的理论和实践问题的分析，作者尝试指出在中国文化下应用西方理论的前提条件和描述解释中国大学生择业行为的局限和困境，并初步提出和阐释了提升职业理论生态文化效度的新思路和未来的理论发展趋势—大学生择业的生态模型理论。

1 强调结果的个人－职业匹配理论

西方职业决策理论主要围绕个人的视角展开，关注个人利益最大、效价最高、最适合和个人自我价值实现。从心理学角度的理论主要是强调个人心理属性与职业要求匹配的特质论，如 Parsons(1908),Williamson(1939), Holland(1976,1980) 等人的理论 P62~69；从社会学角度的理论如 Gottfredson 对择业过程中社会声望和性别角色的作用，提出的范围限制－妥协理论；经济学的研究，则注重劳动力需求与供给的关系，探讨择业过程中供给与需求的关系对择业的影响。这些理论有一个共同特点，都是从个人与职业特点之间的匹配这一角度出发的简单职业决策模型。其特征如图 1 所示：
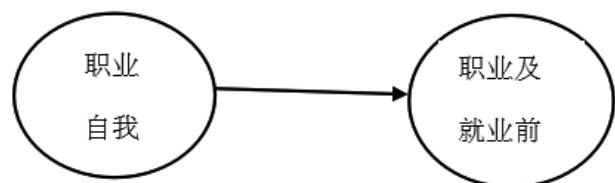
图 1 个人择业决策示意图

强调个人－职业匹配的单维度职业决策理论主要包括 Parsons 的职业决策匹配模型、Holland 的个人—职业相互作用模型、工作调适理论 (theory of work adjustment)、界限妥协理论 (Theory of Circumscription and Compromise)、社会认知职业理论 (social cognitive career theory, SCCT) 等。

个人－职业匹配理论假设匹配的职业决策就是好的或有效的职业决策。但这一假设并没有得到实证的支持，匹配性并不是优化的职业生涯决策结果的良好

预测指标（Furnham, 2001; Spokane et al., 2001;Tinsley, 2000）。Tinsley （2000) 指出，Holland 六边形的一致性系数和职业满意没有显著相关。Holland 职业兴趣的结构，以及是否和如何考虑用这一结构来解释个体的职业兴趣和进行一致性评估等问题目前还没有形成共识。个人 – 职业匹配理论作为一种机械静态的人格观，严重忽视了人的可塑性和职业选择的变通性。

2 注重过程的职业决策研究

在 20 世纪中期，职业发展理论和研究的趋势之一就是对于帮助控制职业行为的认知变量和过程的关注越来越多（Lent, R. W., Brown, S. D.& Hackett, G.2001）。Gati 等（2001) 指出职业生涯决策有如下特点：

（1） 存在必须做出决策的个体和有一个决策要做；

（2）有几个备选项，决策者有某些方面和标准，可以用它们对比和评估各个选项来确定最适合的选项；

（3）要在彼此冲突的目标之间进行权衡。由于许多决策中包含着不确定性和每种结果发生的概率是未知的，因此职业决策是极其复杂的决策过程。

关注过程的职业决策模型的演化趋势和过程是，由早期的受标准化决策理论影响较大的理性决策模型，逐渐发展到强调有限理性思想的理论模型演变，后期的模型逐渐克服了标准化决策模型的理想化和过分理性的特征。标准化决策理论 (normative decisionmaking) 是基于理性人假设的经典决策理论。该理论相信决策者能够加工所有相关信息，遵循着效用最大化原则 (V Neumann &Morgenstern, 1947; E. Edwards,1954) 做出完全理性的选择。并且选项的优势和劣势之间能够互相补偿，某个选项的期望效用是各种效用被各自的概率赋予权重后的效用之和。标准化理论认为， 个体的偏好在决策过程中可以传递 (transitivity) 和保持不变 (consistency)。早期关注过程的职业决策理论都是基于标准化决策理论的思想

提出来的。

如果说匹配理论关注的是什么是理想的职业选择，那么过程理论则关注如何达到这种理想的状态. 所有过程理论的共同的特点就是围绕个体的认知决策的微观机制展开。各种理论也都有不同程度的实证研究的支持。但理论的多样化和概念的不统一本身表明职业决策过程比预想的要复杂。各种过程理论所面临的一个共同难题就是，决策过程极端地依赖决策者对价值的评估、概率的判断等等，而一个人如果能够清晰地意识到自己想要的是什么，那么决策就变得简单了，简单到可能根本就不需要过程。

3 关注职业决策困难（职业未决）的研究

职业决策的研究另外一个思路和主题是职业决策困难及其影响因素。职业决策困难，是指个体在特定的时间里， 无能力作出一个特定的决策。Krumboltz 认为其实质是对决策不满意，或是由于与职业相关的学习经验不够，或是由于个体还没有学会或运用一套作职业决策的系统方法所导致的一种决策状态。决策困难是没有学会某种学习经验的自然结果。Crites 认为是个体无能力挑选或是承诺于一个特殊的，即将准备或进入特定职业的行动过程。龙立荣等认为职业决策困难是个人在职业选择 (进入阶段或职业改变) 过程中，面临最后决策时，不知道要从事什么职业或从几个职业中挑选一个时发生的困难。

这一研究取向的成果，集中体现在以下几个有代表性的职业决策困难的检测工具和量表开发使用， 以及围绕这些量表结构和应用的研究：Career Decision Scale, (CDS) (Osipow, 1987)、Career Factors Inventory, (CFI)、Career Decision Difficulties Questionnaire (CDDQ)。作为经验和实践导向的研究，其主要研究目的是服务于职业辅导的实践，即分拣和筛选出决策困难者，因此有关决策困难的研究在西方理论界也未能够有一个统一的表述，其理论逻辑也并不清楚。虽然国内已经有大量关于上述三个

工具验证和使用的文献，但这些工具基本构念的文化适应性依然缺少有力的验证。

4 国内有关大学生职业价值观的研究

如果说早期的匹配理论是要回答"什么是一个好的职业决策？"关注认知过程的决策理论关注"如何做一个好的决策"的话，那么价值观的研究就是要关注"决策的依据和目标是什么？"

国内近年来关于大学生职业领域的研究相当丰富，如方俐洛、凌文辁等（1996）对中国职业兴趣量表的修订和建构；田宝珍（1997）对大学生择业价值取向的分析；于海波、张进辅（2003）对大学生需要特点的调查；梁龙娟、余贤君对大学生择业标准的研究等等。比较不难发现，国外关于职业决策的研究有着理论和应用两条比较清晰的线索，研究的主题也相对集中和相互关联。而国内的研究多为以国外相关理论为基础的探索，并且国内的多数研究围绕大学生职业价值观、职业兴趣领域的研究展开。

目前国内这一领域的研究虽然数量不少，但大多是复制性的研究或对国外理论的验证性研究。并且对于大学生职业价值观的构成成分和结构业没有一个一致的表述。金盛华（2005）认为我国心理学领域对职业价值观研究尚无稳定的理论基础，并且由于研究方法的差异，研究结果各不相同，很难互相验证。如凌文辁的问卷项目主要来源于文献中筛选的职业价值观条目，宁维卫的工具来源于翻译和修订的国外问卷。因此依然有必要进一步地研究。

5 大学生择业生态模型构想
———职业决策研究的新思路

从上述对西方及国内有关职业决策研究的文献梳理，不难发现，各种理论的关注视角差异很大，在同一视角下的研究也缺乏统一的表述和一致的结论。随着工作世界的持续变化，没有任何一种理论能够完整地描述和有效地解释真实的职业决策过程

和行为，因此也难以在职业辅导实践中应用。同时，由于中西方在社会经济和文化上差异较大，来自西方的理论也难以有效地描述和解释中国人的择业行为。因此，龙立荣等国内学者基于西方理论本身的问题及其在中国应用的生态文化效度的考虑，开始探索新的关注生态效度的理论框架来描述中国人的职业决策行为。

龙立荣（2006）指出在市场经济体系比较完善的制度环境下，个人的职业自我属性是择业的最重要的影响因素。只要个人能够胜任职位的工作（知识经验和能力），与自己的价值观念（包括兴趣）和社会规范（包括社会声望、性别）相符，该职业的就业前景广泛就是合理的。至于这些符合要求的职位或职业在哪种类型的组织、在一个国家的哪个地区，似乎并不重要。因此西方描述和解释择业决策主要考虑（图 1 所示）的模式，即职业自我与职业的关系。随着社会经济发展水平的提高西方人也开始关注组织的差异，但这种考虑的顺序仍然以职业为核心组织的作用是次要的。

基于对中国社会经济、文化发展的现状和市场经济时代人们的价值追求特点的分析，龙立荣提出了的大学生择业的社会生态模型的初步构想：(1)西方的择业理论过于注重择业者的职业自我属性，相对忽视职业所在的组织和地区社会生态环境的作用价值，与中国现实的社会政治经济发展差距太大，难以预测和解释大学生的择业行为；(2)择业决策的影响因素除了个人的职业自我属性外，外部提供的职业机会、外部就业机会的社会生态价值也是人们非常注重的优化匹配因素；(3)在整个择业匹配过程中，个人职业自我与组织环境和地区环境匹配度的预测力，大于与职业本身匹配的预测力；(4)人们看重组织环境是出于薪酬待遇和个人职业生涯的可持续发展双重考虑；人们看重地区环境是突出地区经济、文化的长久回报价值。简言之，地域、行业、组织等生态因子对个体现实职业决策的预测优于和大于个体职业自我属性（如能力、兴趣）的预测力。
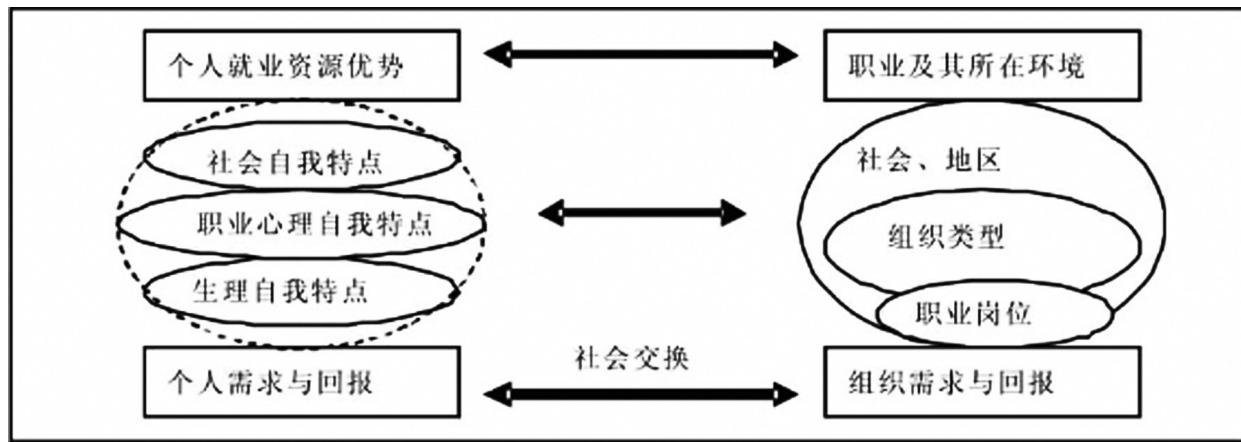
图 2　择业决策的社会生态模型示意图

生态模型理论强调地域、组织等生态因子的突出作用，在决策过程中这些因子是更显著、更容易觉察以及具有比能力、兴趣等个人因子具有更大预测力的因子。同时，一个人的职业心理属性具有一定的弹性，而且同一种职业个性适合的职业往往不是惟一的。只要这些职业需要的个人心理属性相差不大，个人胜任就不会有困难。人们适合从事的职业是一个空间概念而不是一个点的概念（如图 2 所示）。在市场体系的不成熟、地区、行业和组织发展的不平衡性的条件下，相同的付出在不同的组织，地区回报差异十分显著。因此导致人们为了找到理想的组织和预计长期回报较高地区不惜牺牲自己的特长和积累。

最后，龙立荣指出在经济发展水平提高、地区间发展相对平衡的情况下，人们进行职业生涯选择时关注职业自我的可能性才会增加，西方职业生涯决策的模型才会适用。至于在上述条件基本满足的情况下，中国人相对于西方人是否更加突出社会自我在择业中的作用还需要进一步探索。

基于对西方职业本身困难和局限的分析以及对职业决策社会生态模型的初步构想的介绍，研究者认为，从西方职业理论发展本身和建构中国经济文化环境下有解释力的职业理论两方面来看，关注职业决策者决策时所处的包括时代、文化和经济发展水平等更广泛的决策环境，提高职业决策理论的生态效度应该是未来国内职业决策领域研究的发展趋势，惟其如此，才能够真正解决职业理论建构和职业辅导实践脱节的现状，使职业决策理论成为有效地指导职业辅导的理论框架。

参考文献

杜睿，龙立荣 . (2006). 大学生职业决策困难问卷的初步研究 . *中国临床心理学杂志, 14* (3).

金盛华，李雪 . (2005). 大学生职业价值观：手段与目的 . *心理学报* , (5).

凌文轮，方俐洛，白利刚 . (1999). 我国大学生的职业价值观研究 . *心理学报*，(3).

龙立荣，方俐洛 等 . (2002). 社会认知职业理论与传统职业理论比较研究 .*心理科学进展* , (2).

龙立荣，黄小华 . (2006). 大学生择业的社会生态模型：环境的力量 .*高等教育* . (8).

龙立荣，彭永新 . (2000). 国外职业决策困难研究及其启示 .*人类工效学* . 6(4)

Gati, I., Osipow, S. H., Krausz, M., & Saka, N. (2000). Validity ofthe Career Decision Making Difficulties Questionnaire: Counselee versus careercounselor perceptions. *Journal of Vocational. Behavior, 56,* 99–113.

Kevin, R., Kelly, & Wei, C. L. (2002). Mapping the Domain of Career Decision Problems. *Journal of Vocational Behavior, 61*, 302–326.

Lent, R. W., Brown, S. D., & Hackett, G. (2000).Contextual

supports and barriers to career choice: A social cognitive analysis. *Journal of Counseling Psychology, 47*, 36–49.

Phillips, S. D., Pazienza, N. J., & History. (1988).

Theory of the Assessment of Career Development and Decision Making. NewJersey: Lawrence ErlbaumAssociates.

边界弹性与工作 – 家庭冲突、增益的关系：基于人 – 环境匹配的视角 *

马红宇 [1,2]，申传刚 [1,2]，杨　璟 [3,4]，唐汉瑛 [1,2]，谢菊兰 [1,2]

([1] 青少年网络心理与行为教育部重点实验室)

([2] 华中师范大学心理学院暨湖北省人的发展与心理健康重点实验室 , 武汉 430079)

([3] 中国科学院心理研究所 , 北京 100101)

([4] 中国科学院大学 , 北京 100039)

摘　要　本研究从人 – 环境匹配理论的视角探讨工作和家庭边界弹性能力和边界弹性意愿对个体工作 – 家庭冲突和工作 – 家庭增益的交互影响。通过问卷法共获得 494 份有效数据 , 基于多项式回归分析和反应曲面分析的结果表明：工作弹性能力与工作弹性意愿的匹配对工作→家庭冲突有显著的负向效应 , 对工作→家庭增益无显著影响 ; 家庭弹性能力与家庭弹性意愿的匹配对家庭→工作冲突、家庭→工作增益有显著的负向效应。

关键词　人 – 环境匹配 ; 边界弹性能力 ; 边界弹性意愿 ; 工作 – 家庭冲突 ; 工作家庭增益

分类号　B849:C93

1　问题提出

工作和家庭是现今成年人生活中的两个重要组成部分 , 工作角色和家庭角色是个体最主要的社会角色 , 它们之间相互影响 , 密不可分。目前关于工作和家庭角色之间关系的研究主要基于两种视角：一种是消极视角 , 认为由于个体有限的时间、高水平的压力和竞争性的行为期望而引发角色之间的相互冲突 , 即工作 – 家庭冲突 (work–family conflict, WFC) (Eby, Casper, Lockwood, Bordeaux & Brinley, 2005; Greenhaus & Beutell, 1985); 另一种是积极视角 , 认为个体可以从工作和家庭角色的投入中获得有意义的资源 (如自

尊、经济收入等), 继而提升个体在相对角色领域的表现 , 即工作 – 家庭增益 (work–family enrichment, WFE) (Greenhaus & Powell, 2006)。这两种不同的视角为个体处理工作和家庭之间的关系提供了新的思路 , 即个体在履行工作和家庭角色时如何尽可能在减少角色间冲突的同时增加角色间的积极溢出。

边界理论的提出为我们回答上述问题提供了可能。根据边界理论 , 个体会围绕其所在的工作领域或家庭领域建立起不同的角色边界 , 个体每天投身于工作和家庭领域并进行着跨边界的角色转变活动 (Ashforth, Kreiner, & Fugate, 2000; Clark, 2000; Kreiner, 2006)。不同的个体 , 由于价值观、对工作和家庭所

持有的分割 – 整合偏好不同，在面对另一领域的角色需求时，个体是否愿意进行相应角色转变的意愿程度有所不同，这种是否愿意从一个角色向另外一个角色转变的程度称之为边界弹性意愿，包括工作弹性意愿 (work flexibility–willingness, WFW) 和家庭弹性意愿 (family flexibility–willingness, FFW) (Matthews & Barnes–Farrell, 2010; Matthews, Bernes–Farrell, & Bulger, 2010)。另一方面，由于组织政策、家庭责任、管理者和家人支持等外部环境因素的影响，环境对其进行角色转换所提供条件或资源也不相同 (Ashforth et al., 2000; Clark, 2000; Kossek, Lautsch, & Eaton, 2004)。研究者将个体对其能否离开所在领域去满足另外一个领域需求的外部环境特征的认知性评估称之为边界弹性能力，包含着工作弹性能力 (work flexibility–ability, WFA) 和家庭弹性能力 (family flexibility–ability, FFA) (Matthews & Barnes–Farrell, 2010; Matthews et al., 2010)。当个体身处一个领域，而另一领域有需求需要其进行角色转变时，除了会受到组织政策、家庭责任等环境因素的影响之外，还会受到个人的角色转变意愿的影响。这提示我们，个体对其所处的环境所能提供的角色转变的条件或资源的认知性评估与个体进行角色转变的意愿之间的匹配可能是影响个体工作 – 家庭结果变量的重要决定性因素。因此，运用人 – 环境匹配 (person–environment fit, PE fit) (Edwards, 2008; Kristof–Brown, Zimmerman, & Johnson, 2005) 理论来分析边界弹性意愿 (个人需求) 和边界弹性能力 (环境资源) 之间的匹配对工作 – 家庭冲突和工作 – 家庭增益的影响将增进我们对工作 – 家庭界面的理解。

尽管工作 – 家庭界面的研究已经相当丰富，但是以往的研究仍然存在一些不足。首先，尽管人 – 环境匹配一直都是学者们关注的核心理论构念 (Edwards, 2008)，但是纵观国内外的研究，多数研究仍聚焦在情境需求对工作 – 家庭界面的影响上 (Allen et al., 2012; Ford, Heinen, & Langkamer, 2007; Greenhaus, Ziegert, & Allen, 2012; McNall, Masuda, & Nicklin, 2010; Odle–Dusseau, Britt, & Greene–Shortridge, 2012; Straub, 2012)，只有少数研究运用人 – 环境匹配的构念探讨组织环境和个人需求之间的匹配对工作 – 家庭界面重要结果变量的影响，但是这些研究却忽视了家庭作为另一类重要的环境变量与个人需求之间的匹配 (Chen, Powell, & Greenhaus, 2009; Kreiner, 2006; 马丽，徐枞巍，2011)；其次，尽管个体可能同时体验到工作 – 家庭冲突和工作 – 家庭增益，但是多数研究仍然将冲突和增益作为独立的现象来进行分析，少有研究将它们放在同一框架下进行探讨 (Maertz & Boyar, 2011; Powell & Greenhaus, 2006)。因此，本研究的目的是为了针对上述不足，拟以人 – 环境匹配为基础，以边界弹性为切入点，检验工作和家庭领域的边界弹性能力和个体的边界弹性意愿之间的匹配对工作 – 家庭冲突和工作 – 家庭增益的影响，以期更为全面地揭示冲突和增益发生的机制，为工作 – 家庭界面的组织实践和个人应对提供更为有效的意见和建议。

2 假设提出

2.1 弹性能力和弹性意愿之间匹配和工作 – 家庭冲突之间的关系

工作 – 家庭冲突指来自工作与家庭双方的需要在某些方面出现难以调和的矛盾时，个体所经历的一种角色交互冲突，包含着工作→家庭冲突 (work–family conflict, WFC) 和家庭→工作冲突 (family–work conflict, FWC) (Greenhaus & Beutell, 1985)。工作和家庭领域的相关变量对两种不同方向冲突的产生有着不同的效应量。研究表明工作领域的变量与工作→家庭冲突有着更强的相关，而家庭领域的变量与家庭→工作冲突有着更强的相关 (Byron, 2005; Eby et al., 2005)。尽管有元分析研究表明工作领域的变量与家庭→工作冲突有显著的相关以及家庭领域的变量与工作→家庭冲突有显著的相关，但是这些效应很微弱，工作领域的变量与家庭→工作冲突的相关在 –0.04~0.10 之间，家庭领域的变量与工作→家庭的相关在 –0.04~0.14 之间 (Michel, Mitchelson, Kotrba, LeBreton, & Baltes, 2009; Ng & Feldman, 2008)，所以本研究只分析工作弹性能力与工作弹性意愿之间的匹配对工作→家庭冲

突的影响，以及家庭弹性能力与家庭弹性意愿之间的匹配对家庭→工作冲突的影响。

人 – 环境匹配理论认为，当环境不能提供个体需求的条件或资源时，其需求将得不到满足，进而会导致紧张、消极情绪的产生以及冲突的体验(Edwards, 2008; Edwards & Rothbard, 1999; Jansen & Kristof-Brown, 2006)。基于此，在工作 – 家庭领域，一个工作边界弹性意愿高的个体，可能希望组织能够提供相应的资源 (如弹性工作时间或弹性工作地点) 使其能迅速进行角色的转变，当个体知觉到组织提供的资源使其能够从认知或行为上进行角色转变时，就会与个体转变角色的意愿相符，从而使得个体体验到较少的工作→家庭冲突；相反，当个体知觉到组织提供的资源不能使其顺利进行角色转变去满足家庭领域的需求时，即工作弹性能力低于其工作弹性意愿时，个体会体验到较高的工作→家庭冲突。另外，在实际的情境中，还可能会出现此种情况，即一个以工作为中心的个体，其工作弹性意愿可能会很低，倾向于工作时专注于工作任务，尽可能避免工作角色的中断，而此时若组织强制执行弹性工作政策 (弹性工作时间或弹性工作地点)，反而可能使其面临着更多的角色转变，进而引起紧张、体验到较高的冲突。这一点已经得到了相关研究的支持。已有研究发现对那些希望保持工作和家庭高分割的个体而言，当组织提供更多的整合策略时 (如日托中心，弹性工作时间、地点等)，他们会有着更为频繁的角色转变，会导致他们的工作和家庭角色模糊，进而使得他们有着更低的满意感、组织承诺和更高的工作→家庭冲突 (Desrochers, Hilton, & Larwood, 2005; Rothbard, Phillips, & Dumans, 2005)。所以，当个体知觉到工作边界弹性能力高于工作弹性意愿或低于工作弹性意愿时，均会存在一定的冲突，只有当个体知觉到其工作弹性能力与其工作弹性意愿相匹配时，工作→家庭冲突的体验才最低。同理，对个体的家庭弹性意愿而言，如果个体知觉到家庭弹性能力与其家庭弹性意愿正好匹配时，个体会有较低的家庭→工作冲突体验，而知觉到家庭弹性能力高于

或低于家庭弹性意愿时，个体均会有着较高的家庭→工作冲突体验。因此，我们提出以下假设：

H1：工作弹性能力和工作弹性意愿之间的匹配负向影响工作→家庭冲突。即当工作弹性能力朝向工作弹性意愿逐渐增加的时候，个体所体验到的工作→家庭冲突会逐渐降低，但是当工作弹性能力超过工作弹性意愿继续增加的时候，个体所体验到的工作→家庭冲突反而会逐渐上升。

H2：家庭弹性能力与家庭弹性意愿之间的匹配负向影响家庭→工作冲突。即当家庭弹性能力朝向家庭弹性意愿逐渐增加的时候，个体所体验到的家庭→工作冲突将会降低，但是当家庭弹性能力增加超过了家庭弹性意愿继续增加的时候，个体所体验到的家庭→工作冲突反而会逐渐上升。

2.2 弹性能力和弹性意愿之间的匹配和工作 – 家庭增益之间的关系

关于工作 – 家庭冲突产生的原因，研究者多用稀缺假说来解释，认为工作 – 家庭冲突的产生是由于个体的时间和精力等资源的有限而相互竞争所导致的 (Rothbard, 2001)。然而也有学者认为多重角色除了可能会导致角色紧张外，还可能给个体提供获得满意感和自我增益 (self-enrichment) 的机会 (Siber, 1974)。在积极心理学兴起之后，这一观点得以继续发展，研究者们认为个体在某一角色表现中所获得的收益可能大于其角色投入所造成的损失，一个角色所产生的资源 (如知识，技能和能力的发展、社会资本、弹性等) 会增加另外一个角色的表现，即工作 – 家庭增益 (Greenhaus & Powell, 2006; Carlson, Kacmar, Wayne, & Grzywacz, 2006)。与工作 – 家庭冲突类似，工作 – 家庭增益也包括工作→家庭增益 (work-family enrichment, WFE) 和家庭→工作增益 (family-work enrichment, FWE)，并且工作→家庭增益与工作相关的变量有着更强的相关，家庭→工作增益与家庭领域的变量有着更强的相关 (McNall, Nicklin, & Masuda, 2010)。所以，与工作 – 家庭冲突的分析类似，本研究只分析工作弹性能力与工作弹性意愿之间的匹配对工作→家庭增益的影响，以及

家庭弹性能力与家庭弹性意愿之间的匹配对家庭→工作增益的影响。

人 – 环境匹配理论认为，环境的供给或奖赏与个体的需求之间的匹配是以个体的需求是否得到满足为基础的，个体需求的满足影响着个体的态度和行为，当环境的供给与个人的需求之间达到匹配时，会增加个体的满意感、幸福感等积极的体验 (Edwards, 2008; Jansen & Kristof-Brown, 2006)。当个体知觉到组织环境所提供的资源正好与其角色转变意愿相一致时，个体可能会体验到工作满足感等积极情绪，而这种积极情绪可能会使个体更愿意履行家庭角色，进而增加家庭领域的积极体验。当个体知觉到组织环境提供的资源不能使其顺利进行角色转变时，即工作弹性能力低于其工作弹性意愿时，个体的满足感等积极情绪会减少，冲突、紧张等消极情绪会增加，进而不利于家庭角色的履行；同样，当个体知觉到组织所提供的条件或资源使得其工作和家庭领域的整合超出了其期望时，就会导致个体的工作满意感和组织承诺的降低 (Rothbard et al., 2005)，也不利于个体家庭角色的履行。同理，当个体知觉到家庭弹性能力和其家庭弹性意愿一致时，个体同样会体验到幸福、满足等积极情绪，这种积极情绪同样可能会使得个体有着更多的精力和更积极的情绪去履行工作角色，进而增加工作领域的积极体验，使得个体体验到较高家庭→工作的增益，当家庭弹性能力高于或低于家庭弹性意愿时，个体会体验到较低的家庭→工作增益。基于上述分析，我们提出以下假设：

H3：工作弹性能力和工作弹性意愿之间的匹配正向影响工作→家庭增益，即当工作弹性能力朝向工作弹性意愿逐渐增加的时候，个体所体验到的工作→家庭增益会逐渐上升，但是当工作弹性能力超过工作弹性意愿继续增加的时候，工作→家庭增益反而会逐渐降低

H4：家庭弹性能力与家庭弹性意愿之间的匹配正向影响家庭→工作增益。即当家庭弹性能力朝向家庭弹性意愿逐渐增加的时候，个体所体验到的家庭→工作增益将会逐渐上升，但是当家庭弹性能力超过

了家庭弹性意愿继续增加的时候，家庭→工作增益反而会逐渐降低

3 研究方法

3.1 研究对象

本研究采用问卷调查法获取研究数据，对上海某高校的 MBA 已婚学员及湖北、河南、上海、江苏、江西等多家企事业单位的已婚员工发放了问卷。参与调查的员工所在的企业涉及制造、IT、金融、电子等多个行业。问卷采用被调查者自评的方式，总共发放 525 份问卷，在删除无效问卷之后，最终回收有效问卷 494 份，有效回收率为 94.1%。其中男性 261 人，女性 230 人，被试年龄介于 24~53 岁之间，平均年龄 35.39 岁，样本中 348 人有 18 岁以下的小孩需要照顾。

3.2 研究工具

3.2.1 工作家庭边界弹性量表

结合中国的实际情况，对 Matthews 和 Barnes-Farrell (2010) 编制的工作家庭边界弹性量表进行了修订，形成了符合我国实际情况的工作家庭边界弹性量表，共计 16 个条目。该量表包含着工作弹性能力，工作弹性意愿，家庭弹性能力和家庭弹性意愿 4 个维度。问卷采用 Likert 5 点设计，1 表示完全不符合，5 表示完全符合。本研究中 4 个维度的内部一致性系数分别为 0.82，0.78，0.74，0.77，整个问卷的内部一致性系数为 0.82。对本研究中的工作家庭边界弹性量表进行验证性因素分析，结果表明 χ^2/df=3.23<4, GFI=0.92, CFI=0.92, IFI=0.92, NFI=0.88, RSMEA=0.067，说明工作家庭边界弹性量表具有良好的结构效度。

3.2.2 工作 – 家庭冲突量表

采用 Netemeyer 和 Boles (1996) 编制的工作 – 家庭冲突问卷。该量表包含工作→家庭冲突和家庭→工作冲突两个维度，每个维度 5 个条目。问卷采用 Likert 5 点设计，1 表示完全不符合，5 表示完全符合。该问卷的有效性在国内也得到了验证 (李永鑫，赵娜，2009)。本研究中两个维度的内部一致性系数分别为 0.87，0.87，整个问卷的内部一致性系数为 0.89。

3.2.3　工作 – 家庭增益量表

采用 Wayne, Musisca 和 Fleeson (2004) 编制的工作家庭增益量表。该量表包含工作→家庭增益和家庭→工作增益两个维度，每个维度 4 个条目。问卷采用 Likert 5 点设计，1 表示完全不符合，5 表示完全符合。该问卷的有效性在国内也得到了验证 (陈恒盼，2008)。本研究中两个维度的内部一致性系数分别为 0.72, 0.77, 整个问卷的内部一致性系数为 0.74。

3.2.4　控制变量的测量

本研究中，我们将年龄、工龄、工作时间、性别、是否有孩子和配偶的工作状态作为控制变量。其中年龄、工龄、工作时间等是连续测量，男性赋值为 1, 女性赋值为 2; 没有小孩赋值为 0, 有小孩赋值为 1; 配偶没有全职工作赋值为 0, 有工作赋值为 1。

3.3　统计分析

本研究采用 SPSS 18.0 统计软件包和 AMOS 18.0 统计软件包进行统计分析。首先使用验证性因素分析来验证问卷的效度，然后运用 SPSS 18.0 统计软件包进行描述性统计，并分析弹性能力与弹性意愿之间的匹配对工作 – 家庭结果变量的影响。

4　研究结果

4.1　验证性因素分析及共同方法偏差检验

本研究使用 AMOS 18.0 软件对工作弹性能力、工作弹性意愿、家庭弹性能力、家庭弹性意愿、工作→家庭冲突、家庭→工作冲突、工作→家庭增益、家庭→工作增益 8 个构念进行验证性因素分析，并将拟合指数与另外几个模型进行比较。验证性因素分析的具体结果见表 1。研究结果表明，八因子模型各拟合指标均达到了推荐的标准 (温忠麟，侯杰泰，马什赫伯特，2004), 且明显的优于其他备选模型，证明了这 8 个变量确实是 8 个不同的构念。同时，本研究采用了 Harman 的单因素因子分析来进一步的检验了共同方法偏差，结果表明未旋转时，共生成 8 个因子，解释了 64.83% 的变异，第一个因子解释了 18.16%

表 1　验证性因子分析结果 (N=494)

模型	χ^2	df	χ^2/df	RMSEA	GFI	CFI	IFI	NFI
单因素模型	5679.73	527	10.59	0.141	0.45	0.29	0.30	0.28
四因素模型 1	3545.21	521	6.80	0.11	0.60	0.59	0.59	0.55
四因素模型 2	3047.88	521	5.85	0.10	0.66	0.65	0.65	0.61
六因素模型 1	2311.88	512	4.52	0.09	0.71	0.75	0.75	0.71
六因素模型 2	2022.56	512	3.95	0.08	0.78	0.79	0.79	0.74
八因素模型	1276.66	499	2.56	0.056	0.86	0.89	0.89	0.84

注：单因素模型：指所有项目负荷在一个因子上；四因素模型 1：工作弹性能力 + 家庭弹性能力、工作弹性意愿 + 家庭弹性意愿、工作 – 家庭冲突、工作 – 家庭增益；四因素模型 2：工作弹性能力 + 工作弹性意愿、家庭弹性能力 + 家庭弹性意愿、工作 – 家庭冲突、工作 – 家庭增益；六因素模型 1：工作弹性能力 + 家庭弹性能力、工作弹性意愿 + 家庭弹性意愿、工作→家庭冲突、家庭→家庭冲突、工作→家庭增益、家庭→工作增益；六因素模型 2：工作弹性能力 + 工作弹性意愿、家庭弹性能力 + 家庭弹性意愿、工作→家庭冲突、家庭→家庭冲突、工作→家庭增益、家庭→工作增益；八因素模型：项目负荷在各自的理论维度上

表 2　各研究变量的平均数、标准差与相关矩阵 (N=494)

变量	M	SD	1	2	3	4	5	6	7	8
1.WFA	3.39	0.82	1.00							
2.WFW	3.16	0.75	0.45**	1.00						
3.FFA	3.83	0.69	0.36**	0.08	1.00					
4.FFW	3.61	0.76	0.17**	0.12*	0.51**	1.00				
5.WFC	2.68	0.88	−0.23**	0.05	−0.09*	0.14*	1.00			
6.FWC	2.20	0.81	−0.13**	0.18**	−0.29**	−0.07	0.55**	1.00		
7.WFE	3.18	0.73	0.16**	0.12*	0.23**	0.19**	−0.04	0.03	1.00	
8.FWE	3.87	0.61	0.15**	0.14**	0.36**	0.31**	−0.01	−0.22**	0.28**	1.00

注：** $p<0.01$, * $p<0.05$ 其中 WFA 代表工作弹性能力，WFW 代表工作弹性意愿，FFA 代表家庭弹性能力，FFW 代表家庭弹性意愿，WFC 代表工作→家庭冲突，FWC 代表家庭→工作冲突，WFE 代表工作→家庭增益，FWE 代表家庭→工作增益

的方差变异，远小于 Harrison, McLaughlin 和 Coalter (1996) 推荐的 50% 的判断标准，表明本研究中共同方法偏差并不严重。

4.2 描述性统计分析

表 2 呈现了本研究中所有变量的均值、标准差及变量之间相关系数的结果。从表 2 可以得知工作弹性能力与工作→家庭冲突呈显著的负相关，与工作→家庭增益呈显著的正相关；家庭弹性能力与家庭→工作冲突呈显著的负相关，与家庭→工作增益呈显著的正相关，家庭弹性意愿与家庭→工作增益呈显著正相关。

4.3 假设检验

以往关于人 – 环境匹配的研究中，关于两个构念之间匹配程度的衡量主要采用差异分数 (difference scores)，即通过个体在同一属性量表分别评估自己的愿望 (或能力) 和环境的回报 (或要求)，然后计算两个分数的差值，具体的形式有代数差 (X−Y)、差的绝对值 (∣ X−Y ∣)、差的平方 (X−Y)²。但是它们只能反映匹配的整体特征，无法区分超过或不及情况
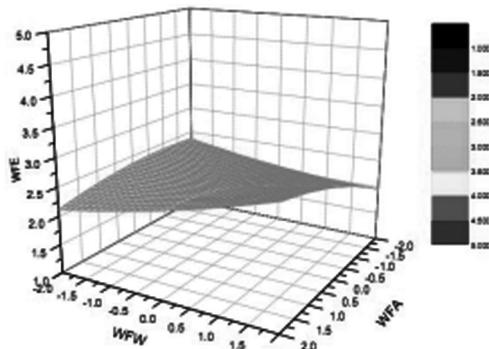
下对结果变量的影响，并且存在着容易丢失信息、对差异性来源不敏感、无法反映单个变量的贡献率等缺点。相关研究者指出人、环境和结果变量之间的关系应被看作 3 个方面，提出应该使用多项式回归 (polynomial regression) 来产生一个三维反应面来代表 P 和 E 的整合关系对结果变量的影响 (Edward & Parry, 1993; Edward & Harrison, 1993; Edward & Rothbard, 1999, 2000)。基于此，我们沿用了 Edwards 所提出来的多项式回归来分析边界弹性能力与边界弹性意愿之间的匹配对工作家庭界面的影响。具体而言，每一个结果变量分别对弹性能力 (A)、个体的弹性意愿 (W)、能力和意愿的交互 (A*W)、弹性能力的平方 (A²)、弹性意愿的平方 (W²) 进行回归，回归方程中的 5 个项目即包含了变量线性匹配对结果变量的影响，同时也包含着曲线匹配效应。具体的表达方程式如下：$Z = b_0 + b_1X + b_2Y + b_3Y^2 + b_4XY + b_5X^2 + e$。其中 Z 代表工作→家庭冲突，家庭→工作冲突，工作→家庭增益和家庭→工作增益；X 代表弹性意愿分数；Y 代表弹性能力分数。为了避免高的共线


图 1　WFA 与 WFW 匹配对 WFC 的影响
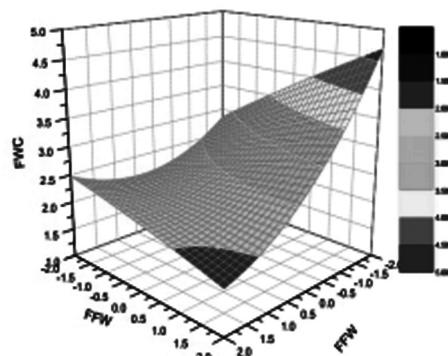
图 2　FFA 与 FFW 匹配对 FWC 的影响


图 3　WFA 与 WFW 匹配对 WFE 的影响
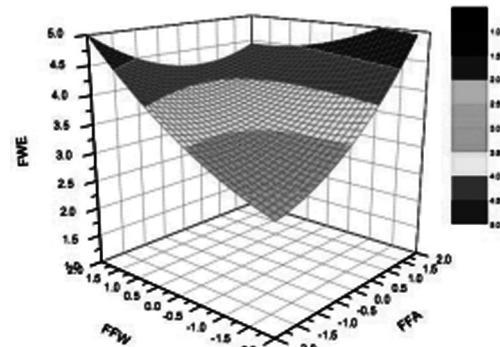
图 4　FFA 与 FFW 匹配对 FWE 的影响

性，我们将所有自变量进行了中心化处理。

在具体的分析过程中，我们均采用了多层回归的方法，将控制变量放在第一层 (M_1)；将边界弹性能力和边界弹性意愿放在第二层，检验边界弹性能力和边界弹性意愿的主效应 (M_2)；将边界弹性能力的平方、边界弹性意愿的平方、以及弹性意愿和弹性能力的交互项纳入到回归方程第三层 (M_3)。在纳入了变量的高阶项之后，R^2 显著地增加，或者方程中的高阶项系数显著，即说明匹配和结果变量之间存在着非线性关系，此时，相关的具体结果可以通过反应曲面方法 (response surface methodology) 呈现出三维图形。在三维图形中，在"弹性意愿 = 弹性能力"对角线曲面的形状表明当个体的弹性能力与其弹性意愿相匹配时结果变量的情形。"弹性意愿 = − 弹性能力"对角线的曲面的形状表明当个体的弹性能力与弹性意愿不匹配时结果变量的变化情况，而所有的假设均是通过此曲面的发展趋势来进行检验的。先计算 $x_1 = b_1 - b_2$，$x_2 = b_3 - b_4 + b_5$，进而对 x_1 和 x_2 进行显著性检验，如果 x_1

显著，则表明沿着"弹性意愿 = − 弹性能力"对角线方向有线性关系，如果 x_2 显著，则表明沿着"弹性意愿 = − 弹性能力"对角线方向有一个曲面 (Edwars & Parry, 1993)，此时可以通过计算曲面的第一主轴和第二主轴来判断曲面的发展趋势。同时，本研究还计算出 $a_1 = b_1 + b_2$，$a_2 = b_3 + b_4 + b_5$ 来进一步的分析"弹性意愿 = 弹性能力"对角线曲面的发展趋势。当 $a_1 > 0$ 且显著时表明高弹性意愿和高弹性能力的匹配比低弹性意愿和低弹性能力的匹配有着更为积极的结果，当 $a_1 < 0$ 且显著时，表明高弹性意愿和高弹性能力的匹配比低弹性意愿和低弹性能力的匹配有着更为消极的结果。当 a_2 显著时，表明"弹性意愿 = 弹性能力"对角线方向有一个曲面。

本研究详细的结果见表 3，除了工作→家庭增益的二次项回归方程中的二次项系数没有达到显著性以外，其他的回归方程中均有二次项系数达到显著。我们对表 2 中的方程运用 origin 8.0 绘图软件绘制出了三维图。具体图形见图 1、图 2、图 3 和图 4。在

表 3 二次项回归方程结果 (N=494)

变量	工作→家庭冲突 (WFC)			工作→家庭增益 (WFE)			家庭→工作冲突 (FWC)			家庭→工作增益 (FWE)		
	M_1	M_2	M_3	M_1	M_2	M_3	M_1	M_2	M_3	M_1	M_2	M_3
常数项	2.37	2.37	2.42	2.47	2.42	2.40	2.73	2.39	2.46	3.25	3.62	3.66
控制变量	--	--	--	--	--	--	--	--	--	--	--	--
WFA		−0.35**	−0.34**		0.11*	0.10*						
WFW		0.23**	0.23**		0.05	0.06						
WFA^2			0.07			−0.03						
WFW^2			0.17**			0.02						
WFA*WFW			−0.29**			0.08						
FFA								−0.44**	−0.41**		0.26**	0.26**
FFW								0.14*	0.14*		0.12**	0.15**
FFA^2									0.13*			0.12*
FFW^2									−0.01			0.07+
FFA*FFW									−0.16*			−0.20**
R^2	0.06	0.14	0.18	0.03	0.05	0.06	0.05	0.15	0.16	0.04	0.18	0.21
F	2.18	5.11	5.44	1.16	1.78	1.62	1.86	5.42	4.91	1.66	6.97	6.80
△R^2		0.08	0.04		0.02	0.01		0.10	0.01		0.14	0.03
A=W 对角线	$a_1=b_1+b_2=-0.11$			$a_1=b_1+b_2=0.16*$			$a_1=b_1+b_1=-0.27**$			$a_1=b_1+b_2=0.41**$		
	$a_2=b_3+b_4+b_5=-0.05$			$a_2=b_3+b_4+b_5=0.07$			$a_2=b_3+b_4+b_5=-0.04$			$a_2=b_3+b_4+b_5=-0.01$		
A=−W 对角线	$x_1=b_1-b_2=0.57**$			$x_1=b_1-b_2=-0.04$			$x_1=b_1-b_2=0.55**$			$x_1=b_1-b_2=-0.11*$		
	$x_2=b_3-b_4+b_5=0.53**$			$x_2=b_3-b_4+b_5=-0.09$			$x_2=b_3-b_4+b_5=0.28**$			$x_2=b_3-b_4+b_5=0.39**$		

注：** $p<0.01$，* $p<0.05$ + $p<0.10$，表中 a_1 和 a_2 代表着"弹性能力 = 弹性意愿"对角线的斜率和曲率，x_1 和 x_2 代表着"弹性能力 = − 弹性意愿"的对角线的斜率和曲率，b_1 代表弹性意愿的系数；b_2 代表着弹性能力的系数，b_3 代表这弹性能力平方的系数，b_4 代表着弹性能力 * 弹性意愿的系数，b_5 代表着弹性意愿平方的系数

图形中，在弹性能力 = 弹性意愿对角线的右边区域表示弹性能力低于弹性意愿，在对角线的左侧则表示弹性能力高于弹性意愿的情况。

4.3.1 假设 1 的检验

根据表 3 中所呈现的工作→家庭冲突的二次项回归方程，以及图 1 所呈现的三维反应曲面，我们发现当工作弹性能力朝向工作弹性意愿逐渐增加时（从 WFW=−WFA 对角线最右端向最左端移动的时候的反应曲面），工作→家庭冲突不断下降；当工作弹性能力超过了工作弹性意愿继续增加时（即从反面曲面的中心移向最左端时），工作→家庭冲突呈上升的趋势。斜率的分析证实了图 1 的结果，在"WFW= −WFA"对角线的形状有一个显著的曲线形状 ($x_2=b_3-b_4+b_5= 0.53$, $p<0.01$)，第一主轴的斜率为 −1.4，第二主轴的斜率为 0.72。进一步的分析发现沿着"WFW=WFA"对角线的趋于一条直线，斜率分析证实了这一结果（$a_1= b_1+b_2 =-0.11$, $p>0.10$; $a_2= b_3+b_4+b_5 =-0.05$, $p>0.10$），说明高工作弹性能力 (WFA) 和高工作弹性意愿 (WFW) 匹配情况下的工作→家庭冲突与低工作弹性能力和低工作弹性意愿匹配情况下的工作→家庭冲突没有显著的差异。假设 1 得到验证。

4.3.2 假设 2 的检验

根据表 3 中所呈现的家庭→工作冲突的二次项回归方程，以及图 2 所呈现的反应曲面，我们可以发现当家庭弹性能力朝向家庭弹性意愿增加时（从 FFW=−FFA 对角线最右端向最左端移动的反应曲面），家庭→工作冲突逐渐降低；当家庭弹性能力超过了家庭弹性意愿继续增加时（即从反应曲面的中心移向最左端的时候），家庭→工作冲突有一个缓慢上升的趋势。斜率的分析证实了此结果，在"FFW=−FFA"对角线的形状有一个显著的曲线形状 ($x_2=b_3-b_4+b_5= 0.28$, $p<0.01$)，第一主轴的斜率为 −0.44，第二主轴的斜率为 2.18。假设 2 得到验证。进一步的分析发现沿着"FFW=FFA"的对角线上，家庭→工作冲突的发展趋势是一条直线 （$a_1=b_1+b_2=-0.27$, $p<0.01$; $a_2= b_3+b_4+b_5=-0.04$, $p>0.10$），且高家庭弹性意愿和高家庭弹性能力匹配情境下家庭→工作冲突比低家庭弹性意愿和低家庭弹性能力匹配情境下的家庭→工作冲突更低。

4.3.3 假设 3 的检验

根据表 3 中所呈现的工作→家庭增益的二次项回归方程，我们发现二次项的系数均不显著，且 $x_1=b_1-b_2=0.04$, $p>0.05$; $x_2= b_3-b_4+b_5=0.09$, $p>0.05$，表明沿着"弹性意愿 = 弹性能力"对角线，整个反应曲面基本上一条直线，此时，工作弹性能力的系数显著，说明工作弹性能力对工作→家庭增益的主效应显著，即随着工作弹性能力的增加，工作→家庭增益也随之增加，这一点在图 3 中也得到证实，如图 3 所示，工作→家庭增益的反应面基本上是一个平面。假设 3 并未得到验证。

4.3.4 假设 4 的检验

根据表 3 中所呈现的家庭→工作增益的二次项回归方程，以及图 4 所呈现的反应曲面，我们发现当家庭弹性能力朝向家庭弹性意愿增加时（从 FFW=−FFA 对角线最右端向最左端移动的反应曲面），此时我们发现家庭→工作增益并非如果我们所期望的那样逐渐增加，而是逐渐降低，当家庭弹性能力超过了家庭弹性意愿继续增加时（即从反应曲面的中心移向最左端时），家庭→工作增益 (FWE) 同样也并非如我们所期望的那样逐渐降低，反而逐渐上升。假设 4 未得到验证。进一步的分析发现沿着"FFW=FFA"的对角线上，家庭→工作增益的发展趋势是一条直线 （$a_1=b_1+b_2=0.41$, $p<0.01$; $a_2=b_3+b_4+b_5=-0.01$, $p>0.10$），且说明高家庭弹性意愿和高家庭弹性能力匹配情境下的家庭→工作增益比低家庭弹性意愿和低家庭弹性能力匹配情境下的家庭→工作增益更高。

5 讨论

5.1 边界弹性能力和边界弹性意愿的匹配对工作 − 家庭冲突的影响

本研究的结果表明工作（或家庭）弹性能力和工作（或家庭）弹性意愿之间的匹配有助于降低员工的工作→家庭冲突（家庭→工作冲突）。从图 1 和图 2

中我们可以看出，当工作 (或家庭) 弹性意愿很高，而其工作 (或家庭) 弹性能力很低的时候，此时个体会体验到最为强烈的工作→家庭冲突 (家庭→工作冲突)；而当个体的弹性能力趋近其弹性意愿时，工作→家庭冲突 (家庭→工作冲突) 呈下降的趋势，当弹性能力超过弹性意愿进一步增加时，其冲突的体验又会有所上升。此结果表明，双向的工作 – 家庭冲突不仅受外部环境资源的影响，还受到个体特征的影响，并且它们还交互作用共同决定了冲突的发生，因此，这一结果比较直观地揭示了工作 – 家庭冲突的发生机制。具体来说，当个体某一领域的弹性意愿很高，而个体知觉到其所在的环境不能提供进行相应的角色转变的资源时，如组织政策不允许，领导不支持或家人反对等，就会导致个体心理紧张以及冲突体验的产生；而当个体在某一领域的弹性意愿很低，此时个体可能倾向于在此领域将工作和家庭分割开来，不愿意将工作和家庭角色进行高度整合，而此时若个体知觉到此领域有着较高的弹性能力，即环境提供的资源会促进领域整合时，可能会导致个体产生较高的角色模糊，进而使个体工作 – 家庭冲突的体验增强 (Glavin & Schieman, 2012)。

5.2 边界弹性能力和边界弹性意愿的匹配对工作 – 家庭增益的影响

本研究结果表明工作弹性能力和工作弹性意愿之间的匹配对工作→家庭增益的作用不显著，只有工作弹性能力能够显著预测个体的工作→家庭增益；而家庭弹性能力和家庭弹性意愿的匹配对家庭→工作增益的影响与假设完全相反，即当家庭弹性能力逐渐向家庭弹性意愿增加的时候，家庭→工作增益呈下降趋势，当家庭弹性能力超过家庭弹性意愿继续增加时，家庭→工作增益逐渐上升，在家庭弹性能力和家庭弹性意愿匹配的情况下，只有当个体的家庭弹性能力和意愿均很低的时候，家庭→工作增益最低。通过图 4 不难发现，只有当个体的家庭弹性能力和家庭弹性意愿均很低的时候，家庭→工作增益最低，随着家庭弹性能力或家庭弹性意愿单方面增加或两者同时增加时，家庭→工作增益也随之增加。这说明了家庭

弹性能力和家庭弹性意愿均能显著地增加个体的家庭→工作增益。究其原因，可能是在我们国家人们普遍持有一种工作优先的行为规范 (张勉，李海，魏钧，杨百寅，2011)，认为工作不仅是个人的事情，而是提升家庭整体利益和荣耀的手段，勤奋工作是一种有家庭责任感的表现 (Aryee, Field, & Luk, 1999; Wang, Lawler, Walumbwa, & Kan, 2004; 李晔，2003)。个体无论在认知上，还是在行为上均可能聚焦在如何更好地履行工作角色上，即使组织提供再多的资源让其能够更容易转变工作角色，个体通常也不太关注家庭角色的履行，因此工作→家庭增益变化不大。反而，只要个体持工作优先规范的原则，或家人给予支持能够让其顺利转变家庭角色去更好地履行工作职责时，均会出现较高的家庭→工作增益，而只有个体在认知和行为上均不以工作优先，且没有足够的资源去转变角色时，才会出现较低的家庭→工作增益。为了更清楚地揭示工作 – 家庭增益的发生机制，将来的研究应充分考虑我国工作优先的规范对员工工作 – 家庭增益的影响，可以通过将深度访谈等质化研究的方法和问卷调查等量化研究方法进行有机的结合，对上述的推论进行进一步的验证。

5.3 理论和实践意义

本研究的理论贡献主要表现为：(1) 从人 – 环境匹配的视角来探索了工作和家庭领域的边界弹性能力和弹性意愿对工作 – 家庭界面结果变量的交互影响，弥补了以往研究多关注情境需求而忽视个体特征同样在其工作 – 家庭管理中发挥着重要作用的不足；而且，同时考虑工作和家庭环境与个人需求的匹配对双向的工作 – 家庭冲突和增益的影响，发现了个体在工作家庭领域的弹性意愿和其知觉到的弹性能力对家庭→工作冲突和增益的交互影响，扩充了以往研究仅关注工作环境和个人需求之间匹配的研究模式，并通过多项式回归和三维反应曲面将工作 – 家庭冲突和增益的发展趋势清晰地展示出来，证实了个体在实际的生活中，工作 – 家庭冲突和工作家庭增益同时存在的假设。通过比较图 1 和图 3，当工作弹性能力朝向工作弹性意愿逐渐增加并进一步超过工作弹性意

愿时，工作→家庭冲突先下降而后呈逐渐上升趋势，而工作→家庭增益的变化趋势却始终不显著；通过比较图2和图4，当家庭弹性能力朝向家庭弹性意愿逐渐增加并进一步超过家庭弹性意愿时，家庭→工作冲突和家庭→工作增益有着同样的变化趋势，均为先逐渐降低而后逐渐升高。以上结果表明，对低家庭弹性能力和高家庭弹性意愿的个体而言，同时有着高的家庭→工作冲突和家庭→工作增益，这一发现为我们更清晰地理解工作－家庭平衡提供了帮助；(2) 本研究为前期研究成果中不一致的结论提供了可能的解释。Kossek 和 Nichol (1992) 研究表明弹性的工作制度、日托中心等家庭友好政策会导致员工工作－家庭冲突显著减少，而 Solomon (1994) 研究却发现家庭友好政策对员工的工作－家庭冲突的影响不大，Desrochers 等 (2005) 研究发现家庭友好政策会增加员工的角色模糊，进而导致员工的工作－家庭冲突增加。基于本文人－环境匹配视角的研究结果，导致上述研究结果不一致的原因可能是组织在实施相关的家庭友好政策时，忽略了在工作－家庭界面中与组织政策共同起作用的个人因素，特别是个体对工作和家庭所持有的价值观、态度和期望的影响作用。

本研究结果为组织管理实践提供了一定的指导。研究发现个体工作弹性能力与工作弹性意愿的匹配会有效地降低其工作→家庭冲突的体验，而不匹配则会导致个体工作→家庭冲突，特别是当个体的弹性意愿很高而其感知到的工作弹性能力很低的时候尤为明显。这提示组织在制定工作和家庭的相关政策制度时，要切实考虑到不同员工的实际需求，做到组织提供的资源和员工个人需求相匹配，而不能"一刀切"式地强制执行某一固定的工作－家庭政策。其次，对于组织内部的员工，要积极宣传组织相关的工作－家庭政策，通过社会化的过程改变员工对工作和家庭的偏好，进而改变其意愿。再次，组织在招聘新进员工时，要尽量录用与组织所能提供的工作家庭边界管理文化相匹配的员工。另外，本研究结果也为促进员工的家庭边界管理提供了途径。通过比较图2和图4，我们不难发现当个体在家庭弹性能

力很低，而家庭弹性意愿很高的情况下，个体会同时体验到较高的家庭→工作增益和家庭→工作冲突；但是当家庭弹能力愿很高而家庭弹性意愿很低的情况下，个体所体验到的家庭→工作增益很高、家庭→工作冲突的增幅却很小；并且当家庭弹性能力和家庭弹性意愿均很高时的家庭→工作冲突要显著低于家庭弹性能力和家庭弹性意愿均很低时的家庭→工作冲突，而家庭→工作增益则相反。这说明了家庭弹性能力在缓解家庭→工作冲突和增加家庭→工作增益上具有独立的效能，所以，家庭成员应该尽可能的给员工提供生活上的支持，减少其在家庭中的负担，当工作有需求时，使其能够顺利及时地进行角色上的转变，这样才能够在有效降低其冲突体验的同时，增加其增益的体验。

5.4 局限与展望

本研究的局限主要表现在以下几方面。首先，本研究的各变量均采用员工自评的方式来进行，对于工作→家庭冲突和工作→家庭增益而言，个体的感受可能与家人存在着一定的差异，对家庭→工作冲突和家庭→工作增益而言，个体的感受也可能与上司或同事的感受存在一定的差异，所以在未来的研究中应该采用多源评价的方式来全面地测量个体的边界弹性能力和工作－家庭界面的结果变量。其次，本研究对工作和家庭环境的测量采用了个体弹性能力这一主观的评估，可能存在同一环境中的不同个体对其所处的环境的感知有所不同的情况，所以在将来的研究中应该检验组织的实际工作－家庭政策、主管和家人实际的行为等与个人弹性意愿的匹配对个体工作－家庭界面的影响。再次，本研究采用了横切面研究，不能进行因果关系的推断，而且工作－家庭冲突和工作－家庭增益在实际的生活中并非一成不变，而是随着不同的时间段有着不同的冲突和增益水平 (Maertz, Jr & Boyar, 2011)，所以在将来的研究中，可以采用纵向追踪或者是日志的研究方法对这些变量的因果关系进行进一步的探讨，以更为清晰、全面地揭示工作－家庭冲突和增益的发生机制。

参考文献

Allen, T. D., Johnson, R. C., Saboe, K. N., Cho, E., Dumani, S., & Evans, S. (2012). Dispositional variables and work–family conflict: A meta–analysis. *Journal of Vocational Behavior, 80,* 17–26.

Ashforth, B. E., Kreiner, G. E., & Fugate, M. (2000). All in a day's work: Boundaries and micor role transitions. *Academy of Management Review, 25,* 472–491.

Aryee, S., Field, D., & Luk, V. (1999). A cross–cultural test of a model of the work–family interface. *Journal of Management, 25,* 491–511.

Byron, K. (2005). A meta–analytic review of work–family conflict and its antecedents. *Journal of Vocational Behavior, 67,* 169–198.

Carlson, D. S., Kacmar, K. M., Wayne, J. H., & Grzywacz, J. G. (2006). Measuring the positive side of the work–family interface: Development and validation of a work–family enrichment scale. *Journal of Vocational Behavior, 68,* 131–164.

Chen, H. P. (2008). The relationship among core self–evaluations and emloyee's satisfaction, turnover intentions: The mediating role of work–family facilitation. Unublished master's thesis, Zhejiang University.

[陈恒盼 . (2008). 核心自我评价与员工满意感、离职意向的关系研究 : 工作家庭促进的中介作用 . 硕士学位论文 , 浙江大学 .]

Chen, Z., Powell, G. N., & Greenhaus, J. H. (2009). Work–to–family conflict, positive spillover, and boundary management: A person environment fit approach. *Journal of Vocational Behavior, 74,* 82–93.

Clark, S. C. (2000). Work/Family border theory: A new theory of work/family balance. *Human Relations, 53,* 747–770.

Desrochers, S., Hilton, J. M., & Larwood, L. (2005). reliminary validation of the work–family integration–blurring scale. *Journal of Family Issues, 26,* 442–466.

Eby, L. T., Caser, W. J., Lockwood, A., Bordeaux, C., & Brinley, A. (2005). Work and family research in IO/OB: Content analysis and review of the literature (1980–2002). *Journal of Vocational Behavior, 66,* 124–197.

Edwards, J. R. (2008). Person environment fit in organizations: An assessment of theoretical progress. *The Academy of Management Annals, 2,* 167–230.

Edwards, J. R., & Harrison, R. V. (1993). Job demands and worker health: Three dimensional reexamination of the relationship between person environment fit and strain. *Journal of Alied sychology, 78,* 628–648.

Edwards, J. R., & Parry, M. E. (1993). On the use of olynomial regression equations as an alternative to difference scores in organizational research. *Academy of Management Journal, 36,* 1577–1613.

Edwards, J. R., & Rothbard, N. P. (1999). Work and family stress and well–being: An examination of person environment fit in the work and family domains. *Organizational Behavior and Human Decision processes, 77,* 85–129.

Edwards, J. R., & Rothbard, N. P. (2000). Mechanisms linking work and family: Clarifying the relationship between work and family constructs. *Academy of Management Review, 25,* 178–198.

Ford, M. T., Heinen, B. A., & Langkamer, C. L. (2007). Work and family satisfaction and conflict: A meta–analysis of cross domain relations. *Journal of Alied sychology, 92,* 57‒80.

Glavin, P., & Schieman, S. (2012). Work–family role blurring and work–family conflict: The moderating influence of job resources and job demands. *Work and Occu ations, 39,* 71–98.

Greenhaus, J. H., & Beutell, N. J. (1985). Sources of conflict between work and family roles. *Academy of Management Review, 10,* 76–88.

Greenhaus, J. H., & Powell, G. N. (2006). When work and family allies: A theory of work–family enrichment. *Academy of Management Review, 31*, 72–91.

Greenhaus, J. H., Ziegert, J. C., & Allen, T. D. (2012). When family supportive supervision matters: Relations between multiple source of support and work–family balance. *Journal of Vocational Behavior, 80*, 266－275.

Harrison, D. A., McLaughlin, M. E., & Coalter, T. M. (1996). Context, cognition, and common method variance: psychometric and verbal protocol evidence. *Organizational Behavior and Human Decision processes, 68*, 246－261.

Jansen, K. J., & Kristof, B. A. (2006). Toward a multidimensional theory of person environment fit. *Journal of Management Issues, 18*, 193–212.

Kossek, E. E., Lautsch, B., & Eaton, S. C. (2004). Flexibility enactment theory: Relationships between type, boundaries, control and work family effectiveness. In E. E. Kossek & S. J. Lambert (Eds.), *Work and life integration: Organizational, cultural and individual persectives* (pp243–261). Mahwah, NJ: Erlbaum.

Kossek, E. E., & Nichol, V. (1992). The effects of on site child care on employee attitudes and performance. *personnel psychology, 45*, 485–509.

Kreiner, G. E. (2006). Consequences of work home segmentation or integration: A person environment fit perspective. *Journal of Organizational Behavior, 27*, 485–507.

Kristof–Brown, A. L., Zimmerman, R. D., & Johnson, E. C. (2005). Consequences of individual's fit at work: A meta–analysis of personpjob, personporganization, personpgroup, and personpsupervisor fit. *personnel psychology, 58*, 281–342.

Li, Y. (2003). Empirical research of Work–family conflict in China. *Ergonomics, 9*, 14–17.

[李晔 . (2003). 工作－家庭冲突的影响因素研究 . 人

类工效学 , 9, 14–17.]

Li, Y. X., & Zhao, N. (2009). Structure and measurement of work–family support and its moderation effect. *Acta psycholigica Sinica, 41*, 863–874.

[李永鑫 , 赵娜 . (2009). 工作－家庭支持的结构与测量及其调节作用 . 心理学报 , 41, 863–874.]

Ma, L., & Xu, C. W. (2011). Boundary management and work–family interface: From personpenvironment fit perspective. *Nankai Business Review, 14,* 41–47.

[马丽 , 徐枞巍 . (2011). 基于个人－环境匹配理论的边界管理与工作家庭界面研究 . 南开管理评论 , 14, 41–47.]

Maertz, Jr. C. –., & Boyar, S. (2011). Work–family conflict, enrichment, and balance under "levels" and "Episodes" approaches. *Journal of Management, 37*, 68–98.

Matthews, R. A., & Barnes–Farrell, J. L. (2010). Development and initial evaluation of an enhanced measure of boundary flexibility for the work and family domains. *Journal of Occupational Health psychology, 15*, 330–346.

Matthews, R. A., Barnes–Farrell, J. L., & Bulger, C. A. (2010). Advancing measurement of work and family domain boundary characteristics. *Journal of Vocational Behavior, 77*, 447–460.

McNall, L. A., Masuda, A. D., & Nicklin, J. M. (2010). Flexible work arrangements, job satisfaction, and turnover intentions: The mediating role of work–to–family enrichment. *The Journal of psychology, 144*, 61－81.

McNall, L. A., Nicklin, J. M., & Masuda, A. D. (2010). A meta–analytic review of the consequences associated with work–family enrichment. *Journal of Business psychology, 25*, 381–396.

Michel, J. S. Mitchelson, J. K., Kotrba, L. M., LeBreton, J. M., & Baltes, B. B. (2009). A comparative test of workpfamily conflict models and critical examination of

work–family linkages. *Journal of Vocational Behavior, 74*, 199–218.

Netemeyer, R. G., & Boles, J. S. (1996). Develo–ment and validation of work–family conflict and family–work conflict scales. *Journal of Applied psychology, 81*, 400–410.

Ng, T., & Feldman, D. (2008). Long work hours: A social identity perspective on meta–analysis data. *Journal of Organizational Behavior, 29*, 853–880.

Odle–Dusseau, H. N., Britt, T. W., & Greene–Shortridge, T. M. (2012). Organizational workpfamily resource as predictors of job performance and attitudes: The process of work pfamily conflict and enrichment. *Journal of Occupational health psychology, 17*, 28–40.

Owell, G. N., & Greenhaus, J. H. (2006). Is the opposite of positive negative? Untangling the complex relationship between work–family enrichment and conflict. *Career Development International, 11*, 650－659.

Rothbard, N. P. (2001). Enriching or depleting? The dynamics of engagement in work and family roles. Administrative Science Quarterly, 46, 655–684.

Rothbard, N. P., Philips, K. W., & Dumas, T. L. (2005). Managing multiple roles: Work–family policies and individuals' desires for segmentation. *Organization Science, 16*, 243－258.

Siber, S. D. (1974). Toward a theory of role accumulation. *American Sociological Review, 39*, 567–578.

Solomon, C. (1994). Work/family's failing grade: Why today's initiatives are not enough. *personnel Journal, 72*, 72–87.

Straub, C. (2012). Antecedents and organizational consequences of family supportive supervisor behavior: A multilevel conceptual framework for research. *Human Resource Management Review, 22*, 15－26.

Wang, P., Lawler, J. J., Walumbwa, F. O., & Kan, S. (2004). Work–family conflict and job withdrawal intentions: The moderating effect of cultureal differences. *International Journal of Stress Management, 11*, 392–414.

Wayne, J. H., Musisca, N., & Fleeson, W. (2004). Considering the role of personality in the work–family experience: Relationships of the big fve to work－family confict and facilitation. *Journal of Vocational Behavior, 64*, 108–130.

Wen, Z. L., Hau, K. T., & Herbert, W. M. (2004). Structural equation model testing: cutoff criteria for goodness of fit indices and Chi–square test. *Acta sycholigica Sinica, 36*, 186–194.

[温忠麟 , 侯杰泰 , 马什赫伯特 . (2004). 结构方程模型检验 : 拟合指数与卡方准则 . *心理学报 , 36*, 186–194.]

Zhang, M., Li, H., Wei, J., & Yang, B. Y. (2011). Cross-over effects or direct effects? The mechanism linking work–family conflict with outcomes. *Acta sychologica Sinica, 43*, 573–588.

[张勉 , 李海 , 魏钧 , 杨百寅 . (2011). 交叉影响还是直接影响？工作－家庭冲突的影响机制 . *心理学报 , 43*, 573–588.]

Boundary flexibility and work-family interface: From person-environment fit perspective

MA Hongyu[1,2]; SHEN Chuangang[1,2]; YANG Jing[3,4]; TANG Hanying[1,2]; XIE Julan[1,2]

([1]Key Laboratory of Adolescent Cyberpsychology and Behaivor (CCNU), Ministry of Education, Wuhan 430079, China)

([2]School of Psychology, Central China Normal University, and Hubei Human Development of Mental Health Key Laboratory, Wuhan 430079, China)

([3]Institute of Psychology, Chinese Academy of Sciences, Beijing, 100101, China)

([4]University of Chinese Academy of Sciences, Beijing 100039, China)

Abstract Work and family are two important domains in an individual's life. How to balance work and family domains have become an increasingly compelling and pressing issue for both the organizational scholars interested in theoretical advances, and for human resources practitioners seeking to promote the employee's daily life. Individuals negotiate the boundaries between work and family in their daily activities. There are differences between the individual's preference and the resource that provided by the organization in boundary management of work and family domain. In our study, using a person-environment (PE) fit theoretical base, we explored how the interaction between an individual's boundary flexibility willingness and the perceived flexibility ability of the domain boundary affects work-family conflict and work - family enrichment. Specifically, we predict that the fit of domain boundary flexibility-ability and individual's domain boundary flexibility-willingness would be associated with lower work-family conflict and higher work-family enrichment. Data were collected from a sample of 494 fulltime married employees from different industries. The questionnaire for employee included work-family boundary flexibility scale, work-family conflict and work-family enrichment. Among the major measures, the 16-items boundary flexibility scale was adopted from Matthews and Barnes-Farrell (2010), WFC was measured via 10 items that was adopted from Netemeyer and Boles (1996), the eight item WFE scale was adopted from Wayne, Musisca and Fleeson (2004). Results show that the Cronbach's alpha coefficients for the above measures range from 0.72 to 0.89. Polynomial regression and response surface methodology were utilized to examine the proposed hypotheses. In line with the predictions, results of polynomial regression and response surface methodology demonstrate that work-to-family conflict decreased as work flexibility-ability (WFA) approached work flexibility-willingness (WFW), and increased as WFA exceeded WFW, family-to-work conflict decreased as family flexibility-ability (FFA) approached family flexibility-willingness (FFW), and increased as FFA surpassed FFW. The results also showed that the fit of WFA and WFW has no effect on work-to-family enrichment, and the fit of FFA and FFW has the significant effect on family-to-work enrichment, but it is opposite to the hypotheses. Specifically, family-to-work enrichment decreased as FFA approaching FFW, and increased as FFA exceeded FFW. The present study extends to our understanding the mechanism of the process of the work-family conflict and work-family enrichment happens. Finally, the theoretical and managerial implications of the

findings, limitations and future research directions were also discussed.

上司不当督导与下属绩效：反馈寻求行为和学习目标定向的作用 *

申传刚 [1,2]，马红宇 [1,2]，杨 璟 [3,4]，刘腾飞 [1,2]

([1] 青少年网络心理与行为教育部重点实验室)

([2] 华中师范大学心理学院暨湖北省人的发展与心理健康重点实验室 , 武汉 430079)

([3] 中国科学院心理研究所 , 北京 100101)

([4] 中国科学院研究生院 , 北京 100039)

摘 要 本研究从下属反馈管理行为的视角来探索领导与下属的社会交换过程。具体为探讨下属的反馈寻求行为在上司不当督导与下属绩效之间的中介作用，下属的学习目标定向对上述过程中的调节作用。通过问卷法获得 306 名下属与上司的对偶数据，基于层级回归和 Bootstrap 分析的结果表明：上司不当督导不仅直接影响下属的绩效，还能通过抑制下属的反馈寻求行为间接地影响员工的绩效；下属的学习目标定向调节着上司不当督导与下属的反馈寻求行为的关系，当下属的学习目标定向越低，上司不当督导对反馈寻求行为的抑制作用更加明显。

关键词 上司不当督导；员工绩效；反馈寻求行为；学习目标定向

分类号 BB49：C93

1 问题提出

领导力一直都是组织行为学研究的热点，以往关于领导力的研究多是从积极的视角来研究有效的领导行为。而领导在与下属互动的过程中，为了影响下属的行为，除了通过建立愿景、个别关怀等积极领导行为以外，还可能会经常做出一些非常细微、很难捕捉并且容易被社会所接受的负面领导行为，如奚落挖苦、控制资源等。相对于前面的积极领导行为，

后面被称为不当督导 (abusive supervision) 的消极领导行为尚未得到深入的研究。

Tepper（2000）针对组织中上述的奚落挖苦、控制资源等负性领导行为，首次提出了不当督导的概念：下属感知到的被上司反复表现出来的具有敌意性的言语和非言语行为，但不包括身体上的接触。此概念提出之后，相关学者将此概念与职场欺负 (workplace bullying) 等其他一些类似的职场负性行为进行了相应的区分（Tepper, 2007）。职场欺负指一

个个体长期遭受难以反抗的负性行为，这些行为来源于一人或多人，如侮辱，贬低、戏弄、孤立某人（Einarsen, 1999）。从概念和测量内容上看，它们之间存在着一定的重叠，均指个体反复地暴露于职场中的敌对行为情境中，均包含着被戏弄、隐私被侵犯、受侮辱、承担费力不讨好的工作、工作成绩得不到肯定等方面(吴隆增，刘军，刘刚，2009; 李永鑫，聂光辉，李艺敏 等，2011)。但它们之间也存在着一些差异。首先，不当督导行为发起者是上司，是一对一的行为，而欺负行为的发起者，可以是上司、同事或下属，可以是一对一的行为，也可以是多对一的行为；其次，欺负行为包含着行为的消极结果，但是不当督导并未预测行为的结果；再次，上司不当督导不包含对个体身体接触的行为，如打耳光，推拉等，而职场欺负行为包含着身体伤害的行为。

作为负性领导的典型代表，自不当督导的概念提出以后，得到了国内外学者的积极关注。一系列研究结果表明，上司不当督导会影响众多工作结果变量，如工作满意度、情绪耗竭、心理紧张、自尊、建言行为、抵抗、攻击、偷窃、酗酒、组织公民行为和绩效等（Tepper, 2000; Tepper, Duffy & Shaw, 2001; Duffy, 2002; Inness, Barling & Turner, 2005; Burton & Hoolber, 2006; Bamberger & Bacharach, 2006; Harvey et al, 2007; Mitchell & Ambrose, 2007; Harris，Kacmar & Zivnuska, 2007; Tepper, et al. 2006, 2008; Tsung-Yu Wu & Changya Hu, 2009; 吴宗祐，2008; 李锐，凌文轻，柳士顺，2009; 刘军，吴隆增，林雨，2009; 吴隆增等 2009）。其中，有关上司不当督导与绩效之间的关系得到了相关学者的重视。上司不当督导对下属绩效的负向影响已经在中西方文化下都得到了验证（Harris，et al. 2007; 吴隆增等，2009）。而有关不当督导对下属绩效的影响是如何产生以及如何有效应对上司不当督导提升下属绩效的实证研究还相对较少。所以，探讨上司不当督导影响下属绩效的作用机制，更全面地揭示有效应对上司不当督导、改善个人绩效的方法，是本研究的出发点。

社会交换理论和互惠准则（Blau, 1964）是用来解释上司和下属互动的重要理论之一。对员工而言，上司是其重要的交换对象之一。在员工与上司的交换过程中，主管提供下属完成任务的资源，下属以好的绩效回报主管。下属可能为了个人职业的发展，期望能够与主管形成良好的互惠关系，以及能够从主管那里得到更多的支持和资源，他们可能会在完成任务的过程中，可能会寻求如何满足他们主管要求、期望的反馈。而主管的这些反馈信息主要是关于下属绩效或者表现，下属根据此反馈来判断其行为是否正确与核实，减少关于工作和自我的不确定，这种行为即反馈寻求行为（feedback seeking behavior）（Ashford & Cummings,1983）。从工具性的视角来看，一些研究表明反馈对员工增强他们的绩效来说确实是一个相当有价值的资源（Asfhord & Cumming, 1983; VandeWalle; VandeWalle, Ganesan, Challagalla & Brown, 2000,）。已有研究表明，与那些不经常做出反馈寻求行为的员工相比，经常做出反馈寻求行为的员工的目标达成情况会更好，上司对其给予的绩效评价也更高（Deshon, Kozlowski, Schmidt, Milner & Wiechmann, 2004; Locke & Latham, 2004; Northcraft & Ashfrod, 1990），所以，不难推测员工主动向主管寻求反馈的行为会促进主管和员工之间的社会交换过程。而作为员工最重要、最有价值的交换对象和反馈源的上司，他们的领导方式反过来又影响着员工向其寻求反馈的行为（VandeWalle, et al. 2000; Madzar, 2001; Levy, Cober, Miller, 2002）。研究表明，当下属在与上司互动的过程中，知觉到上司对其表现出信任、尊重、关心、支持、鼓励时，他们会更频繁地向其上司询问关于工作绩效的信息（Chen, Lam & Zhong, 2007; Lam, Huang & Snape, 2007）；相反，当下属知觉寻求反馈所付出的代价大时，会抑制个体的反馈寻求频率（Ashford, 1986）。因此，不难推测，当下属在面对一个经常做出不信任、不支持、甚至辱骂、贬损、不理不睬、冷嘲热讽等不当督导行为的上司时，会减少向其寻求反馈的频率，但是减少了寻求反馈的频率，也同时让他们丧失了获得一些提升个人绩效的有价值信息的机会，不利于其绩效

的提高。基于以上的分析，我们提出以下假设：

H1：上司不当督导通过影响下属的反馈寻求行为间接影响下属的绩效

领导的有效性一直都是领导力研究的焦点。领导力通过影响下属得以体现，下属对领导的接受或拒绝，决定了领导的效能，这充分的体现了领导和下属的对立统一的关系，在研究领导对下属的过程中，要充分的考虑下属的特性（Fiedler, 1967; House & Mitchell, 1974; Jago & Vroom, 1980）。根据 Kerr 和 Jermeir（1978）的替代领导理论（substitute for leadership theory），下属的行为除了受到正式的层级领导的影响之外，还会受到个体、群体、任务和组织因素的影响，此理论强调了个体、任务和组织方面的因素会抵消（neutralize）或替代（substitute）领导行为，从而降低了领导行为对下属的影响。在本研究中，就下属的反馈寻求行为本身而言，除了受领导这一外界的因素的影响之外，员工自身的人格特质——学习目标定向——也是影响其寻求反馈信息的重要影响因素（VandeWalle & Cummings, 1997; VandeWalle, 2003）。学习目标定向的个体将绩效反馈视为激发完成当前任务和将来任务所需能力的一种手段，是提升任务掌控力的诊断性信息（Dweck & Leggett, 1988; VandeWalle, 2003）。高学习目标定向型的个体更多地关注反馈信息对个人能力提升和发展的价值，而较少地关注做出反馈寻求行为所付出的代价（VandeWalle & Cummings, 1997）。所以，根据 Kerr 和 Jermeir（1978）的替代理论的思想，我们推测，当主管展现出不当督导等负性行为时，下属的学习目标定向会起到一定的替代作用，即下属的学习目标定向这一人格特质会削弱不当督导对下属反馈寻求行为的影响。具体表现为对于高学习目标定向的个体而言，当面对一个经常做出不当督导的上司时，即使面临着受辱和丢面子的风险，他们对能力提升和任务控制的期望以及对反馈和困难所持有的信念仍会驱使他们向上司寻求绩效的反馈。相反，对于低学习目标定向的个体而言，他们本身就不会频繁地主动向上司寻求反馈，如果再面对一个经常贬损、

辱骂下属的不当督导类型的上司，这加大了寻求反馈的代价，会导致他们更少做出寻求反馈的行为。基于以上分析，提出以下假设：

H2：下属学习目标定向在上司不当督导与反馈寻求行为之间起调节作用。下属学习目标定向越低，不当督导对其反馈寻求行为的抑制作用越明显，进而对下属的绩效产生更大的负面效应。

综上所述，具体的研究框架如图 1 所示。
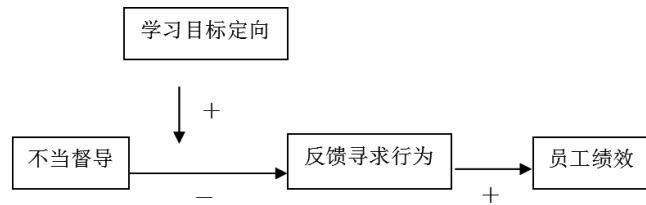
图 1　研究框架

2　研究方法

2.1　被试与调查程序

本研究采用上司 – 下属的二元对偶研究设计。调查对象来自湖北、广州、厦门、宁波的 7 家企业中的不同工作部门的上司及其直接下属。调查问卷分为员工问卷（问卷 A）和上司问卷（问卷 B），问卷 A 包括员工对上司的领导行为方式（不当督导），反馈寻求行为，和自我学习目标定向的评价，问卷 B 包括上司对下属一般任务绩效的评价。为了尽量消除被试的疑虑和保证问卷的隐匿性，员工问卷装在带有双面胶的信封内，指导语中强调该问卷调查以不记名的方式进行，并提醒被试填写完毕后将信封进行封口后，交还给调查小组成员或人力资源部门负责人。为了保证数据的配对，我们采用了相应的编码系统，以配对上司评定与下属的回答。上司调查采取方便抽样的原则确定参与调查上司 140 位，再通过随机的方式各选择 3–5 名直接下属参与调查。

此调查共发放上司问卷和员工问卷各 500 份，回收上司问卷 451 份，员工问卷 439 份，回收率分别为 90.2% 和 87.8%，剔除无效问卷后，最终得到上司 – 下属匹配数据 306 份。除去未填写项目后，调查样本中

员工的平均年龄为 33.5 岁（SD=8.53），平均工龄为 8.52 年（SD=6.75），其中男性占 35.7%，女性 65.3%。

2.2 研究工具

2.2.1 不当督导 采用 Aryee（2007）研究中所使用的 10 条目单维结构量表。此量表是 Ayree 等人根据中国文化情境，从 Tepper（2000）的原始量表中选出的 10 个条目组成的量表。该 10 条目量表的有效性在李锐，凌文辁，柳士顺（2009）的研究中得到了再次验证。该量表采用该量表采用 5 级计分，1 代表我不记得他（她）对我做出过这种行为，5 代表他（她）经常的对我做出这种行为。在本研究中，该量表的内部一致性系数为 0.91。经过验证性因素分析，所得结果为 χ^2/df=4.86<5，拟合指标 GFI=0.89，CFI=0.91，NFI=0.90，IFI=0.92，RMSEA=0.11。该结果表明不当督导的单维度结构拟合良好。

2.2.2 反馈寻求行为 关于反馈寻求行为的测量工具，主要有 Ashford 和 Cummings（1983）年编制的反馈寻求量表或其后来修订版本，Moss 等（2003）从印象管理的角度编制的反馈寻求量表，以及日本学者 Yanagizawa（2008）编制的关于目标达成情况的反馈寻求量表。Ashford 和 Cummings（1983）所编制的反馈寻求量表主要是测量员工的于反馈寻求行为的数量、努力程度等方面，Moss 等（2003）所编制的量表主要是测量个体做出反馈寻求行为的动机和意愿。因为本研究中的目的是为了探索不当督导对下属绩效的影响，为了突出反馈源以及员工寻求反馈的目的，所以我们采用日本学者 Yanagizawa（2008）所编制的反馈寻求量表，并在测量的过程中将反馈源进行了相应的确定。该量表由员工自评，共 6 个条目，例如，"我向上司寻求在处理工作问题时我的判断是否正确的信息"，"我向上司寻求关于我完成的工作有多好的信息"等。该自评问卷采用 6 点计分，1 代表从不，6 代表非常频繁。该问卷为单维结构，本研究中该量表的内部一致性系数为 0.88。

2.2.3 学习目标定向 采用 Vandewalle（1997）开发的工作领域成就目标定向量表中的学习目标定向分量表。该量表包含 5 个条目，如"我愿意选择一个能从中学习

很多东西且具有挑战性的工作任务"、"我经常寻找机会来丰富自己的知识和提高自己的技能"等。该问卷采用 7 点计分。1 代表非常不同意，7 代表非常同意。该问卷的有效性在国内也得到验证（王雁飞，凌文辁，朱瑜，2004）。本研究中该量表的内部一致性系数为 0.75。

2.2.4 员工绩效 上司评价绩效问卷采用 Tsui 等（1997）编制的绩效评估问卷。该问卷测量的是员工的一般绩效问卷，而不是具体的某个行业的绩效问卷。该问卷包含 6 个条目，单维结构，6 个条目分别从员工工作的数量、质量和效率等方面与平均水平的比较来测量员工的基本任务绩效。例如，"该员工的工作数量高于平均水平"，"该员工的工作质量远高于平均水平"。该问卷测量采用 7 点计分，1 代表非常不同意，7 代表非常同意。该问卷的单维性和有效性得到了相关学者的验证（Aryee, Budhar & Chen, 2002; 韦慧民，龙立荣，2009）。本研究中，该量表的内部一致性系数为 0.80。

2.3 统计分析

本研究采用 SPSS15.0 和 Amos7.0 进行统计分析。首先使用验证性因素分析来验证问卷的效度，然后运用逐步回归来分析不当督导对员工绩效的影响，以及反馈寻求行为的中介作用，最后运用 Edwards 和 Lambert（2007）提出的"总效应调节效应"来分析学习目标定向的调节作用。

3 数据分析和结果

3.1 验证性因素分析结果

3.3.1 员工测量量表验证性因素分析结果

运用 Amos7.0 对员工调查数据进行了验证性因素分析，比较了单因素模型（三个变量同属于一个因素），两因素模型（学习目标定性与反馈寻求行为同属于一个因素），三因素模型（不当督导，学习目标定向，反馈寻求行）从而确认研究中员工调查变量（不当督导，学习目标定性、反馈寻求行为）的区分效度。结果显示，三因素对数据的拟合程度最好，卡方与自由度之比小于 4, RMSEA 为 0.07，低

表 1　验证性因素分析结果（N=306）

模型	χ^2	df	χ^2/df	RMSEA	GFI	CFI	IFI	NFI
单因素模型	1366.91	189	7.23	0.14	0.55	0.60	0.61	0.57
两因素模型	545.55	188	2.90	0.08	0.84	0.88	0.88	0.83
三因素模型	436.59	186	2.34	0.07	0.88	0.92	0.92	0.86

表 2　各研究变量的平均数、标准差与相关矩阵（N=306）

	M	SD	1	2	3	4
1. 不当督导	2.11	0.87	1.00			
2. 反馈寻求	3.59	0.95	−0.38**	1.00		
3 学习目标定向	5.19	0.92	−0.31**	0.57**		1.00
4. 员工绩效	5.17	0.68	−0.33**	0.35**	0.36**	1.00

注：** p<0.01，* p<0.05

表 3　层级回归结果：员工反馈寻求行为的中介（N=306）

变量	Step1		Step2		Step3	
	绩效		反馈寻求		绩效	
	β	β	β	β	β	β
控制变量				0		
年龄	0.02	−0.03	0.03	−0.02	0.02	−0.03
员工性别	0.16	0.10	0.09	0.04	0.16	0.09
上司性别	−0.06	0.02	0.03	0.11	−0.06	−0.01
工龄	−0.03	−0.10	0.32**	0.25**	−0.03	−0.17
与上司工作年限	−0.10	−0.01	−0.30**	−0.21**	−0.10	0.05
自变量						
不当督导		−0.33**		−0.32**		−0.24**
中介变量						
反馈寻求行为						0.30**
R^2	0.30	0.13	0.13	0.22	0.03	0.20
F	1.66	7.09**	8.63**	14.46**	1.66	10.53**
ΔR^2		0.10		0.09		0.17

注：** p<0.01，* p<0.05

于 0.080，而 GFI，CFI，IFI，NFI 等介于 0.86 和 0.92 之间，接近于 0.90，说明三因素模型可以接受，所以，员工测量量表测量了三个不同的因素，具有较高的区分效度（见表 1）。

3.2　各研究变量的描述性统计分析

运用这些配对数据，对不当督导、反馈寻求行为、反馈逃避行为、员工绩效进行描述性统计和相关分析，其结果见表 2。从结果中可以看出，不当督导与反馈寻求、员工绩效呈显著的负相关，反馈寻求行为与员工绩效呈显著的正相关关系，这为后面的中介效应检验提供了良好的基础。

3.3　研究假设的检验

3.3.1　反馈寻求行为的中介效应检验

采用层级回归对假设 1 进行检验。检验员工反馈寻求行为的中介效应时，遵循 Baron 和 Kenny（1986）提出的中介效应检验步骤，各步骤标准化的回归系数和方程检验结果见表 3。

第一步检验了上司不当督导对员工绩效的显著影响。由表 3 结果可知，在控制了人口学变量之后，不当督导对员工绩效有显著的影响（$\beta = -0.35$, $P<0.01$）。第二步检验了上司不当督导对员工反馈寻求行为的影响。结果显示，上司不当督导对员工的反馈寻求行为达到显著影响（$\beta = -0.33$, $P<0.01$）。第三步将不当督导和反馈寻求行为同时纳入回归方程来解释员工绩效时，检验不当督导的效应消失（完全中介效应）或减弱（部分中介效应）。而结果表明，将不当督导和反馈寻求行为同时纳入方程来预测员工绩效时，反馈寻求行为的效应显著（$\beta =0.29$,

$P<0.01$），不当督导的效应有所减弱，但是依然显著（$\beta=-0.25, P<0.01$）。可以得出，反馈寻求行为在上司不当督导和员工绩效之间起部分中介作用。进一步的 sobel 检验表明反馈寻求行为在不当督导和员工绩效之间的间接效应显著（$Z=3.87, p<0.01$），假设 1 得到检验。

3.3.2 学习目标定向的调节效应检验

本研究调节效应的分析采用 Edwards 和 Lambert（2007）所提出的总效应调节模型（Total effect moderation model）的方法，将调节效应和中介效应纳入到同一个分析框架中进行分析。因为在上述的中介效应分析中，反馈寻求行为起部分中介作用，所以，在总效应调节的过程中，上司不当督导→反馈寻求行为，反馈寻求行为→员工绩效，以及上司不当督导→员工绩效三条路径都有可能受到调节变量学习目标定向的影响。因此，采用此方法将直接效应（上司不当督导→员工绩效）和间接效应（上司不当督导→反馈寻求行为→员工绩效）结合起来进行调节分析，克服了以往研究中将中介效应和调节效应进行分开的弊端，以便从更完整的角度来探讨员工学习目标定向对整个中介模型的调节效应（Edwards & Lambert, 2007; 李锐，凌文辁，柳士顺，2009）。根据 Edwards 和 Lambert（2007）所提出的分析程序，将建立以下两个回归方程：（1）回归方程一：$FSB=a_{05}+a_{X5}AS+a_{Z5}LGO+A_{xz5}(AS \times LGO)+e_{m5}$；（2）回归方程二：$Pf=b_{020}+B_{X20}AS+b_{M20}FSB+b_{Z20}LGO+B_{XZ20}(AS \times LGO)+B_{MZ}$

$_{20}(FSB \times LGO)+e_{Y20}$。其中 AS、FSB、LGO、Pf 分别代表不当督导，反馈寻求行为，学习目标定向和绩效。根据上述两个回归方程，得出或计算出下列系数或效应：（1）第一阶段：由不当督导到反馈寻求行为的回归系数（根据回归方程一得到）；（2）第二阶段，由反馈寻求行为到员工绩效的回归系数（根据回归方程二得到）；（3）直接效应：由不当督导到员工绩效的回归系数（根据回归方程二）；（4）间接效应：由第一阶段和第二阶段的回归系数相乘得到；（5）总效应：直接效应和间接效应相加得到；（6）差异：高学习目标定向个体（1 个标准差之上）情况下的系数或效应减去低学习目标定向个体（1 个标准差之下）情况下的系数或效应所得的差。

因为上面所得的系数或效应（路径系数、直接效应、间接效应、总效应等）都是从回归方程中得到的，其中单一路径（第一阶段、第二阶段以及直接效应）系数的显著性检验遵循简单斜率检验流程，高、低学习目标定向下第一阶段、第二阶段和直接效应的差异的显著性检验等同于单一系数的显著性检验，对于间接效应和总效应以及其差异的显著性检验使用拔靴法（bootstrap），以本研究中的 306 个样本为"母本"，采用有放回的抽样方式从母本中随机抽取 306 个样本，共抽 1000 组样本，根据抽取的 1000 个样本计算出单纯路径系数、间接效应、总效应的估计值。然后根据这 1000 组估算值，推导出"偏差校正置信区间"，最后根据这些置信区间来确定各路径系数、

表 4　学习目标定向调节效应分析结果

阶段与效应	低学习目标定向	高学习目标定向	差异
阶段			
第一阶段：			
不当督导→反馈寻求行为	−0.34**	−0.12	−0.22**
第二阶段：			
反馈寻求行为→员工绩效	0.20**	0.01	0.19**
效应			
直接效应	−0.16**	−0.14*	−0.02
间接效应	−0.07**	0	−0.07**
总效应	−0.23**	−0.14**	−0.09**

注：（1）表格中的数字为回归系数（第一阶段、第二阶段及直接效应），以及运用这些回归系数计算而得到的数值（间接效应、总效应及差异）。低学习目标定向是将学习目标定向的均值减一个标准差，高学习目标定向是将学习目标定向的均值加一个标准差；（2）** $P<0.01$，* $P<0.05$

间接效应、总效应及差异的显著性。分析结果见表4。

从表4中可以得出，对于低学习目标定向个体而言，反馈寻求行为对上司不当督导与员工绩效之间的关系具有中介作用，而对高学习目标定向的个体而言，反馈寻求行为对上司不当督导与员工绩效之间的关系中介作用不显著。在低学习目标定向情况下，不当督导对反馈寻求的负向效应为较强（–0.34，$p<0.01$）；在学习目标定向较高的情况下，不当督导对反馈寻求的负向效应较弱，且变得不显著（–0.12，$p>0.05$），两者的差异显著（–0.22, $P<0.01$）；相应地，通过反馈寻求行为传递的间接效应的差异也同样达到了显著水平（–0.07，$p<0.01$）；在学习目标定向高、低两种情况下，不当督导对员工绩效的负向效应都达到了显著，但是差异并不显著；在低学习目标定向情况下，模型的总效应较高，（–0.24，$p<0.01$），在高学习目标定向情况下，模型总效应较弱（–0.13，$p<0.01$），并且差异也达到了显著 (–0.11, $p<0.01$)。因此 H3 得到验证。

4 讨论

4.1 不当督导对员工绩效的影响

本研究结果表明，上司不当督导既直接地影响着员工的绩效，也通过抑制员工的反馈寻求行为间接地影响员工绩效。此结果与 Harris 等（2007）在西方情景下的研究结论一致，与吴隆增等（2009）在中国情境下所得结果也基本相同，这说明了在中西方的企业中，上司的不当督导都会对下属绩效产生消极的影响。同时，本研究结果与 Chen 等（2007）的研究结果在某种程度上具有一致性。即员工的反馈寻求行为在领导与下属社会交换的过程中扮演着重要的角色。本研究结果表明上司不当督导会通过抑制下属的反馈寻求行为进而影响下属的绩效。这一结果让我们从新的视角来理解上司与下属之间的社会交换过程。根据 Blau（1964）的社会交换原则以及 Cropanzano 和 Mitchell（2005）的消极互惠原则的观点当个体受到他人的消极对待时，也同样会以的消极的方式来对待他人，即

所谓的以眼还眼，以牙还牙。因此，当下属遭到了来自上司的奚落挖苦、控制资源等不当对待行为时，他们会以同样的方式给予上司消极的回报。但是，在我国文化背景下，上下级之间权距较大，等级制度深严，"上尊下卑"的思想严重，上级控制着下级的各种资源和命运（龙立荣，刘亚，2004），直接与上司发生冲突可能会给自己带来"厄运"，所以，个体消极回报上司时可能会有更多的顾虑，不敢轻易地以上司表现出来的不当对待的方式反过来对待上司，而会通过直接地降低个人绩效这种比较"隐蔽"的方式来报复上司，或者通过降低个人的反馈寻求行为这样一种更为"隐蔽"的方式，不主动地去寻求获得提升个人绩效的一些有价值的信息和方法，进而影响个人绩效的完成，以达到报复主管的目的。本研究结果表明下属的反馈寻求行为在上司不当督导和下属绩效之间起部分中介作用，这说明了在上司不当督导对下属的绩效影响的过程中，并非只有通过影响下属的反馈寻求行为才能进一步的影响下属的绩效，它一方面可能直接地影响了下属的绩效（Harris et al., 2007），另一方面，可能通过影响下属的其他方面进而影响下属绩效，如通过影响下属对主管的信任（吴隆增，等2009）。

4.2 学习目标定向的调节作用

本研究采用了 Edwards 和 Lambert（2007）的分析方法和程序，分析了下属的学习目标定向是否会调节不当督导对下属绩效的影响效应。研究结果表明，下属的学习目标定向可以调节不当督导和反馈寻求行为之间的关系，并能进一步的调节相应的中介链效应（"不当督导→反馈寻求行为→员工绩效"）。对低学习目标定向的下属而言，不当督导与反馈寻求行为之间的负向关系越强，进而对下属的绩效产生不利影响；对高学习目标定向的下属而言，不当督导与反馈寻求行为之间的负向关系越弱，但是并未进一步的影响其绩效。

首先，此结果证实了下属的个体特质对上司领导过程的影响。即上司不当督导对下属的影响过程受下属学习目标定向的影响。这可能是由于学习目标定向在某种程度上决定着员工的工作态度、工作

动机、努力程度以及影响着下属对工作中困难和失败的态度。学习目标定向型的个体在面对困难时会做出控制导向（mastery-oriented）的反应模式（Dweck,1986）。因此，高学习目标定向的员工可能会将上司的贬损行为理解为自己的绩效仍未"达标"，会削弱上司不当督导的消极效应。相反，对低学习目标定向的个体而言，他们可能认为完成工作仅仅是为了与领导进行交换，他们关注的焦点在于领导如何对待他们，而不是如何去提升自身的能力和技能，因此，他们不会太主动地去向上司寻求关于自己工作绩效的一些反馈，再加上要面对一个对自我形象等构成威胁的不当对待型上司的时候，他们更不会主动向上司询问关于他们绩效的相关信息。

另外，对于学习目标定向的不同水平下，反馈寻求行为的中介作用的差异。这可能是因为下属的学习目标定向水平不同，反馈寻求行为对绩效的影响不同（具体见表4第二阶段）。前人研究表明，学习目标定向与自我效能感、工作目标的设定水平呈正相关（VandeWalle, William, & John, 2001），因此，有可能低学习目标定向的下属对工作的控制感和目标设定比较低，所以，当他们从上司那里寻求反馈时，会获得一定提升绩效的有用信息，有利于其绩效的提高；而对于高学习目标定向的下属而言，他们本身对工作有着很高的控制感，通过主动寻求获得的反馈信息对其控制感的增强不是很大，对其绩效的影响也不是很大。

总的来说，本研究结果证实了 Howell 和 Shamir（2005）的观点：尽管上司和下属在权力上不平衡，但是在形成他们相互的关系的过程中均扮演着重要的角色，并且共同决定着组织的结果变量，即下属在上司的领导过程中同样起着相当关键的作用。同时，本研究证实了替代领导理在负性领导情境中的存在，为我们更为全面地理解不当督导的作用机制提供了一个新的视角。

4.3 研究意义与局限

本研究采取了上司和下属互评的方式收集数据，克服了员工单一数据所造成的共同方法偏差，使得本研究的结果可信度更高。另外，本研究结果为组织的管理实践提供了一定的指导。本研究结果显示上司的不当督导会抑制员工的反馈寻求行为，并进一步的影响员工的工作绩效，并且员工个人的学习目标定向会影响上司的领导效能。因此，为了让员工更加积极主动的寻求关于他们绩效的反馈信息，消除顾虑，进一步提高他们的工作绩效，组织和管理者应从以下两个方面来进行：一方面组织管理者应该充分认识到不当督导的消极影响，并应在实际的管理中减少和杜绝不当督导现象的发生，例如可以从制度，文化氛围等方面对上司的领导方式进行约束，还可以通过对上司进行培训，让他们知道如何人性化地去管理他们的下属以及哪些领导行为可能会导致下属产生不当对待的知觉；另一方面，应该积极地出台相应的制度和奖惩措施来支持员工主动寻求反馈的愿望和行动，构建员工帮助计划平台，帮助员工形成应对组织负性事件的技巧，以及在将来再次受到领导不当对待时，如何更好地保护自己。

本研究还存在一些局限：（1）本研究虽然采用的是上司–下属对偶设计，在一定程度上控制了共同方法偏差，但是由于测量内容本身的敏感性，社会称许性在所难免，所以，在将来的研究中，可以尝试同时从下属和第三方来收集上司的不当督导水平，然后将所搜集数据进行整合作为不当督导的水平来探讨不当督导对下属和第三方的影响及其机制，同时也可以运用深度访谈等质化研究的方式，进一步的探索不当督导对下属绩效的影响过程；（2）本研究采用的横切面研究，不能进行因果关系的推断，所以在将来的研究中，可以采用纵向研究，在收集自变量，中介变量和因变量的数据时，有一定的时间跨度或者是采取实验室模拟实验来对这些变量的因果关系进行进一步的探讨。

5 结论

本研究结果表明：（1）上司的不当督导不仅对员工的绩效有直接的消极影响，还通过抑制员工的

反馈寻求行为间接对员工的绩效产生影响；（2）员工的学习目标定向对不当督导与员工反馈寻求行为的关系，以及"不当督导→员工反馈寻求行为→员工工作绩效"中介效应链和不当督导对员工绩效的总效应均具有调节的作用，当下属的学习目标定向越低时，上述关系及效应也越强。

参考文献

Aryee, S., Budhwar, P. S., & Chen, Z. X. (2002). Trust as mediator of the relationship between organizational justice and work outcomes: Test of a social exchange model. *Journal of Origanizational Behavior, 23,* 267–285.

Aryee, S., Chen, Z. X., Sun, L., & Debrah, Y. D. (2007). Antecedents and outcomes of abusive supervision: Test of a trickle–down model. *Journal of Applied Psychology, 92,* 191–201,

Ashford, S. J. (1986). Feedback–seeking in individual adaptation: A resource perspective. *Academy of management Journal, 29,*465–487.

Ashford, S. J., & Cummings, L.L. (1983). Feedback as an individual resource: personal strategies of creating information. *Organizational Behavior and Human Performance,32,* 370–398.

Baron, R. M., & Kenny, D. A.(1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Bamberger, P. A., & Bacharach, S. B. (2006). Abusive supervision and subordinate problem drinking: Taking resistance, stress, and subordinate personality into account. *Human relations,59,* 1–30.

Blau,P. M. (1964). *Exchange and power in social life.* New York: John Wiley.

Burton.P. J. & Hoolber, M. J.(2006), Subordinate

self–esteem and Abusive supervision. *Journal of Managerial Issues,3,* 340–355.

Chen, Z. G., Lam, W. & Zhong, J, A. (2007). Leader–Member exchange and member performance: A new look at individual–level negative feedback–seeking behavior and team–level empowerment climate. *Journal of Applied Psychology. 92,* 202–212.

Cropanzano, R, & Mitchell, M. S. (2005). Social exchang theory: An interdisciplinary review. *Journal of Management, 31,* 1–27.

Deshon, R., Kozlowski, S. W. J., Schmidt, A. M., Milner, K. R., & Wiechmann, D. (2004). A multiple–goal, multilevel model of feedback effects on the regulation of individual and team performance. *Journal of Applied Psychology, 89,* 1035–1056.

Duffy, M. K., Ganster, D., & Pagon, M., C. (2002). Social undermining in the workplace. *Academy of Management Journal, 45,* 331–351.

Dweck, C. S. (1986). Motivational processes affecting learning. *American psychologist, 41,* 1040–1048.

Dweck, C. S., & Leggett, E. L. (1988). A social–cognitive approach to motivation and personality. *Psychological Review, 95,* 256–273.

Edwards, J. R. & Lambert, L. S. (2007). Methods for intergrating moderation and mediation: A general analytical framework using moderated path analysis. *Psychological Methods, 12,* 1–11.

Einarsen, S. (1999). The nature and causes of bullying at work. *International Journal of Manpower, 20,*16–27.

Fiedler, F. E. (1967). *A theory of leadership effectiveness.* New York: Academic Press.

Harris, K. J., Kacmar, K. M., & Zivnuska, S. (2007). An investigation of abusive supervision as a predictor of performance and the meaning of work as a moderator of the relationship. *Leadership quarterly, 18,* 252–263.

Harvey, P., Stoner, J. Hochwarter, W., & Kacmar,C. (2007). Coping with abusive bosses: The neutralizing effects

of ingratiation and positive affect on negative emplyee outcomes. *Leadership Quarterly, 18*, 264–280.

House, R.J., & Mitchell, T. R. (1974). Path － goal theory of leadership. *Journal of Contemporary Business, 5*, 81–97.

Howell, J. M., & Shamir, B. (2005). The role of followers in the charismatic leadership process: relationships and their consequences. *Academy of Management Review, 30*, 96–112.

Inness, M., Barling, J., & Turner, N. (2005). Understanding supervisor–targeted aggression: A within–person, between–jobs design. *Journal of Applied Psychology, 90*, 731–739.

Jago, A. G., & Vroom, V. H. (1980). An evaluation of two alternatives to the Vroom/Yetton Normative Model. *Academy of Management Journal, 23*, 347–355.

Kerr, S., & Jermeir, J. (1978). Substitutes for leadership: their meaning and measurement. *Organizational Behavior and Human Performance, 22*, 375– 403.

Lam, W., Huang, X., & Snape, E.(2007). Feedback– seeking behavior and LMX: do supervisor–attributed motives matter? *Academy of Management Journal. 50*, 348–363.

Levy, P. E., Cober, R. T. & Miller, T. (2002). The effect of transformational and Transactional leadership perception on Feedback– seeking intentions. *Journal of Applied Psychology, 32*, 1703–1720.

Li, R., Ling, W. Q. & Liu, S. S. (2009). The Mechanisms of How Abusive Supervision Impacts on Subordinates' Voice Behavior. *Acta Psychologica Sinica, 12*, 1189–1202.

[李锐 , 凌文辁 , 柳士顺 . (2009): 上司不当督导对下属建言行为的影响及其作用机制 . *心理学报 , 41*, 1189–1202.]

Li, Y. X., Nie, G. H., Li, Y. M., Wang, M. H., & Zhao, G. X. (2011). The structure and measurement of workplace bullying in china. *Journal of Psychological Science, 34*, 1201–1208

[李永鑫 , 聂光辉 , 李艺敏 , 王明辉 , 赵国祥 . (2011). 工作场所欺负的内容结构与测量 . *心理科学 , 34*, 1201–1208.]

Liu, J., Wu, L. Z., & Lin, Y. (2009).Coping with Abusive Supervision: Mechanisms of Ingratiation and Political Skill. *Nankai Business Review, 2*, 52–58.

[刘军 , 吴隆增 , 林雨 . (2009). 应对辱虐管理：下属逢迎与政治技能的作用机制研究 . *南开管理评论 , 2*, 52–58.]

Locke, E. A. & Latham, G. P. (2004). What should we do about motivation theory? Six reconmmendations for the twenty–first century. *Academy of Management Review, 29*, 388–403.

Madzar, S. (2001). Subordinate's information inquiry: Exploring the effect of perceived leadership style and individual difference. *Journal of Occupational Psychology, 74*, 221–232.

Mitchell, M. S., & Ambrose, M. L. (2007). Abusive supervision and workplace deviance and the moderating effects of negative reciprocity beliefs. *Journal of Applied Psychology. 92*, 1159–1168.

Moss, S. E., Valenzi, E. R., & Taggart, W. (2003). Are you hiding from your boss? The development of a taxonomy and instrument to assess the feedback management behaviors of good and bad. *Journal of Management, 29*,487–510.

Northcraft, G. B., & Ashford, S. J. (1990). The preservation of self in everyday life: The effects of performance expectations and feedback context on feedback inquiry. *Organizational Behavior and Human Decision Processes, 47*, 42–64.

Tepper, B. J., (2000). Consequences of abusive supervision. *Academy of Management Journal, 43*, 178–190.

Tepper, B. J., Duffy, M. K., & Shaw, J. D., (2001). Personality moderators of the relationships between

abusive supervision and subordinates' resistance. *Journal of Applied Psychology, 86*, 974–983.

Tepper, B. J., Duffy, M. K., Henle, C. A., & Lambert, L. S. (2006). Procedural injustice, victim precipitation, and abusive supervision. *Personnel psychology, 59*, 101–123.

Teeper, B. J., Henle, A. C., Lambert, S. L. Duffy, M. K. & Giacalone, A. R. (2008). Abusive supervision and subordinates' organization deviance. *Journal of Applied Psychology. 93*, 721–732.

Tsui A. S., Pearce, J. L. Porter, L. W. Tripoli, A. M. (1997). Alternative approaches to the employee– organization relationship: does investment in employees pay off? *Academy of Management Journal, 40*, 109–121.

Tsung–Yu Wu. (2008). Abusive Supervision and Emotional Exhaustion: The Mediating Effects of Subordinate Justice Perception and Emotional Labor. *Chinese Journal of Psychology, 2*, 201–221.

[吴宗祐 . (2008). 由不当督导到情绪耗竭：部属正义知觉与情绪劳动的中介效果。 *中华心理学刊, 2*, 201–221.]

Tsung–Yu Wu, Changya Hu (2009) Abusive supervision and employee emotional exhaustion: dispositional antecedents and boundaries. *Group & Organizational Management 34*, 143–169.

VandeWalle, D. (1997). Development and validation of a work domain goal orientation instrument. *Educational and Psychological Measurement,56*, 995–1015.

VandeWalle, D. (2003). A goal orientation model of feedback seeking behavior. *Human Resource Management Review,13*, 581–604.

VandeWalle, D., & Cummings, L. L..(1997). A test of the influence of goal orientation on the feedback–seeking process. *Journal of Applied Psychology, 82,* 390–400.

VandeWalle, D., Ganesan, S., Challagalla, G. N., & Brown, S. P. (2000). An integrated model of feedback–seeking behavior: Disposition,context and cognition. *Journal of Applied Psychology, 85*, 996–1003.

VandeWalle, D., & William, L. C. & John, W. S. (2001). The role of goal orientation following performance feedback. *Journal of Appilied Psychology, 86*, 629–640.

Wang, Y. F., Ling, W. Q., & Zhu, Y. (2004). The relationship among achievement goal orientation, self–efficacy and feedback–seeking behavior. *Psychological Science, 27*, 31–33.

[王雁飞 , 凌文辁 , 朱瑜 . (2004). 成就目标定向、自我效能与反馈寻求行为的关系 . *心理科学, 27*, 31–33.]

Wei, H. M., Long, L. R. (2009). Effects of cognition– and affect–base trust in supervisors on task performance and OCB. *Acta Psychologica Sinica, 41*, 86–94.

[韦慧民 , 龙立荣 . (2009). 上司认知信任和情感信任对员工行为及绩效的影响 . *心理学报, 41,* 86–94.]

Wu, L. Z., Liu, J., & Liu, G. (2009). Abusive Supervision and Employee Performance: Mechanisms of Traditionality and Trust. *Acta Psychologica Sinica, 6*, 510–518.

[吴隆增 , 刘军 , 刘刚 . (2009). 辱虐管理与员工表现：传统性与信任的作用 . *心理学报, 41*, 510–518.]

Yanagizawa, S. (2008). Effect of goal difficulty and feedback seeking on goal attainment and learning. *Japanese Psychological Research. 50*, 137–144.

Abusive supervision and employee' performance: Mechanisms of FSB and learning goral orientation

SHEN Chuangang[1,2], MA Hongyu[1,2], YANG Jing[3,4], LIU Tengfei[1,2]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan 430079, China)

([2] School of Psychology, Central China Normal University, and Hubei Human Development and Mental Health Key Laboratory,Wuhan 430079, China)

([3] Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China)

([4] Graduate School, Chinese Academy of Sciences, Beijing 100039, China)

Abstract   The literature on abusive supervision has consistently demonstrated the negative relationship between member perception of supervisor's abusive behavior and member performance. The process through which relationship supervisor's abusive behavior influences subordinates' performance, however, is still not fully understood. The present study provides a mechanism for the process. Specifically, we predict that the feedback seeking behavior(FSB) of members mediates these relationships, and learning goal orientation moderates the relationship between abusive supervision and FSB. In order to avoid the common method variance problem, two sources of survey were administrated. Data was from a total of 306 matched supervisor–subordinate dyads in 7 enterprises located in Hubei, Zhejiang, Xiamen. Two structured questionnaires were employed as the research instrument for this study. one consisted of three scales designed to measure abusive supervision, FSB and learning goal orientation, Among the major measures, the 15–items abusive supervision was adopted from Tepper(2000) study; FSB was measured via 6 items that was adopted from Saori Yanagizawa(2008) study; the five item learning goal orientation scale was adopted from   Vandewalle & Cummings(1997) study. The other questionnaire was consisted of one scale which was adopted from Tusi et al(1997) study to measure supervisor–rated subordinates' performance. Results show that the Cronbach's alpha coefficients for above measures were from 0.75 to 0.94. Hierarchical regression and the total effect moderation model were utilized to analyze the date for testing the hypotheses proposed. In line with predictions, results of hierarchical regression demonstrate that abusive supervision had negative influence on FSB, supervisor–rated performance, and FSB played partially mediated role in the relationship between abusive supervision and supervisor–rated performance. Specifically, the negative effect of abusive supervision on subordinates' performance was partially mediated by subordinates' FSB. In addition, results of total effect moderation model analysis reveal that subordinates' learning goal orientation moderated the relationship between abusive supervision and FSB. Abusive supervision was more strongly related to FSB when subordinates' learning goal orientation was low.The present study extends our understanding of social exchange between supervisor and subordinate in the link between abusive supervision and subordinate's performance. Finally, the theoretical and managerial implications of the findings, limitations and future research directions were discussed.

Keywords   abusive supervision; performance; feedback seeking behavior; learning goal orientation

企业员工的工作满意度与反生产行为：传统价值观的调节作用 *

唐汉瑛，马红宇

（青少年网络心理与行为教育部重点实验室，湖北武汉 430079；

湖北省人的发展与心理健康重点实验室 暨 华中师范大学心理学院，湖北 武汉 430079）

摘 要 反生产行为是指员工在工作场所中有意实施的所有可能给组织或者组织利益相关者造成伤害的行为，是增加组织运营成本、干扰组织内部和谐问题定的重要因素。研究实证考察中国组织情境下，各种领域性工作满意度与反生产行为的关系，并探讨传统价值观的调节作用。结果发现，相比在工作报酬、晋升、奖励和工作本身等方面的的满意度，对主管领导、同事的满意度是预测员工反生产行为发生率相对更加重要的因素；谦让守分、团结和谐两种传统价值观除了可以直接负向预测反生产行为的发生率外，还可以通过调节主管满意度与反生产行为之间关系的方式间接抑制反生产行为的发生。

关键词 反生产行为；工作满意度；传统价值观；谦让守份；团结和谐

反生产行为是指员工在工作场所中有意实施的所有可能给组织或者组织利益相关者（同事、领导、客户、股东等）造成伤害的行为。具体行为表现如，恶意取笑同事、盗窃公司财物、泄露公司机密等。有统计表明，在美国，每年仅因员工的盗窃和欺诈行为给公司造成的直接财务损失就高达 4000 亿美元之巨。除了导致组织的直接经济损失，反生产行为还会严重损害组织的声誉，并可能在更广泛的领域引发严重后果，已经成为增加组织运营成本、干扰组织内部和谐稳定的重要因素。而随着经济全球化、组织间竞争的加剧，如何最大化降低组织的生产和运营成本，同时维护组织内部的和谐稳定成为保持组织持续竞争力的关键。因此，关注工作场所中的反生产行为，并对其影响因素和发生机制进行广泛、深入的考察具有重要的现实意义。

1 文献回顾和研究假设

基于个体‑环境交互作用（individual‑context interaction）的视角，可以从因素来源（外部、内部）和作用性质（促进、抑制）两个维度将反生产行为的影响因素区分为以下四种类型，即：源自外部的诱发性因素（Triggers）和机会性因素（Opportunities），以及源自内部的倾向性因素（Propensities）和内控性因素（Internal controls）。其中，诱发性因素（如，分配不公）和倾向性因素（如，敌意性归因风格）对反生产行为的发生起到促进作用，而机会性因素（如，组织监控）和内控性因素（如，团结和谐价值观）

则对反生产行为的发生起到抑制作用。源自外部的诱发性因素常常被视作视作反生产行为发生的源头。但是，如果仅仅只有外部诱因的存在，反生产行为并不一定发生。任何反生产行为的发生都是个体与与外部诱因之间相互作用的结果。

大量研究表明，日常工作中的压力源（工作负荷、组织监控、人际关系等）是反生产行为最主要的诱发性因素。作为个体实际工作状况的直接体现，工作满意度（job satisfaction）一直被用作评价和衡量个体工作状况时最重要的指标之一。工作场所中的领导、同事关系、报酬分配和奖励/晋升机会等状况都可以在通过个体的工作满意度直接反映出来。而领导、同事关系、报酬分配和奖励/晋升等状况往往是企业员工最常见的压力来源。因此，可以将工作满意度作为评估个体日常工作中压力源状况的有效指标。研究表明，员工的整体工作满意度确实与其反生产行为发生频率密切相关。例如，Dalal 在一项元分析中发现，员工的整体工作满意度与其反生产行为发生率之间呈中等程度的负相关（$r=-0.37, p<0.01$）。然而，在整体工作满意度之外，研究者对各种领域性工作满意度（如，报酬满意度、主管满意度、同事满意度等）与反生产行为之间的关系却了解不多。诸如"员工工作状况的不同方面对其反生产行为表现的预测作用是否相同"，"员工工作状况的哪些方面对其反生产行为表现具有相对更重要的预测价值"等问题，仍然较少有实证研究方面的证据。各种领域性工作满意度反映的是员工对于各方面工作状况最直接的知觉和体验。对这些领域性工作满意度与反生产行为之间的关系进行全面、深入的考察无疑有助于人们更好地理解企业员工的工作状况与其反生产行为之间的关系。可见，通过考察各种领域性工作满意度与反生产行为之间的关系有助于实现对反生产行为影响因素的精准定位，进而为组织管理实践提供更具针对性的参考和指导。基于此，本研究以 Spector（1985）确定的工作满意度多维结构为依据，实证探讨报酬、晋升、主管、利益、奖励、流程、同事和工作本身等九个方面的工作满意度与反生产行为之间的关系。

这九个方面的工作满意度几乎涵盖了员工对其日常工作各个方面的评价状况。鉴于迄今尚未有足够的研究证据表明究竟哪些领域性工作满意度对反生产行为的预测作用更强或更弱，本研究不对各个领域性工作满意度与反生产行为之间的关系进行具体假设，仅提出一个体现探索性的总体假设：

H1 员工的各种领域性工作满意度对其反生产行为的预测作用具有差异性

依据 Marus 等人（2004）提出的个体 – 环境交互作用的观点，各种领域性工作满意度代表的只是反生产行为众多影响因素中的诱发性因素（triggers）。对这些因素的挖掘是反生产行为影响因素研究的基础性工作。但是，这些诱发性因素只是反生产行为发生的必要条件而非充分条件，它们对反生产行为的预测作用会受到一些内控性因素（internal control）的调节。员工内心持有的价值观念（如，为人处事要谦让）或具有的某些人格特质（如，尽责性）被视作可能影响反生产行为发生的典型内控性因素。目前，有关人格特质与反生产行为之间的关系已经得到不少实证研究的考察，但是专门考察个体的价值观与反生产行为之间关系的研究几乎没有。近年来，国内的众多研究者在分析中西方文化背景下下企业员工的反生产行为表现可能存在的差异时都曾提出类似的疑问，即：中国传统文化素来提倡的"与人谦让"、"以礼待人"、"维护人际和谐"等传统价值观念是否能够对国内企业员工的反生产行为的发生具有一定的抑制作用？

曾有不少西方学者试图用儒家思想来解释东亚经济的发展。Berger（1983）认为如家思想包括一整套引导人民努力工作的信仰、价值观和一套重视纪律、节俭的规范，并据此提出了"世俗化儒家"（vulgar confucianism）的概念。Berger 认为，世俗化儒家的思想有助于促成一种良好工作伦理（work ethic）的形成，进而对人们在工作中的行为表现（尤其是那些与儒家思想相悖的行为）发挥相应的约束作用，并最终促进经济的健康发展。与此同时，儒家思想中重视群体内部团结、和谐的规范，也会从传统的

中国家庭成功渗透到现代企业组织中来。杨国枢和郑伯壎（1987）对中国人的传统价值观与组织行为之间关系的研究发现，儒家传统价值观较强的员工更加偏好组织结构明确、工作环境良好、工作报酬丰厚、上司体恤部属、同事关系和睦及工作特性丰富的工作；并且，儒家传统价值观较强的员工在工作中的表现以及守规尽职方面都更好一些。因此，本研究选取儒家思想中颇具特色的"谦让守分"和"团结和谐"两种传统价值观，实证考察它们与反生产行为之间的关系，并进一步探讨二者在工作满意度预测反生产行为过程中发挥的调节作用。具体假设如下：

H2 谦让守分、团结和谐两种传统价值观能够负向预测员工的反生产行为发生率

H3 谦让守分、团结和谐两种传统价值观能够负向调节员工工作满意度与反生产行为发生率之间的关系，亦即，员工谦让守分、团结和谐的价值观越强，其工作满意度与反生产行为之间的关系越弱

2 研究方法

2.1 样本

本研究被试为来自全国多个省市（深圳、广州、东莞、长沙、重庆、武汉）的 377 名企业员工。其中，男性 182 人（48.3%）、女性 195 人（51.7%）；国有企业 45 人（11.9%）、民营企业 241 人（63.9%）、中外合资企业 14 人（3.7%）、外资企业 61 人（16.2%），信息缺失 16 人（4.2%）；中学学历者 52 人（13.8%）、专科学历者 142 人（37.7%）、本科学历者 100 人（26.5%）、硕士学历者 43 人（11.4%）、博士学历者 2 人（0.5%）；被试年龄介于 18 到 53 岁之间，平均年龄 26.8 岁；被试在当前所在公司的平均工作年限为 27.3 月。

2.2 测量工具

反生产行为。反生产行为的测量工具为唐汉瑛（2011）编制的"中国企业员工的反生产行为问卷"。该问卷共 28 道测题，包括人际指向的反生产行为（CWB_I）和组织指向的反生产行为（CWB_O）两

个维度，各 14 题。该问卷具有良好的信度和效度。在本研究中，人际指向和组织指向的反生产行为的一致性 α 系数分别为 0.76 和 0.89。

工作满意度。工作满意度的测量工具为 Spector（1985）编制的多维工作满意度问卷（Job Satisfaction Survey）[10]。该量表共 36 题，从九个方面（报酬、晋升、主管、利益、奖励、流程、同事、人际和工作本身）对企业员工的工作满意度进行全方位测量。考虑到研究目的的实际需要，本研究实际测量的内容只包括其中六个维度，分别为报酬、晋升、主管、奖励、同事和工作本身。问卷题目采用六级计分方式，"1"–"6"依次表示"非常不同意"–"非常同意"。该问卷的一致性 α 系数为 0.89（Blau, 1999）。在本研究中，工作满意度问卷的总体一致性 α 系数为 0.89；报酬、晋升、主管、奖励、同事和工作本身等 6 个维度的一致性 α 系数分别为 0.75、0.74、0.77、0.77、0.78 和 0.80。

谦让守分与团结和谐价值观。本研究用于测量谦让守分和团结和谐价值观的工具选自杨国枢和郑伯壎（1987）编制的"中国人传统价值观量表"。其中，谦让守分和团结和谐的测题分别为 10 道和 8 道，均采用四级评分方式，"1"–"4"依次代表"很不重要"–"非常重要"。在本研究中，谦让守分与团结和谐的一致性 α 系数分别为 0.76 和 0.89。

3 研究结果

3.1 共同方法偏差检验

由于本研究中多个变量数据均由同一个被试提供，且主要采用问卷调查法，有可能存在共同方法偏差效应。研究主要通过验证性因素分析对共同方法偏差的大小进行考察。结果表明，10 因素模型（10 个因素分别为 6 个领域性工作满意度 + 谦让守分、团结和谐 2 种传统价值观 + 人际指向和组织指向 2 类反生产行为）的各项拟合指数良好：RMSEA 低于 0.08，CFI、NFI 都高于 0.90，卡方与自由度的比值小于 3，说明上述变量具有良好的区分效度，确实是

表 1　各研究变量的均值、标准差及两两间的相关系数矩阵

	M	SD	1	2	3	4	5	6	7	8	9	10
1 报酬满意度	3.21	0.87	1									
2 晋升满意度	3.26	0.89	0.40**	1								
3 主管满意度	4.06	0.84	0.36**	0.13**	1							
4 奖励满意度	3.44	0.92	0.37**	0.16**	0.61**	1						
5 同事满意度	4.38	0.76	0.45**	0.21**	0.68**	0.58**	1					
6 工作满意度	3.89	0.94	0.39**	0.38**	0.31**	0.40**	0.39**	1				
7 谦让守分	2.81	0.53	0.22**	0.14**	0.11*	0.05	0.13**	0.09	1			
8 团结和谐	3.36	0.59	0.27**	0.25**	0.09	0.14**	0.15**	0.12*	0.47**	1		
9 组织指向反生产行为	1.73	0.64	−0.29**	−0.24**	−0.13*	−0.18**	−0.18**	−0.18**	−0.44**	−0.54**	1	
10 人际指向反生产行为	1.41	0.59	−0.31**	−0.31**	−0.14**	−0.13*	−0.15**	−0.09	−0.51**	−0.61**	0.77**	1

注：** 表示 $p<0.01$，* 表示 $p<0.05$；下同

10 个不同的构念。同时，本研究还采用了 Harman 的单因素因子分析来进一步的检验了共同方法偏差，结果表明未旋转时，第一个因子解释了 23.06% 的方差变异（所有因子解释的总体方差变异量为 65.75%），小于 Hair 等（1998）推荐的 50% 的判断标准，表明研究测量中的共同方法偏差并不严重。

3.2 描述性统计分析

表 1 呈现的是各变量的描述性统计分析结果。可知，除工作满意度与人际指向的反生产行为之间相关不显著外，其他各个领域性工作满意度与反生产行为之间均显著负相关；谦让守分和团结和谐价值观与两种不同指向的反生产行为之间也显著负相关。研究假设 1 和假设 2 得到初步验证。

3.3 反生产行为对领域性工作满意度的回归分

表 2　反生产行为对各领域性工作满意度和谦让守份、团结和谐价值观的回归分析结果

	CWB_I		CWB_O		CWB_I		CWB_O	
控制变量	M1	M2	M1	M2	M1	M2	M1	M2
性别	.044	.010	.066	.033	.048	.008	.067	.032
年龄	.051	.015	−.025	−.038	.050	.022	−.025	−.050
学历	−.120	−.068	.036	.100	−.123	−.074	.035	.078
工龄	−.008	.032	.010	.025	−.009	.014	.010	.030
公司性质 1	.264	.067	.270	.051	.264	.137	.270	.161
公司性质 2	.205	.022	.193	−.026	.206	.145	.193	.142
公司性质 3	.089	.055	.139	.082	.089	.025	.139	.084
公司性质 4	.406	.260	.397	.200	.401	.253	.400	.271
Ajusted R^2	.081		.077		.079		.078	
R^2	.081**		.077**		.079**		.078**	
F	2.93**		2.77**		2.86**		2.79**	
自变量								
报酬满意度		−.077		.047				
晋升满意度		.051		−.004				
主管满意度		−.245**		−.227**				
奖励满意度		.076		−.049				
同事满意度		−.249**		−.141*				
工作满意度		−.126		.002				
谦让守分						−.256**		−.229**
团结和谐						−.470**		−.406**
Ajusted R^2		.207		.174		.461		.369
R^2		.126**		.097**		.411**		.292**
F		4.84**		3.89**		22.59**		15.45**

析研究通过多元层级回归分析方法考察各领域性工作满意度对两种不同指向的反生产行为的预测作用，回归分析的第一层放入的被试的性别、年龄、学历、工龄及所在公司性质等控制变量，第二层放入的才是自变量领域性工作满意度。回归分析的结果如表 2 所示，可知：在控制了人口统计学变量的可能影响后，能够显著预测人际指向的反生产行为的领域性工作满意度有主管满意度（β =–0.25, $p<0.01$）和同事满意度（β =–0.30, $p<0.01$）；能够显著预测组织指向的反生产行为的领域性工作满意度（β =–0.23, $p<0.01$ ）同样只有主管满意度和同事满意度（β =–0.14, $p<0.05$）两项。

3.4 反生产行为对谦让守分、团结和谐价值观的回归分析

采用与反生产行为对各领域性工作满意度的回归分析时一样的方法，考察谦让守分、团结和谐两种传统价值观对反生产行为的预测作用，结果如表 2 所示。可知：在控制了人口统计学变量的可能影响后，谦让守分价值观能够显著负向预测人际

指向（β =–0.26, $p<0.01$）和组织指向（β =–0.23, $p<0.01$）两种反生产行为；同时，团结和谐价值观也能显著负向预测人际指向（β =–0.47, $p<0.01$）和组织指向（β =–0.41, $p<0.01$）两种反生产行为。假设 2 得到验证，员工持有的谦让守分和团结和谐两种价值观在其反生产行为的发生过程中确实扮演内控性因素的角色，发挥抑制作用。除了直接的抑制作用，这两种传统价值观是否还可以通过调节工作满意度与反生产行为之间关系的方式间接地抑制反生产行为的发生？因此，接下来研究对两种传统价值观在工作满意度与反生产行为关系间的调节作用进行检验。

3.5 传统价值观的调节作用分析

由前面的回归分析结果可知，只有主管满意度和同事满意度对反生产行为的预测作用显著，因此接下来仅考察谦让守分、团结和谐价值观在这两种领域性满意度预测反生产行为过程中的调节作用。开始进行回归分析前，对除因变量之外的所有预测变量（工作满意度和传统价值观）均进行中心化处理。

表 3 谦让守分价值观的调节效应

控制变量	人际指向反生产行为（CWB_1）			组织指向反生产行为（CWB_0）		
	M1	*M2*	*M3*	*M1*	*M2*	*M3*
性别	0.05	0.01	0.01	0.07	0.03	0.03
年龄	0.05	0.03	0.02	–0.03	–0.05	–0.05
学历	–0.12	–0.04	–0.04	0.04	0.11	0.11
工龄	–0.01	–0.01	0.02	0.01	0.01	0.03
公司性质 1	0.27	0.15	0.10	0.27	0.15	0.12
公司性质 2	0.21	0.11	0.03	0.19	0.08	0.04
公司性质 3	0.09	0.05	0.02	0.14	0.10	0.08
公司性质 4	0.40	0.27	0.20	0.40	0.26	0.23
Ajusted R^2	0.052			0.050		
R^2	0.079**			0.079**		
F	2.86**			2.86**		
自变量和调节变量						
主管满意度		–0.09	–0.11*		–0.15*	–0.16**
同事满意度		–0.19**	–0.19**		–0.13*	–0.13*
谦让守分		–0.43**	–0.44**		–0.37**	–0.38**
Ajusted R2		0.339			0.270	
R^2		0.277**			0.222**	
F		13.20**			10.20**	
调节效应						
主管满意度 × 谦让守分			0.14*			0.10
同事满意度 × 谦让守分			0.09			0.02
Ajusted R^2			0.366			0.311
R^2			0.040			0.012
F			13.12**			9.06**

表 4　团结和谐价值观的调节效应

控制变量	人际指向反生产行为（CWB_I）			组织指向反生产行为（CWB_O）		
	$M1$	$M2$	$M3$	$M1$	$M2$	$M3$
性别	0.05	−0.01	−0.01	0.07	0.02	0.02
年龄	0.05	0.02	0.02	−0.03	−0.06	−0.06
学历	−0.12	−0.06	−0.06	0.04	0.10	0.09
工龄	−0.01	0.04	0.04	0.01	0.05	0.05
公司性质 1	0.27	−0.03	0.01	0.27	−0.01	0.01
公司性质 2	0.21	−0.02	0.03	0.19	−0.03	−0.01
公司性质 3	0.09	−0.01	0.01	0.14	0.04	0.05
公司性质 4	0.40	0.12	0.13	0.40	0.13	0.14
Ajusted R^2	0.052			0.050		
R^2	0.079**			0.079**		
F	2.86**			2.79**		
自变量和调节变量						
主管满意度		−0.09	−0.14**		−0.15**	−0.18**
同事满意度		−0.13*	−0.12*		−0.08	−0.07
团结和谐		−0.55**	−0.49**		−0.47**	−0.44**
Ajusted R^2		0.420			0.336	
R^2		0.364**			.286**	
F		18.97**			13.63**	
调节效应						
主管满意度 × 团结和谐			0.15**			0.10
同事满意度 × 团结和谐			0.10			0.02
Ajusted R^2			0.456			0.342
R^2			0.038**			0.010
F			18.60**			11.94**

表 3 和表 4 分别呈现了谦让守分和团结和谐两种价值观的调节作用分析结果。如表 3 可知，谦让守分价值观仅在主管满意度预测人际指向的反生产行为时发挥调节作用（β=0.144, $p<0.05$）；由表 4 可知，团结和谐价值观也仅在主管满意度预测人际指向的反生产行为时发挥调节作用（β=0.148, $p<0.01$）。假设 H3 得到部分验证。具体来说，当主管满意度较低时，持有高水平谦让守分和团结和谐价值观的员工倾向于表现出较少的人际指向的反生产行为，如图 1 和图 2 所示。
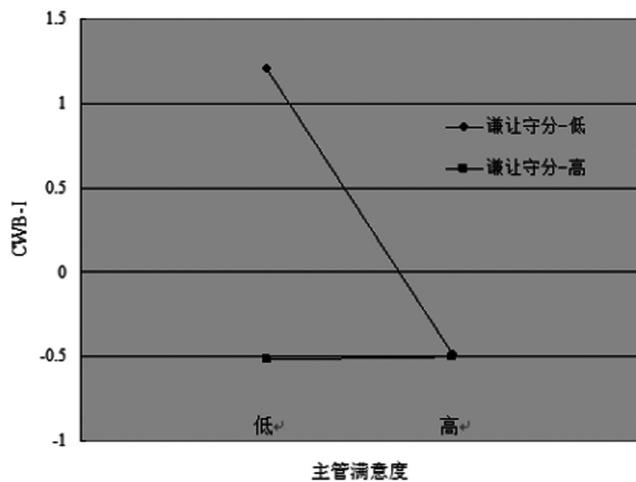

图 1　谦让守分价值观的调节效应
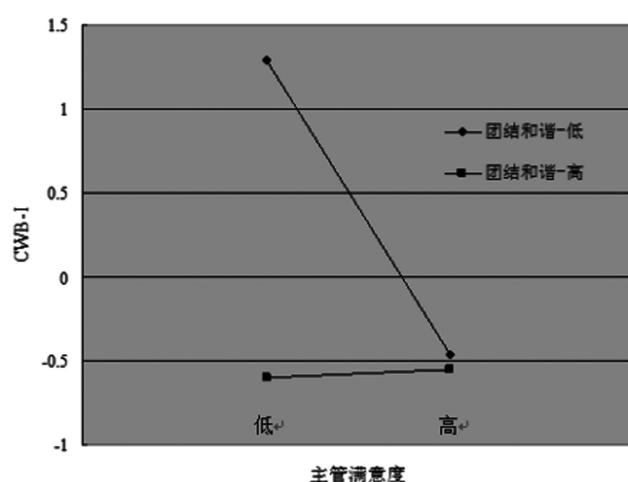
图 2　团结和谐价值观的调节效应

4 讨论

4.1 工作满意度与反生产行为

工作满意度描述的是个体从其工作或工作经历中所体会到的积极情绪状态，同时也是个体在评估工作环境是否满足其需要及满足程度基础上产生的心理体验。基于社会交换理论（social change theory）的观点，当对组织提供的工作条件、管理制度、报酬及人际关系等感到满意时，员工将会通过对组织或者组织中其他成员有利的行为表示回报；反过来，员工会通过对组织或者组织中其他成员有害的行为（反生产行为）对不满意的工作条件、报酬、及人际关系等进行消极回应。例如，退缩行为（反生产行为的一种）就被视作"一系列由对工作不满意的个体被动实施的旨在逃避工作的行为"。

本研究发现，并非所有方面的工作不满意都是员工反生产行为的有效预测因素。除了主管满意度和同事满意度，其他方面（报酬、奖励、晋升和工作本身）的工作满意度并不能直接有效地预测员工的反生产行为发生率。之所以如此，可能有以下两方面的原因。一方面，在组织中涉及员工报酬、奖励和晋升等的决策几乎都是由其主管领导做出，以至于当员工对其报酬、奖励和晋升等不满意时，首先想到的是主管领导的原因。据此可以推测，员工在报酬、奖励和晋升等方面的工作满意度之所以不能有效预测其反生产行为发生率，可能是因为它们的影响已经部分地包含在主管满意度的作用里面了。另一方面，社会取向（social orientation）是中国人适应方式的重要特点，任何时候、任何地点，人际关系对中国人心理与行为的影响都至关重要。在企业组织中，领导和同事是员工最重要的他人，相比于报酬、奖励、工作条件等物质因素，与领导、同事之间的关系是影响员工心理与行为最重要的因素。例如，在中式企业组织中，从来不乏个人因为喜欢某位领导或者喜欢工作部门的人际氛围而舍弃高薪和优越条件的例子。

4.2 传统价值观与反生产行为

与研究假设一致，谦让守分、团结和谐两种传统价值观确实能够对员工反生产行为的发生起到有效的抑制作用。一方面，员工持有的谦让守分、团结和谐两种传统价值观能够直接有效地预测其反生产行为的发生率（表2）；另一方面，这两种传统价值观还可以通过调节主管不满意与反生产行为之间关系的方式间接抑制反生产行为的发生。以及，当主管满意度较低时，持有高谦让守分和团结和谐价值观的员工出现反生产行为的可能性更低。在儒家传统价值观念中，"团结和谐"强调的是人与人之间交往要重视和谐，团队成员之间必须要有团队精神，能热爱国家，为团队绩效共同努力；"谦让守分"强调的则是个人要自守本分、与人无争、凡事谦冲为怀，容忍别人，在日常生活中则表现为坚持中庸之道、牺牲小我、成全大我、容忍别人、长幼有序、遵守规范及尊重传统等。这两种传统价值观本质上都是为了营造良好的人际关系，维持人际间的和谐相处。在工作场所中，任何人际指向的反生产行为都可能破坏组织内部的团结和谐。因此，即使员工对自己的主管领导不满意时，他们内心持有的谦让守分和团结和谐两种价值观会有效抑制其实施可能破坏人际和谐和群体团结的反生产行为。

虽然谦让守分与团结和谐价值观都能调节主管满意度与人际指向的反生产行为之间的关系，但是笔者认为这两种传统价值观念调节个体行为的途径可能存在些微的差异。关系取向是中国人社会取向的重要特征，其中的一项内容就是关系和谐性；同时，追求均衡与和谐也是中国文化的基本运作法则。杨国枢认为，为了维持关系的和谐，个人必须努力去做对方期望他/她做的事，不去做对方期望他/她不做的事；个人必须在做人方面处处小心，尽力保护他人的面子，避免可能的冲突[30]。人际间如果互动良好，将会产生投合、愉悦等积极情绪；如果互动不良，则会导致愤怒、敌意或羞耻等消极情绪，

而为了缓解或消除这些消极情绪，中国人倾向于采取自卫性投射、合理化或者直接发泄等应对方式。在这个时候，个体内心所坚持的谦让守分与团结和谐两种价值观念就可能发挥其作用。二者所不同的是，谦让守分发挥作用的途径是提醒个体"应该怎么做"，而团结和谐发挥作用的途径则是提醒个体"需要怎么做"。也就是说，当遇到不满意的主管时，谦让守分的人觉得自己"不应该"做出可能损害对方利益的行为，因为内心坚持的谦让守分价值观告诉他们在人际交往中"应该"与人无争、谦冲为怀、容忍别人；而坚持团结和谐价值观念的人觉得自己"需要"不做出可能伤害对方的行为，因为伤害对方就意味着损害团结、破坏和谐，为了不付出破坏团结、和谐的代价，他们"需要"控制自己的行为。可见，虽然发挥作用的过程可能不同，但是在抑制员工反生产行为方面，谦让守分与团结和谐两种传统价值观念具有"殊途同归"的效果。

5　结论

本研究旨在考察各种领域性工作满意度与反生产行为的关系，并在此基础上探讨谦让守分、团结和谐两种传统价值观在其间发挥的调节作用。研究得到的主要结论如下：

其一，对国内企业员工而言，相比在工作报酬、晋升、奖励和工作本身等方面的的满意度，对主管领导、同事的满意度是预测员工反生产行为发生率相对更加重要的因素。员工对主管领导或同事越不满意，其实施反生产行为的可能性越大。

其二，谦让守分、团结和谐两种传统价值观对反生产行为的发生具有抑制作用。除了可以直接负向预测反生产行为的发生率，二者还可以通过调节主管满意度与反生产行为之间关系的方式间接抑制反生产行为的发生。亦即，当主管满意度较低时，持有高水平谦让守分和团结和谐价值观的员工倾向于表现出较少的人际指向的反生产行为。

参考文献

郭晓薇, 严文华. (2008). 国外反生产行为研究述评. *心理科学, 31*, 936–939.

李敏龙, 杨国枢. (1998). 中国人的忍：概念分析与实证研究. *本土心理学研究, 10*, 3–68.

李亦园. (1993). 从民间文化看文化中国. 台北：*台大考古人类学刊, 49*, 7–17.

林玲, 唐汉瑛, 马红宇. (2010). 工作场所中的反生产行为：及其心理机制. *心理科学进展, 18*(1), 151–161.

龙立荣, 周浩. (2007). 职场攻击行为研究探讨. *外国经济与管理, 29*(9), 42–49.

唐汉瑛. (2011). *工作场所中的反生产行为及其心理机制*. 华中师范大学博士学位论文.

杨国枢. (1992). 中国人的社会取向. 社会互动的观点 //. 杨国枢, 徐安邦. 中国人的心理与行为： 理念与方法篇. 中国台北： 桂冠.

杨国枢, 郑伯壎. (1987). 传统价值观、个人现代性及组织行为： 后儒家假说的一项微观验证. "*中央研究院*"（*中国台湾*）民族学院研究所集刊*64*, 1–49.

Berger, P. L. (1983). *Cultural identity and modernization in Asian countries.* Kokugakuin University Centennial Symposium.

Bowling, N. A., & Eschleman, K. J. (2010). Employee personality as a moderator of the relationships between work stressors and counterproductive work behavior. *Journal of Occupational Health Psychology, 15*(1), 91–103.

Cropanzano, R., Howes, J. C., Grandey, A. A., & Toth, P. (1997). The relationship of organizational politics and support to work behaviors, attitudes , and stress. *Journal of Organizational Behavior, 18,* 159–180.

Dalal, R. S. (2005). A meta–analysis of the relationship between organizational citizenship behavior and counterproductive work behavior. *Journal of Applied Psychology, 90,* 1241–1255.

Dawis, R., & Lofquist, L. H. (1984). *A psychological theory of work adjustment: An individual differences*

model and its applications. Minneapolis ,MN: University of Minnesota Press.

Douglas, S. C., & Martinko, M. J. (2001). Exploring the role of individual differences in the prediction of workplace aggression. *Journal of Applied Psychology, 86,* 547–559.

Francesco, A. M., & Chen, Z. X. (2004). Collectivism in action ： its moderating effects on the relationship between organizational commitment and employee performance in China. *Group and Organizational Management, 4,* 425–441.

Gruys, M. L., & Sackett, P. R. (2003). Investigating the dimensionality of counterproductive work behavior. *International Journal of Selection and Assessment, 11,* 30–42.

Hanisch, K. A., & Hulin, C. L. (1990). Retirement as a voluntary organizational withdrawal behavior. *Journal of Vocational Behavior , 37,* 60–78.

Kaplan, S., Bradley, J. C., Luchman, J. N., & Haynes, D. (2009). On the role of positive and negative affectivity in job performance A meta–analytic investigation. *Journal of Applied Psychology, 94,* 162–176.

Locke, E. A. (1976). *The nature and causes of job satisfaction*. In M. D. Dunnette(Ed.), Handbook of industrial and organizational psychology of industrial and organizational psychology (pp.1297 – 1349). Chicago: Rand McNally.

Marcus, B. & Schuler, H. (2004). Antecedents of counterproductive behavior at work: A general perspective. *Journal of Applied Psychology, 89*(4), 647–660.

Martinko, M. J., Gundlach, M. J., & Douglas, S. C. (2002). Toward an integrative theory of counterproductive workplace behavior A causal reasoning perspective. *International Journal of Selection and Assessment, 10,* 36–50.

Miles, D. E., Borman, W. E., Spector, P. E., & Fox, S. (2002). Building an integrative model of extra role work behaviors ： A comparison of counterproductive work behavior with organizational citizenship behavior. *International Journal of Selection and Assessment, 10,* 51–57.

Penney, L. M., & Spector, P. E. (2002). Narcissism and counterproductive work behavior ： Do bigger egos mean bigger problems. *International Journal of Selection and Assessment, 10,* 126–134.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psycholog, 88*(5), 879–903.

Robinson, S. L., & Bennet, R. J. (1995). A typology of deviant workplace behaviors ： A multidimensional scaling study. *Academy of Management Journal, 38,* 555–572.

Spector, P. E., & Fox, S. (2002). An emotion–centered model of voluntary work behavior: Some parallels between counterproductive work behavior(CWB) and organizational citizenship behavior (OCB). *Human Resources Management Review, 12,* 269–292.

Spector, P. E., & Fox, S. (2005). The stressor–emotion model of counterproductive work behavior. In S. Fox & P. E.Spector(Eds.), Counterproductive work behavior: Investigations of actors and targets: pp.151 –174 ） .Washington , DC: American Psychological Association.

Spector, P. (1997). Job satisfaction. Thousand Oaks , CA ： Sage.Copyright ◎ 197 by Sage Publications , Ins. Items were taken from the appendix, 75–76.

Spector, P. (2010). The relationship of personality to counterproductive work behavior(CWB): An integration of perspectives. Human Resource Management Review, 12(1), 143–155.

Wells, J. T. (1999). A fistful of dollars. *Security Management, 43*(8), 70–75.

Enterprise employees' job satisfaction and counterproductive work behavior: The moderation of traditional values

TANG Hanying , MA Hongyu

(Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan 430079, China;

Key Laboratory of Human Development and Mental Health of Hubei Province, School of Psychology, Central China Normal University, Wuhan

430079, China)

Abstract　Counterproductive work behaviors (WBs) are defined as volitional behaviors in the workplace that harm or are intended to harm specific individuals and/or the organization. The relationships between job satisfaction and CWBs and the moderation of traditional values are investigated in this study. Results showed that, satisfaction with supervisors and satisfaction with colleagues are important predictors of employees' CWBs. And, modesty value and harmony value moderates the relationship between job satisfaction and CWBs. Specifically, relationship between job satisfaction and CWBs was weaker for employees with lower modesty or harmony value.

Keywords　counterproductive work behaviors; job satisfaction; traditional values; modesty value; harmony value

仕途"天花板"：公务员职业生涯高原结构、测量与效果 *

王忠军¹，龙立荣²，刘丽丹³，黄小华⁴，贾文文¹，李 璐¹，马红宇¹

(¹ 青少年网络心理与行为教育部重点实验室 , 华中师范大学心理学院暨湖北省人的发展与心理健康重点实验室 , 武汉 430079)

(² 华中科技大学管理学院 , 武汉 430074)

(³ 湖北中医药大学人文学院 , 武汉 430065)

(⁴ 江西省委统战部 , 南昌 330006)

摘 要 中国公共组织中基层公务员的职业生涯高原是值得关注的重要问题。首先通过深度访谈和开放式问卷调查 , 归纳出公务员职业生涯高原的典型特征与表现 , 然后通过问卷调查获得三批处级以下公务员的数据。基于样本一 (n=279) 的探索性因素分析发现 , 公务员职业生涯高原为两维度的结构：升迁停滞、职位边缘化。基于样本二 (n=517) 的验证性因素分析证实了公务员职业生涯高原两维度结构模型 , 公务员职业生涯高原测量工具有较理想的信度和效度。基于样本三 (n=520)、样本四 (n=230) 的研究表明 , 职业生涯高原对公务员的组织承诺、职业倦怠、工作退缩行为均有不同程度的消极的影响 , 但相对于升迁停滞 , 公务员职业边缘化对于上述效果变量的影响更为显著。

关键词 公务员 ; 职业生涯高原 ; 升迁停滞 ; 职位边缘化

分类号 B849：C93

1 问题的提出

公务员是我国社会中的一个特殊群体 , 指"我国各级党政机关、人大政协、法院检察院 , 以及承担行政职能的事业单位和群众团体的人员"（依据《公务员法》）。与其他职业群体相比 , 公务员拥有较多的社会、文化和经济资源 , 掌握较大的社会公共权力 , 承担公共事务的管理责任 (董鑫 , 2007)。职业生涯高原 (career plateau) 指的是个体在当前组织中职业生涯发展出现停滞的一种现象 (Smith-Ruig, 2009; Miles, Gordon, & Storlie, 2013)。即在职业生涯的某一阶段 , 个体进一步①"晋升" (Ference, Stoner, & Warren, 1977)、②"工作流动" (Veiga, 1981)、③"承担更大或更多责任" (Feldman & Weitz, 1988)、④"学习新知识与新技能" (Lee, 2003) 的可能性很小。鉴于职业生涯高原对个人和组织效能的消极影响 , 近年来 , 中国企业员工的职业生涯高原问题受到较多的关注 (谢宝国 , 龙立荣 , 2008; 余琛 , 2006; 李华 , 2006; 王竹青 ,

张慧, 2007), 也有少量针对教师职业生涯高原的研究 (寇冬泉, 2007; 高峰, 2011; 惠善康, 曹健, 2010), 但针对中国公共组织中公务员群体的职业生涯高原现象, 尚未有系统的研究。

1.1 公务员仕途的"天花板"现象

公务员群体中是否存在职业生涯高原现象？从我国公务员管理体制的历史与现状来看, 答案应是肯定的。近 10 年来, 中国公务员职业生涯发展所依托的制度基础是 2006 年颁布实施的《中华人民共和国公务员法》, 该法对公务员录用、分类管理、考核、职务任免、升降、交流与退出做出系统规定。与中国公务员法相配套的法规还有《党政领导干部选拔任用工作条例》、《公务员职务与级别管理规定》等。以上法规构成了中国公务员群体职业生涯发展的基本政治与制度背景。但没有任何制度是万能的, 或完美无缺。许多研究者认为, 我国公务员群体的职业生涯发展, 存在许多制度性的瓶颈制约, 比如：①晋升渠道单一, 发展空间狭小；②人力资源配置不够优化；③官本位现象严重, 个人目标与组织目标脱节；④内外流动不畅, 人才难尽其用；⑤培训重形式轻实效；⑥"强者流失、弱者易留", "逆淘汰"现象严重 (梁文懋, 杨龙兴, 2006; 梁丽芝, 郑凤娇, 2007)。直到今天, 这些问题都没有得到很好地解决。

公务员管理体制的瓶颈, 制度变革的滞后效应, 加上公务员队伍年轻化、知识化的发展趋势, 加剧了公务员职业发展的难题, 比如职业路径单一、晋升通道狭窄、晋升空间有限、工作流动困难 (张文勤, 2006; 张再生, 2005; 杜兴洋, 田进, 2011)。由于我国各类党政机关为高耸的"金字塔"式组织结构, 我们推测, 大量有潜质的公务员的职业生涯发展会过早地停滞不前, 他们在不同人生阶段, 都长期陷入职业生涯高原状态, 也被媒体称为官员或干部仕途的"天花板现象", 即在干部成长过程中, 大多数人达到一定级别后, 晋升空间便会越来越小, 从而在不同阶段遇到自身仕途的"天花板"的状况 (杜凤娇, 2009)。例如, 人民论坛杂志社曾联合多家网络媒体和研究机构进行了一项"干部成长天花板"的大型调查 (受调查人数

总计 8311 人), 结果显示, 64% 的被调查者认为县处级"天花板"干部最多, 尤其是年龄在 45-55 岁之间, 而全国绝大部分公务员都在处级以下, 在县乡两级公务员中, 更多的人只能在"科员"与"办事员"这"两级台阶"上走完"仕途" (杜凤娇, 2009)。但遗憾的是, 学术界尚未对这一重要的现象开展过系统的研究。

1.2 公务员职业生涯高原的结构

公务员的职业生涯高原是否有别于企业员工？这是值得研究的问题。从前文所述, 职业生涯高原概念的内涵十分丰富, 导致有关职业生涯高原概念结构的研究长期存在较大的争议。比如职业生涯高原概念先后出现：①单维观 (指晋升的可能性小) (Tremblay & Roger, 1993, 2004); ②二维观 (层级高原、工作内容高原) (Milliman, 1992); ③三维观 (结构高原、内容高原、生活高原) (Bardwick, 1986); ④四维观 (结构高原、内容高原、个人选择和工作技能) (Joseph, 1996)。其中, 结构高原和层级高原均指在组织中难以晋升；内容高原指在工作中难以获得新知识和技能；生活高原指个体所感受到在所有生活领域的低成就感或停滞感。国内研究者谢宝国、龙立荣和赵一君 (2008) 通过实证研究, 提炼出三维结构：层级高原、内容高原和中心化高原。其中, 中心化高原指很难被赋予更多的责任, 向组织中心转移的可能性很小。但以上都是针对企业员工的研究结果。

根据《公务员法》, 中国公务员分为领导职务 (包括乡科级、县处级、市厅级、省部级、国家级) 和非领导职务 (包括科员与办事员、正副主任科员、正副调研员、正副巡视员) 两大类别或系列, 但两类公务员之间存在交错流动的现象 (施康, 2006)。概括而言, 主要有两种职业生涯运动形式：第一, 跨越职务、职级向上流动, 即"晋升"。值得注意的是, 向上流动可能会有两种结果：一是"被重用", 地位上升、资源更多、职权和职责扩大、待遇更好；二是职责与权力没有明显变化, 甚至下降或缩小, 即被"边缘化"、"明升暗降"。第二, 跨越公共部门职能边界平行或横向流动, 即"交流"。《公务员法》规定, 公务员需要在不同地区、部门、职务进行交流、挂职锻炼、

调任或转任。相对于企业员工在组织内部的转岗、轮岗而言，公务员的交流范围更大、形式较多、内涵更丰富，并伴随着职权、职责的变化。并且，交流有一定"趋高性"，往往成为晋升的前提，没有交流的机会，也表明晋升希望渺茫 (施康，2006; 刘欣，2014)。同样地，交流可能会给公务员带来或"起"或"落"的感受。

鉴于公共组织的职位设计与公务员职业生涯运动轨迹的复杂性，我们推测，公务员职业生涯高原的概念结构也有别于企业员工。比如，晋升与交流的停滞与机会缺失可能是公务员职业生涯高原的核心，并伴随着权力资源的缺乏，以及被组织"冷落"或"边缘化"的感受。但以上推测还有待实证研究的检验。

1.3 公务员职业生涯高原的测量

对职业生涯高原的实证与干预研究都必须建立在对这一概念的有效测量的基础上。20 世纪 90 年代以前，研究者主要采用从他人判断的角度对职业高原进行测量，并提出了一些客观测量指标，如年龄、任职时间、两次晋升的间隔时间等 (Slocum, Jr, Cron, Hanson, & Rawlings, 1985; Evans & Gilbert, 1984)。其中任职时间是一个相当重要的指标，如 Tremblay 和 Roger (1993) 通过不同样本的测量，均发现如果员工的任职时间达到或超过 5 年，便可判定其达到职业生涯高原期。虽然使用任职时间、年龄、晋升间隔期等指标较为客观，但无法解释诸如为什么有些员工在同一职位干了很长时间，却依然能保持较高的工作热情和不俗的绩效等现象。

事实上，员工的工作态度和行为的基础，是基于他们对客观现实的主观认知，而非客观现实本身。因此上世纪 90 年代以后，学者们更多地从个人感知的角度，对职业生涯高原进行主观测量。Chao (1990) 就明确指出采用主观知觉测量比用客观指标更为恰当，因为个体对其职业生涯发展的主观评估，决定了其对当前工作的反应，Chao 因此使用了包含两个条目的量表来测量知觉的结构高原。随后，采用主观测量方法的研究逐渐成为主流。比如 Joseph (1996) 和 Milliman (1992) 各自开发了一些条目来测量结构高原和内容高原。谢宝国等 (2008) 也基于中国企业员工开发了一

套包含 16 个条目的主观测量工具。作为一项探索性的研究，我们对公务员职业生涯高原也将采用主观测量的方式，同时结合一些客观指标进行交互验证。

综上所述，探究基层公务员职业生涯高原现象，是深入了解中国公务员队伍职业生涯发展现状，揭示公务员职业心理与行为规律的一个独特视角。虽然学术界对职业生涯高原现象的研究已有了一些积累，但这些研究主要基于营利性的企业组织和员工样本，缺乏对非营利性的公共管理组织与公务员队伍的研究。且在以往文献中，有关职业生涯高原的理论与概念结构并不统一，缺乏统整的、权威的理论或结构模型作为实证研究的基础。另一方面，国内外文献中缺乏对公职人员职业生涯高原的理论与实证研究，由于文化与体制的差异，国外的研究结果也很难被直接适用于中国公务员对象。

考虑到中国文化中的高权力距离、"官本位"的传统思想，以及公共组织的"金字塔式"层级结构，"职位"与"权力"可能是中国公务员职业发展特殊"生态环境"中最为核心的两大要素，而这一点与企业员工也有较大差异 (如薪酬回报、知识经验、工作技能、可雇佣性对于企业员工职业发展尤其重要)。因此，本研究拟采用"归纳式"研究取向，即从现象、经验和数据出发，"自下而上"地建构公务员职业生涯高原的理论结构。由于处级及其以下级别的公务员是职业生涯高原多发的阶段，且在人数上也占中国公务员队伍的绝大多数 (杜凤娇，2009),本研究主要以处级以下的基层公务员为研究对象，拟初步探索公务员职业生涯高原的内涵结构与测量，进一步考察职业生涯高原与公务员职业心理与行为的关系。

2 研究 1：公务员职业生涯高原的质性研究及结构要素的探索

2.1 研究方法

2.1.1 公务员职业生涯高原的访谈、开放式问卷调查及问卷的初步编制

第一步：深度访谈。首先，研究人员对职业生涯

高原的相关文献进行了分析，制定了访谈提纲。访谈提纲分为两个部分：第一部分为受访者的基本信息，比如性别、年龄、工龄、学历、单位、职务、职责、职级等；第二部分为公务员职业生涯高原的表现，比如要求受访者回答以下问题：在党政机关里工作，陷入职业发展停滞状态的公务员是否普遍，哪些人员表现比较突出，有何征兆，请举例说明等。访谈是半结构化的深度访谈，以个别访谈的方式进行，每名访谈对象的访谈时间为 1 个小时左右。

考虑到访谈对象身份与访谈内容的特殊性，同时为了充分地收集信息，首先研究人员确定选取 16 名处级以下公务员（领导职务、非领导职务各 8 名）作为访谈对象，其中，领导职务中有 2 名正处、2 名副处、2 名正科、2 名副科；非领导职务中有 2 名调研员、2 名副调研员、2 名主任科员、1 名副主任科员和 1 名科员。接下来，在江西省南昌市和新余市的市委组织部门的支持和帮助下，研究人员在两市公务员名册上，每间隔一定数量的页码选取一页，然后基于该页名单中的个人信息与研究需要，从中选取 1 人作为访谈对象，并通过电话与其约定访谈时间，这样直到选取所需的 16 名对象为止。在访谈对象选取时，注意涵盖了不同性别、年龄、职务、级别和组织。在性别上，男性 10 名，女性 6 名；访谈对象的年龄在 35 ~ 47 岁之间；其工作单位包含林业局、农业局、审计局、经贸委、民政局、教育局、科技局、统战部、人大、司法局等多家单位。

第二步：开放式问卷调查。为了在更大范围内搜集公务员职业生涯高原的信息，在江西省南昌市某干部培训班上进行了开放式问卷调查。调查内容与访谈研究相同，比如公务员陷入职业生涯发展停滞状态的具体信号、征兆、表现等，要求调查对象对每一问题至少列出 3 项信息或事例。被试是来自江西省不同地市的公务员，总共发放 55 份问卷，回收有效问卷 43 份。其中，男性占 71%，女性占 29%；平均工龄 24 年；专科占 11%，本科占 76%，研究生以上占 13%；正科级占 23%，副处级占 68%，正处级占 9%。

第三步：编码、条目归类与汇总。首先由 3 名研究人员独立对访谈和开放式问卷调查所得资料进行编码，然后集中讨论编码结果，在达成一致的基础上将所有条目输入计算机，进行筛选工作，然后归类、汇总。筛选标准为：(1) 被试的描述必须有清楚的含义；(2) 对频次和重要性排序，选取出现频次大于 3 的描述。根据以上标准，最后得到 30 个有关公务员职业生涯高原的描述性陈述。

第四步：问卷条目的初步编制。在确定了公务员职业生涯高原所包含的特征之后，由 2 名人力资源管理专家参照国内外职业生涯高原测量问卷的描述，并根据前一研究阶段所汇总的公务员职业生涯高原现象描述，经分析讨论后确定了公务员职业生涯高原测量问卷的 20 个初始条目。为了确保条目的内容效度，另请 8 名研究人员对这 20 个初始条目进行了评定与修改。在综合考虑了问卷的内容效度，条目的单维性、重要性、文字表述清晰和简洁性，以及是否适合公务员对象这几个方面的因素后，最终缩减为 14 个条目构成公务员职业生涯高原预试问卷。采用 Likert 5 点等级量表计分，1 表示"完全不符合"，5 表示"完全符合"。

第五步：试测。在江西省南昌市某公务员学习班上对初编问卷进行了试测。试测时，共发放 50 份问卷，回收有效问卷 41 份。在样本中，男性占 68.3%，女性占 31.7%；正科级占 4.9%，副处级占 90.2%，正处级占 4.9%；大专占 4.9%，本科占 61.0%，研究生及其以上占 34.1%。在填答问卷的同时，要求调查对象对条目中描述不合适、表述不恰当或理解不清的地方进行标记。问卷填写完毕后，研究人员随机对其中 6 名公务员进行了回访，征求他们对问卷的意见。最后综合上述信息对部分条目的用词表达进行了调整，并将包含 14 个条目的职业生涯高原问卷用于下一阶段的大样本问卷调查。

2.1.2 大样本问卷调查

大样本问卷调查被试一部分是来自江西省南昌市的一个公务员培训班学员，学员是来自该省不同地市的公务员，另一部分来自江西省新余市党校培训班的学员，其学员是来自该市所辖各区、县、乡的公务员。

问卷由研究人员在培训期间集中发放，并当场回收。

本次调查共发放 320 份问卷，回收有效问卷 279 份，有效回收率为 87.19%。在有效样本中，男性占 82.4%，女性占 17.2%；已婚占 96.8%，未婚占 0.7%；中专及其以下占 0.4%，大专占 17.6%，本科占 60.9%，研究生及其以上占 20.4%；乡镇占 16.8%，县（市、区）占 20.8%，设区市占 40.5%，省直占 19.0%，其他地区占 1.4%；正乡科级占 30.1%，副县处级占 49.8%，正县处级占 18.6%，其他职级占 0.4%；领导职务占 89.9%，非领导职务占 7.6%；30 岁（含）以下占 1.9%，31～35 岁占 7.6%，36～40 岁占 30.5%，41～45 岁占 41.8%，46～50 岁占 12.4%，51～55 岁占 4.4%，55 岁以上占 1.5%；调查对象的平均工龄为 22.65 年，标准差为 6.07。

2.2 探索性因素分析结果

为了寻求公务员职业生涯高原的概念结构，采用 SPSS16.0 统计软件对 14 个问卷条目进行探索性因素分析，提取因子的方法为主成分分析法 (principal components analysis)，转轴的方法是极大方差法，选取特征根大于等于 1 作为保留因子的标准。第一次探索性因素分析后删除了因素负荷过低的条目"调任至外单位任职的希望不大"（因素负荷小于 0.35）；第二次探索性因素分析后删除了因素负荷过低的条目"在单位内部调动工作的可能性不大"（因素负荷小于 0.35），以及存在双重负荷的条目"我不想继续升职"；第三次探索性因素分析后删除了存在双重负荷的条目"上

级领导不关注我"；第四次因素分析的结果比较理想，10 个测量条目呈现出比较清晰的两因素结构，各测量条目的因素负荷在 0.62~0.81 之间，解释的方差总量为 60.61%。具体见表 1。

因素一包括 6 个条目，反映公务员知觉到在现任职位上得不到上级和组织的重视与关怀，缺乏权力和影响力，很难获得更大的职权，缺乏成就感，并倍感失落的一种状态，体现的是在目前职位与权力体系中被"边缘化"后的感受。因此，我们将该因素命名为"职位边缘化"。因素二包括 4 个条目，反映公务员知觉到的其职务、级别、工作岗位在公职系统中难以得到进一步晋升、交流的状况。因此，我们将该因素命名为"升迁停滞"。

经探索性因素分析所得两因素间具有中等程度的相关 ($r = 0.51$, $p < 0.01$)，职位边缘化、升迁停滞分别与总分的相关为 0.86、0.88 ($p < 0.01$)，说明两因素既共同反映了相同构念的内容，又具有一定程度的独立性。职位边缘化、升迁停滞的内部一致性信度系数 (Cronbach α) 分别为 0.85 和 0.81，总问卷的内部一致性系数为 0.87，表明职业生涯高原问卷的信度较好。

3 研究 2：公务员职业生涯高原结构的验证及信效度检验

3.1 研究方法

选取江西省南昌、九江、新余、景德镇、萍乡、宜春、赣州共 7 个地市的研究样本，采用本研究所开发的公务员职业生涯高原问卷，将 10 个测量条目在不同维度上进行交叉排列，以 Likert 5 点等级量表来计分。通过各地公务员培训与学习班、党校、机关单位来搜集数据，样本覆盖面宽、来源广泛，具有代表性，基本涵盖了该省各地区、层级、单位的公务员。采用集中发放问卷，当场回收，研究人员进行现场指导，并强调本调查的学术目的和匿名性。总共发放 1250 份问卷，回收有效问卷 1037 份，有效回收率为 82.96%。由于本次问卷调查所搜集的样本量较大，我们通过计算机将其随机分为两半，一半样本 ($n=517$)

表 1 公务员职业生涯高原结构的探索性因素分析

测量条目	因素 1：职位边缘化	因素 2：升迁停滞
Q 1：不被上级领导关怀和重视	0.81	0.22
Q 2：已很少得到组织的关怀和重视	0.80	0.15
Q 3：在工作中缺乏成就感	0.77	0.23
Q 4：在单位里经常感到很失落	0.74	0.14
Q 5：很难获得更大的职权	0.66	0.19
Q 6：所在的职位缺乏权力和资源	0.62	0.20
Q 7：继续升迁调动的可能性不大	0.13	0.78
Q 8：对得到上级的提拔不抱希望	0.37	0.78
Q 9：职务级别很难得到提升	0.33	0.77
Q 10：晋升与交流的愿望不再强烈	0.09	0.73
解释的方差（变异量）（共计60.61%）	35.07%	25.54%
内部一致性 Cronbach α 系数	0.85	0.81

注：表中数据为各测量条目的因素负荷值。

用于验证性因素分析，另一半样本 (n=520) 用于考察职业生涯高原与效果变量的关系。

在样本一 (n=517) 中，男性占 67.4%，女性占 32.6%；已婚占 88.3%，未婚占 11.7%；中专及其以下占 5.7%，大专占 32.0%，本科占 53.9%，研究生及其以上占 8.4%；乡镇占 10.5%，县区市占 17.6%，设区市占 56.7%，省直占 13.4%，其他地区占 1.8%；副乡科级占 19.4%，正乡科级占 33.3%，副县处级占 19.2%，正县处级占 11.7%，其他职级占 16.4%；领导职务占 70.3%，非领导职务占 29.7%；30 岁（含）以下占 18.1%，31 ～ 35 岁占 14.2%，36 ～ 40 岁占 21.5%，41 ～ 45 岁占 21.7%，46 ～ 50 岁占 10.4%，51 ～ 55 岁占 11.4%，55 岁以上占 2.8%；调查对象的平均工龄为 19.89 年，标准差为 10.003。采用 SPSS16.0 和 LISREL8.30 统计软件对数据进行统计分析。

3.2 研究结果

3.2.1 验证性因素分析

采用验证性因素分析 (CFA) 来验证公务员职业生涯高原两因素结构。验证性因素分析技术的关键在于通过比较多个模型之间的优劣，来确定最佳匹配模型。从前面的探索性因素分析 (EFA) 结果可知，公务员职业生涯高原是一个两因素的结构，包括升迁停滞和职位边缘化。但这两个因素之间具有显著的相关，那么，公务员职业生涯高原是否为单维度结构，即单因素模型？本研究对单因素模型和双因素模型进行比较。选择的拟合指数包括 χ^2、df、χ^2/df、RMSEA、SRMR、IFI、CFI、GFI。验证性因素分析结果见表 2、表 3 所示。

从表 2 所呈现的模型拟合指数的结果来看，单因素模型的 χ^2/df 值超过了 5，RMSEA 值超过了 0.08，SRMR 值超过了 0.06，说明单因素模型的拟合情况较差。相对而言，双因素模型的各项拟合指数明显较好，达到可接受的临界水平，并且与单因素模型比较，$\Delta\chi^2$ 为 128.45、$\Delta df=1$，达到 0.001 的显著水平，我们接受拟合效果更好的双因素模型，这一结果也说明我们编制的公务员职业生涯高原问卷具有较好的结构效度。

3.2.2 信度分析

对样本二 (n=517) 的数据进行信度分析。其中，升迁停滞、职位边缘化两个分量表的内部一致性 Cronbach α 信度系数分别为 0.82、0.85，总问卷的内部一致性系数为 0.89，所有的信度系数均高于 0.70 的推荐值。这表明，公务员职业生涯高原问卷的信度质量比较理想。同时也说明，研究 2 和研究 1 中，公务员职业生涯高原问卷信度比较稳定。

3.2.3 效标检验

根据《公务员法》和其他相关法规，如《党政领导干部选拔任用工作条例》、《公务员职务与级别管理规定》，我国选任制和委任制公务员大多具有工作任期制度。选任制公务员每一职务任期大约 5 年，在任期届满后，如果不能晋升，便可能体会到"仕途停滞"感，而委任制公务员如果在同一职位或职级上连续工作满 5 年以上，如果无法改变，也同样可能存在"仕途停滞"感。在本研究的访谈和开放式问卷调查中，接受调查的公务员也普遍认为，如果在某一职级或职务连续工作满 5 年，还得不到提拔升迁的话，晋升的希望就比较小。在以往文献中，Slocum 等 (1985)、Savery (1989)、Tremblay 和 Roger (1993) 也通过不同的企业样本进行测量均发现，如果企业员工在当前职位上的任职时间达到或超过 5 年，便可判定其达到了职业生涯高原阶段。综上考虑，我们选择了三个效标。

表 2　公务员职业生涯高原的验证性因素分析结果

模型	χ^2	df	χ^2/df	RMSEA	SRMR	IFI	CFI	GFI
虚模型	4979.02	45	110.65					
单因素模型	283.51	35	8.10	0.12	0.06	0.95	0.95	0.89
双因素模型	155.06	34	4.56	0.08	0.04	0.98	0.98	0.94

表 3　双因素结构模型中各潜变量在外源变量及误差上的负荷

升迁停滞			职位边缘化		
测量项目	负荷	误差负荷	测量项目	负荷	误差负荷
T1	0.72	0.49	Q1	0.76	0.43
T2	0.76	0.42	Q2	0.77	0.41
T3	0.70	0.52	Q3	0.67	0.55
T4	0.76	0.43	Q4	0.63	0.60
			Q5	0.73	0.46
			Q6	0.64	0.58

首先，在问卷中编制了一个题目，要求调查对象根据实际情况评估自己"在未来 5 年内晋升的可能性"。该题采用 5 级计分，1 表示"完全可能晋升"，5 表示"完全不可能晋升"。通过对样本数据进行相关分析，结果发现，公务员的"升迁停滞"与其感知的"未来 5 年内晋升的可能性"显著负相关 ($r=-0.57$, $p<0.01$)，公务员的"职位边缘化"与其感知的"未来 5 年内晋升的可能性"显著负相关 ($r=-0.28$, $p<0.01$)。

其次，考虑到中国公务员管理体制的特殊性，如公务员的职级与职务既有联系又有区别，职务指公务员所具有的岗位头衔称谓，主要体现工作能力和职责大小，如科长、处长、县长、局长、厅长等；职级指一定职务层次所对应的级别，主要体现资历，如县长所对应的职级一般是县处级正职，与正处级、调研员可能为相同级别。职务与职级不一定具有"同步性"，比如存在"高职低就"和"低职高就"的情况 (施康，2006; 刘欣，2014)，我们采用独立样本 t 检验方法进行差异分析，结果发现：(1) 任目前职级 5 年以上的公务员，其升迁停滞得分 ($M=3.18$, $SD=0.96$) 显著高于 5 年以内者 ($M=2.85$, $SD=0.99$)($t=3.79$, $p<0.01$)，其职位边缘化得分 ($M=2.85$, $SD=0.90$) 也高于 5 年以内者 ($M=2.70$, $SD=0.87$)(边缘显著, $t=1.77$, $p<0.10$)。(2) 任目前职务 5 年以上公务员，其升迁停滞得分 ($M=3.24$, $SD=0.97$) 显著高于 5 年以内者 ($M=2.85$, $SD=0.98$)($t=4.17$, $p<0.01$)，其职位边缘化得分 ($M=2.90$, $SD=0.92$) 也显著高于 5 年以内者 ($M=2.69$, $SD=0.86$)($t=2.49$, $p<0.05$)。以上结果表明，公务员职业生涯高原测量问卷，对公务员晋升和工作任期具有较好的区分度，如果 5 年没有得到升迁，公务员容易出现职业生涯高原状态。同时，上述结果表明公务员感知的升迁停滞与职位边缘化程度有差异，这也在一定程度上说明公务员职业生涯高原双因素结构的合理性。

最后，职业成功 (career success) 是个体职业生涯发展最重要的结果，指的是个体在工作经历中逐渐积累和获得的积极的心理感受 (主观职业成功)，以及与工作相关的成就 (客观职业成功) (Seibert, Grant, & Kraimer, 1999)。在当代职业生涯管理领域中，个体主观的或心理成功感日益受到学者们的重视。职业满意度 (career satisfaction) 是个体从自身的角度，采用内在的标准对其职业生涯发展状况的总体性评价与感受，是衡量个体主观职业成功的最重要的、普遍采用的指标 (Heslin, 2005)。理论上，职业生涯高原是公务员知觉到的一种严重的负性职业发展状态，与职业成功应是相反的结果。

为了检验上述假定，同时考虑到问卷的长度问题，我们在问卷中加入一道题目，请公务员评价"对目前职业发展现状的总体满意度"，采用 Likert 5 点计分，1 表示"非常不满意"，5 表示"非常满意"。该题目来源于 Greenhuas、Parasuraman 和 Wormley (1990) 编制的职业满意度问卷，该指标目前被学术界广泛用于测量主观职业成功。我们假设：公务员职业生涯高原与自评的职业满意度负相关。对公务员职业生涯高原和职业满意度进行相关分析结果表明，公务员的升迁停滞与职业满意度呈现显著的负相关 ($r=-0.41$, $p<0.01$)，公务员的职位边缘化与职业满意度同样呈现显著的负相关 ($r=-0.54$, $p<0.01$)。该结果验证了我们的假设，说明当公务员处于职业生涯高原状态时，对自身职业生涯发展满意度的评价也不高。

3.2.4 公务员职业生涯高原在人口学因素上的特征

进一步通过独立样本 t 检验、单因素方差分析和相关分析，探讨公务员职业生涯高原在人口统计学因素上的特征。结果表明：(1) 公务员的升迁停滞 在 性 别 ($M_男=2.99$, $SD_男=0.98$, $M_女=2.94$, $SD_女=0.97$, $t=0.56$, $p>0.05$) 上不存在显著差异，公务员的职位边缘化在性别 ($M_男=2.75$, $SD_男=0.86$, $M_女=2.78$, $SD_女=0.92$, $t=0.30$, $p>0.05$) 上也不存在显著差异;(2) 公务员的升迁停滞在地区因素上不存在显著差异 ($F(4,486)=0.51$, $p>0.05$)，职位边缘化在地区因素上差异显著 ($F(4,478)=3.36$, $p<0.05$)，经事后多重比较发现，县区市公务员 ($M=2.69$, $SD=0.90$) 的职位边缘化显著高于乡镇公务员 ($M=2.69$, $SD=0.90$);(3) 已婚公务员的升迁停滞 ($M=2.99$, $SD=0.96$) 与未婚公务员 ($M=2.78$, $SD=1.03$) 的差异不显著 ($t=1.52$, $p>0.05$)，已婚公务员的职位边缘化 ($M=2.72$, $SD=0.84$) 与未婚公务员

($M=2.87$, $SD=0.97$) 的差异不显著 ($t=1.07$, $p>0.05$);(4) 领导职务公务员的升迁停滞 ($M=2.97$, $SD=0.96$) 与非领导职务公务员 ($M=3.01$, $SD=1.02$) 的差异不显著 ($t=0.37$, $p>0.05$),但是,非领导职务公务员的职位边缘化 ($M=3.01$, $SD=0.88$) 显著地高于领导职务公务员 ($M=2.67$, $SD=0.87$)($t=3.92$, $p<0.01$);(5) 最后,公务员的升迁停滞与其年龄显著的正相关 (r=0.32, $p<0.01$),与其工作年限显著地正相关 (r=0.28, $p<0.01$),与其教育程度显著的负相关 (r=−0.22, $p<0.01$),公务员的职位边缘化与其年龄的相关不显著 (r=0.08, $p>0.05$),与其工作年限的相关也不显著 (r=0.01, $p>0.05$),但与其教育程度显著的负相关 (r=−0.20, $p<0.01$)。以上结果不仅能初步反映中国基层公务员的职业生涯高原现状,也为本研究开发的公务员职业生涯高原问卷的有效性提供了更多的证据。

4 研究 3:职业生涯高原与公务员职业心理与行为的关系

职业生涯发展是人们获得成就感和价值感的重要来源。而较早进入职业生涯的停滞期,往往会使他们发生机能失调,产生挫败感,进而降低工作积极性和组织的效能。正是由于职业生涯高原所引发的一系列问题,使得该现象成为人力资源管理研究的重要内容。以往研究证实,职业生涯高原会对企业员工的工作满意度、组织承诺、组织认同、工作绩效、缺勤、工作投入、工作压力、离职等一系列工作态度与行为变量产生消极影响 (Lemire, Saba, & Gagnon, 1999; McCleese, Eby, Scharlau, & Hoffman, 2007; Ji-hyun & Jinkook, 2008; 余琛, 2006; 李华, 2006; 谢宝国, 龙立荣, 2008; 白光林, 凌文辁, 李国昊, 2011)。

但值得注意的是,以上研究主要针对的是企业员工,职业生涯高原与公务员职业心理与行为关系的研究尚是空白。更重要的是,从中国公务员的历史与现状来看,与企业员工的高流动性不同,公务员队伍的离职率极低,这是不争的事实。例如,在我们的前期访谈中,管理部门和访谈对象普遍反映,即使职位多

年无法升迁,也很少有公务员会选择离职。此外,公务员的工作绩效的标准、绩效考核与评价的机制问题一直是困扰中国公共管理的一个难题,这一点也与企业员工有着明显的差异 (王骚, 2011)。因此,作为一项探索性的研究,为了揭示职业生涯高原与公务员心理与行为的关系,同时检验公务员职业生涯高原问卷的预测效度,在综合考虑以上因素后,我们拟选择组织承诺、职业倦怠与工作退缩行为作为公务员职业生涯高原的效果变量。

4.1 研究假设

组织承诺 (organizational commitment) 反映个人对所属组织的参与、忠诚和认同 (Meyer & Allen, 1984)。高组织承诺的员工会主动增加工作投入,创造更高的工作绩效,从而提升组织绩效。公务员对政府的组织承诺也是保持公务员队伍稳定、提高行政效率的重要手段,提高公务员对政府的组织承诺,意义深远 (安世民, 蔺全录, 2007)。总所周知,中国公务员的职业生涯运动轨迹与发展结果受组织与体制因素的影响较大,且公务员向组织外部流动、主动离职的机会也较少 (梁文懋, 杨龙兴, 2006; 梁丽芝, 郑凤娇, 2007)。一方面,根据归因理论 (attribution theory) 和人类成就归因中的"自我服务偏差"倾向 (self-service bias) (Weiner,1985) 来推测,公务员很可能将高原状态归因于与组织、政策、体制等相关的外因,并因此而降低对组织的承诺。另一方面,根据社会交换理论 (Blau, 1964),长期处于职业生涯高原状态的公务员,也容易产生受到组织不公正对待的感受,会降低自己回报组织的义务感,可能表现为降低对组织的承诺。例如,针对企业员工样本的研究证实,职业生涯高原与情感承诺存在负相关关系 (Ji-hyun & Jinkook, 2008; 谢宝国, 龙立荣, 2008),但对继续承诺和规范承诺的影响还缺乏研究。根据组织承诺的三维度理论,情感承诺强调对组织的情感依恋,规范承诺强调留在当前组织的义务感知,情感承诺与规范承诺间一般存在显著的正相关 (Allen & Meyer, 1990),可推测职业生涯高原与公务员对组织的情感承诺、规范承诺间存在正相关关系。而继续承诺强调对离职的成本感知,是一种

功利性的结果，由于公共组织在中国社会的先天"优越性"，职业发展越"停滞"，越可能强化公务员对于离开公共组织所带来利益损失的感知，而不得不继续留在组织中。因此，我们提出如下假设：

假设1：职业生涯高原与公务员的组织承诺相关。具体而言，升迁停滞、职位边缘化分别与其对组织的情感承诺负相关，与继续承诺正相关，与规范承诺负相关。

从职场压力的视角来看，职业生涯高原也是一种与组织和工作相关的工作压力源，职业倦怠则是目前被广泛关注的一种职业压力症状。职业倦怠 (occupational burnout)，也称工作倦怠、工作耗竭、职业枯竭等，指一种情绪耗竭、犬儒主义及个人专业效能降低的现象，是对工作中长久的情绪和压力的持续反应，表现为身体上的耗竭与长期性的疲倦，感觉无助与无望，进而对工作、生活或其他人产生负面的观念及态度 (Maslach & Jackson, 1981)。其中，情绪耗竭是指个体情绪与情感处于极度疲劳状态，乃至无法应付工作之需要；犬儒主义又称为讥讽态度、情感疏离等，是指个体以消极、否定或麻木不仁的态度对待工作，以应对耗竭，这是对工作本身的态度；专业效能包含对过去和现在的成就感的满足，可评估个体未来继续努力工作的可能 (Maslach, Jackson, & Leiter, 1996)。

对公务员而言，升迁的停滞，以及职位的边缘化，会在一定程度上导致公务员体验到不同程度的职业发展压力。一般而言，人们都带着期望到工作中，每个人都有成长的动机，希望在组织中不断成长。对于那些具有较高期待的公务员来说，如果期望没有得到满足，在工作中没有获得认可与提升，同时受体制的限制，又没有办法进行灵活的职业发展规划，人生中又缺乏其他的机会时，很容易产生压力，增加焦虑、恐惧等不良情绪。在以往文献中，"工作压力→工作倦怠"的模型已得到广泛支持 (Schaufeli & Taris, 2014)。根据资源保存理论 (the conservation of resources, COR) (Hobfoil, 1989) 来推测，公务员的升迁停滞与职位边缘化，会导致个体资源的实际的、潜在的损失，因而感知到压力和威胁，工作资源的相对缺乏，会进一步对身心反应产生负面影响，引发工作倦怠感。这一推论在其他样本的研究中也得到初步的证实，如研究发现，企业员工和教师的职业生涯高原与工作倦怠正相关 (陈子彤，金元媛，李娟，2011；赵寅汝，2012)。因此，我们提出以下假设：

假设2：公务员职业生涯高原与职业倦怠相关。具体而言，升迁停滞、职位边缘化分别与情绪耗竭正相关，与犬儒主义正相关，与专业效能负相关。

工作退缩行为 (Withdrawal behavior) 是组织成员在维系组织运作与工作角色关系的前提下，逃避工作角色和工作任务的一种行为，例如不合理的迟到、早退，找借口逃离工作、擅离岗位、缺勤、假装生病等 (Hanisch & Hulin, 1990)。由于这些行为背后的原因与机制存在某些共性，研究者统称之为退缩行为。人们的行为会受到情绪和认知的影响，当员工处于消极情绪时，更容易引发工作退缩行为 (Judge, Scott, & Ilies, 2006)。如研究发现，退缩行为可能是源于工作满意度的下降，以"闹情绪"方式表现出来，也可能是个体面临角色冲突和工作压力的一种被动性应对方式 (Griffeth, Gaertner & Sager, 1999; Taris, Schreurs, & van Iersel-van Silfhout, 2001)。

如同前面的分析，从归因的角度来看，公务员很可能将职业生涯高原归因于与公共管理组织、政策、体制等相关的外因。根据社会交换理论，处于职业生涯高原状态的公务员，会体验到组织不公正对待的感受，但在公共组织中，公开的报复或离职行为一般较为少见，公务员更容易做到的是减少对工作的投入，减少在组织中的主动参与行为，即更多地表现出退缩行为。中国文化强调以和为贵、遇事忍为先，因此在长期面对职业发展停滞时，基层公务员更可能出现消极情绪和负面认知，增加公务员身心双方面回避工作的倾向，例如消极怠工、拒绝帮助、"出工不出力"、"在其位不谋其政"等。作为国家公共管理系统中的公务员，退缩行为不仅会给个人和组织带来消极影响，还有可能给社会和民众带来长久的负面效应。因此，我们提出以下假设：

假设 3 : 公务员职业生涯高原与工作退缩行为正相关。具体而言 , 公务员升迁停滞与工作退缩行为正相关、职位边缘化与工作退缩行为正相关。

4.2 方法

4.2.1 样本

(1) 样本一来自江西省南昌、九江、新余、景德镇、萍乡、宜春、赣州共 7 个地市的共 1037 名处级 (含) 以下的基层公务员中随机分配的另一半样本 (n = 520), 问卷调查方法和数据收集过程与研究 2 相同 , 该样本量也相对较大 , 基本覆盖了各级、各类公务员。在该样本中 , 我们同时测量与收集了公务员职业生涯高原与组织承诺的数据 , 我们基于该样本来检验假设 1。其中 , 男性占 67.1%, 女性占 32.9%; 已婚占 90.2%, 未婚占 9.8%; 中专及其以下占 2.2%, 大专占 31.1%, 本科占 59.9%, 研究生及其以上占 6.8%; 乡镇占 8.6%, 县区市占 13.5%, 设区市占 61.3%, 省直占 15.3%, 其他地区占 1.2%; 副乡科级占 20.7%, 正乡科级占 32.1%, 副县处级占 18.6%, 正县处级占 11.8%, 其他职级占 16.8%; 领导职务占 69.7%, 非领导职务占 30.3%; 30 岁以下占 18.4%, 31~35 岁占 17.9%, 36~40 岁占 16.3%, 41~45 岁占 22.7%, 46~50 岁占 11.7%, 51~55 岁占 10.5%, 55 岁以上占 2.5%; 调查对象的平均工龄为 19.51 年 , 标准差为 9.64。

(2) 样本二来自湖北省某厅级单位及其下属的 2 家单位的公务员 , 研究人员对不同部门的公务员进行现场问卷调查 , 并当场收回问卷。问卷内容主要涉及公务员职业生涯高原、职业倦怠和工作退缩行为。共发放问卷 250 份 , 有效回收问卷 230 份 , 有效回收率为 92%。样本二集中于某一特定的机关系统 , 主要为处级以下公务员 , 虽然样本量小于样本一 , 但被试在性别、学历、年龄、工龄与职级上分布比较均匀 , 样本代表性较强 , 我们基于该样本的数据来检验假设 2 和假设 3。其中 , 男性占 57.8%, 女性占 42.2%; 学历为中专及以下占 2.6%, 大专占 8.3%, 本科占 48.7%, 研究生及以上占 40.4%; 在年龄上 , 25 岁以下占 4.3%, 2630 岁占 22.2%, 3135 岁占 22.2%, 3640 岁占 16.5%, 4145 岁占 12.6%, 46 岁以上占 22.2%; 工龄为 5 年以下者占 29.1%, 610 年者占 17.4%, 1115 年者占 13%,

1620 年者占 12.2%, 21 年以上占 28.3%; 办事员占 28.7%, 科员占 43.9%, 副科级占 11.7%, 正科级占 10%, 副处级占 4.8%, 正处级占 0.9%。

4.2.2 测量工具

职业生涯高原。采用本研究所编制的公务员职业生涯高原问卷 , 包含 "升迁停滞" (4 个测量项目) 和 "职位边缘化" (6 个测量项目) 两个维度 , 将 10 个测量项目在不同维度上进行交叉排列 , 以 Likert 5 点计分 , 1 表示 "完全不符合", 5 表示 "完全符合"。在样本一中 , 升迁停滞、职位边缘化两个分量表的内部一致性 α 系数分别为 0.83 和 0.85, 总问卷的内部一致性系数为 0.89; 在样本二中 , 升迁停滞、职位边缘化两个分量表的内部一致性 α 系数分别为 0.91 和 0.89, 总问卷的内部一致性系数为 0.92。这表明 , 本研究开发的公务员职业生涯高原问卷的信度较为稳定。

组织承诺。采用 Allen 和 Meyer (1990) 编制的三维度组织承诺问卷 (即情感承诺、继续承诺和规范承诺), 该问卷被国内外的众多研究普遍使用。采用 Likert 5 点等级量表计分 , 1 表示 "完全不同意"; 5 表示 "完全同意"。在本研究的样本一中 , 情感承诺、继续承诺和规范承诺的 α 系数分别为 0.76、0.79 和 0.80, 总问卷的 α 系数为 0.89。

职业倦怠。采用 Maslach 等开发的 MBI-GS (第三版) 量表 (Maslach Burnout Inventory-General Survey), 该量表是国际上应用最广的职业倦怠测量工具 , 适用更广泛职业领域的员工职业倦怠水平的测评 (Maslach et al., 1996)。该量表包括三个维度 , 分别是情绪耗竭、犬儒主义、专业效能 , 共 16 个测量项目 , 采用 Likert 5 点计分 , 1 表示 "非常不同意", 5 表示 "非常同意"。在本研究的样本二中 , 情绪耗竭、犬儒主义、专业效能的的 α 系数分别为 0.81、0.72 和 0.76, 总问卷的 α 系数为 0.87。

工作退缩行为。采用 Lehman 和 Simpson (1992) 编制的退缩行为问卷 , 共 4 个测量项目 , 如 "我不想上班"、"我对目前的工作不愿全力以赴", 采用 Linkert 4 点记分 , 1 表示 "从不如此", 4 表示 "始终如此"。在本研究的样本二中 , 该问卷的内部一致性

α 系数为 0.74。

4.2.3 变量测量的验证性因素分析

对样本一中的变量构建了一个五因素的测量模型 (升迁停滞、职位边缘化、情感承诺、继续承诺、规范承诺), 基于样本一的数据 (n=520) 进行验证性因素分析, 五因素模型的各项拟合指数如下：χ^2= 955.75、df=265、χ^2/df=3.61、RMSEA=0.07、SRMR=0.06、IFI=0.95、CFI=0.94、NFI=0.93、NNFI=0.94。模型拟合的结果比较理想, 这说明在样本一中变量测量之间具有较好的区分效度。

同样, 我们对样本二中的变量构建了一个六因素的测量模型 (升迁停滞、职位边缘化、情绪耗竭、犬儒主义、专业效能、工作退缩行为), 基于样本二的数据 (n=230) 进行验证性因素分析, 各项拟合指数如下：χ^2=869.68、df=390、χ^2/df=2.23、RMSEA=0.07、SRMR=0.07、IFI=0.96、CFI=0.96、NFI=0.93、NNFI=0.96。模型拟合的结果同样比较理想, 这说明在样本二中变量测量之间也具有较好的区分效度。

4.2.4 共同方法偏差的控制与检验

由于本研究采用问卷调查法, 所有的问卷题目均由被试本人填答, 因此测量中可能存在共同方法偏差 (common method bias) 问题。在程序控制上：(1) 本研究在数据收集过程中强调匿名性、保密性以及数据仅限于科学研究的说明；(2) 在不同变量的测量项目上采用不同的计分方式, 例如有的测量问卷采用 Likert

5 点计分, 有的则采用 Likert 4 点计分；(3) 对不同问卷采用不同的反应方式, 如有的采用同意程度, 有的采用符合程度, 有的采用行为频次。由于程序控制只能对共同方法偏差起到部分的修正作用, 我们采用统计方法进一步检验共同方法偏差效应。

首先根据 Podsakoff, MacKenzie, Lee 和 Podsakoff (2003) 的建议, 我们进行了 Harman 单因子检验, 也就是同时对所有变量的测量项目进行未旋转的主成分因素分析。如果得到了多个因子, 且第一个因子解释的变异量没有超过 40%, 则表明共同方法变异问题并不严重。本研究中, 样本一的未旋转的主成分因素分析结果表明, 有 4 个因子的特征根值大于 1, 且第一个因子解释的变异量只有 24.17%；样本二有 6 个因子的特征根值大于 1, 第一个因子解释的变异量只有 27.07%。这一结果初步说明本研究中的共同方法偏差问题并不严重。

考虑到单因子方法检验共同方法偏差的局限性, 进一步通过控制非可测潜在因子影响法检验共同方法偏差 (Anderson & Williams, 1992)。即将共同方法偏差作为一个潜变量进入结构方程模型, 允许所有的标识变量在该方法潜变量上负载, 通过比较含有共同方法偏差潜变量与不含方法潜变量的两种情况下的拟合程度, 来检验共同方法偏差效应。通过比较样本一中两模型的 χ^2 和 df 的差值可知, 当在五因素模型中加入一个共同方法变异因子后, 模型的卡方值并没有

表 4 变量的描述性统计与相关分析结果

变量 (n = 520)	M	SD	1	2	3	4	
1. 升迁停滞	3.02	0.98	(0.84)				
2. 职位边缘化	2.77	0.91	0.64**	(0.86)			
3. 情感承诺	3.14	0.82	−0.03	−0.23**	(0.76)		
4. 继续承诺	2.75	0.83	0.09*	0.16**	0.45**	(0.78)	
5. 规范承诺	2.76	0.87	0.04	−0.03	0.69**	0.64**	(0.81)

变量 (n = 230)	M	SD	1	2	3	4	5	
1. 升迁停滞	3.39	1.13	(0.91)					
2. 职位边缘化	2.83	0.94	0.67**	(0.89)				
3. 情绪耗竭	2.92	0.91	0.54**	0.60**	(0.81)			
4. 犬儒主义	2.86	0.82	0.51**	0.62**	0.67**	(0.72)		
5. 专业效能	3.78	0.66	−0.27**	−0.50**	−0.44**	−0.52**	(0.76)	
6. 工作退缩行为	1.96	0.57	0.39**	0.56**	0.60**	0.56**	−0.53**	(0.74)

注：括号中的数据为 Cronbach α 系数；**p < 0.01, *p < 0.05。

表 5　公务员职业生涯高原与各效果变量的层级回归分析

自变量	组织承诺 (n=520)			职业倦怠与工作退缩行为 (n=230)			
	情感承诺	继续承诺	规范承诺	情绪耗竭	犬儒主义	专业效能	工作退缩行为
第一步：控制变量							
性别	−0.09*	−0.07	−0.04	0.17*	0.06	−0.11	0.25**
年龄	0.15*	0.06	0.13*	0.16*	0.17*	0.10	−0.04
教育程度	−0.02	−0.10*	−0.10*	−0.11	−0.15*	0.04	−0.09
职务	−0.18**	−0.21**	−0.23**	0.05	−0.04	0.17*	0.07
第二步：预测变量							
升迁停滞	0.11	0.01	0.05	0.22*	0.19*	−0.05	0.09
职位边缘化	−0.37**	0.15*	−0.11*	0.48**	0.52**	−0.46**	0.52**
$Adj. R^2$	0.14	0.07	0.06	0.43	0.44	0.25	0.35
$\triangle R^2$	0.10	0.02	0.01	0.35	0.38	0.22	0.30

注：$^{\uparrow} p < 0.10$, *p < 0.05, **p < 0.01。

得到显著的改变 ($\Delta df=25$, $\Delta \chi^2=35.63$)。当在样本二的六因素模型中加入一个共同方法变异因子后，模型的卡方值也没有显著的改变 ($\Delta df=30$, $\Delta \chi^2=41.02$)。此外，样本一与样本二在纳入共同方法因子前后两模型的 RMSEA、CFI、IFI、NFI、NNFI 等主要拟合指标的改变在 0.01~0.02 之间，其改变并不显著。因此可判定，加入共同方法偏差因子后，模型并未得到十分明显的改善，即共同方法偏差对模型中的变量关系并未产生显著影响，基于本研究的数据得出的职业生涯高原与各效果变量的关系是可信的。

4.3　结果

表 4 呈现的是公务员职业生涯高原、组织承诺、职业倦怠、工作退缩行为的描述性统计及相关分析结果。

以职业生涯高原为自变量，分别以组织承诺的三个维度（情感承诺、继续承诺、规范承诺）、职业倦怠的三个维度（情绪耗竭、犬儒主义、专业效能）和工作退缩行为作为因变量，采用层级回归(Hierarchical Regression) 分析方法探讨职业生涯高原与各结果变量之间的关系。控制变量包含性别(1 为男，0 为女)、年龄、教育程度和职务(1 为领导职务，0 为非领导职务)。表 5 为层级回归分析的汇总结果。

层级回归分析的结果表明，控制人口学变量后，公务员的升迁停滞与组织情感承诺 (β =0.11, p>0.05)、继续承诺 (β =0.01, p>0.05)、规范承诺 (β =0.05, p>0.05) 的相关均不显著；公务员的职位边缘化与组织情感承诺显著地负相关 (β =−0.37,

p<0.01)，与继续承诺显著地正相关 (β =0.15, p<0.05)，与规范承诺的负相关边缘显著 (β =−0.11, p<0.10)，本研究的假设 1 得到部分支持。其次，公务员的升迁停滞与情绪耗竭显著地正相关 (β =0.22, p<0.01)，与犬儒主义同样显著地正相关 (β =0.19, p<0.05)，但与专业效能的相关不显著 (β =−0.05, p>0.05)；公务员的职位边缘化与情绪耗竭显著地正相关 (β =0.48, p<0.01)，与犬儒主义同样显著地正相关 (β =0.52, p<0.01)，与专业效能显著地负相关 (β =−0.46, p<0.01)，本研究的假设 2 基本得到支持。最后，公务员的升迁停滞与工作退缩行为的相关不显著 (β =0.09, p>0.05)，但职位边缘化与工作退缩行为具有显著的正相关 (β =0.52, p< 0.01)，假设 3 得到部分支持。此外，在上述结果中不难发现，公务员升迁停滞虽然能在一定程度上引发职业倦怠 (如与情绪耗竭、犬儒主义正相关)，但相对而言，职位边缘化却是一种更为严重的职业生涯高原状态，其消极影响更为严重。

5　讨论

5.1　公务员职业生涯高原现象及概念结构

某一构念的结构模型一般通过两条途径来建构：一是理论驱动型，即在文献分析与理论的基础上进行建构；二是数据驱动型，主要通过因素分析来实现。鉴于以往针对企业员工职业生涯高原的研究中，缺乏统一、权威的理论基础及结构模型，加上受文化与制度的约束，中国公务员职业生涯发展轨迹比较特殊，

本研究主要基于经验数据"自下而上"地建构公务员职业生涯高原的结构模型。本研究的"升迁停滞"维度主要是从外显的、可见的"流动"角度来刻画公务员的职业生涯高原,不仅包含针对企业员工研究所提出的层级高原 (Milliman, 1992; Bardwick, 1986; 谢宝国等, 2008)、结构高原 (Joseph, 1996) 以及针对中小学教师研究中的层级高原 (寇冬泉, 2007) 所表达的晋升或向上流动的含义,即职业生涯纵向运动的"晋升",还包含中国公务员职业生涯发展中较为特殊的横向交流、调动与转任等职务的变化,即横向运动的"交流"。无论"升"还是"迁"对公务员而言都有着同等的意义。在中国现行公务员管理体制下,也唯有"升迁"二字最能概括公务员职业生涯运动的特殊、丰富的含义。

本研究的"职位边缘化"维度主要是从内隐的、不可见的"权力"角度来刻画公务员的职业生涯高原,具有更强烈的主观心理体验色彩,体现了公务员职业生涯高原的特殊性,与谢宝国等人 (2008) 提出的"中心化高原"也有区别。在企业中,员工即便没有获得职位晋升,也可能被组织赋予更大的工作职权与责任,或者被纳入到核心的决策团队中。但在党政机关等公共组织中,组织结构为典型的"金字塔"式科层结构,职务、级别往往与特定的权力、责任和待遇挂钩,等级森严,不容错乱,故有"在其位,谋其政","不在其位,则不谋其政也"的说法,因而在公务员中难以出现类似于企业员工的中心化高原。对于公务员而言,职务级别的升迁往往伴随着权力资源与社会影响力的变化,二者如影随形。换言之,公务员在升迁停滞的同时,容易感知到所拥有权力和资源的相对静止或减少,即影响力下降。更进一步,在遭遇升迁停滞与影响下降后,公务员会感受到不被上级和组织重视,由此产生低成就感与失落感,"门前冷落车马稀"正是其后果的写照。

不过,在本研究中并没有发现企业员工中广泛存在的内容高原维度 (Milliman, 1992; Bardwick, 1986; Joseph, 1996; 谢宝国等, 2008)。究其原因,可能与组织背景的差异有关。目前,企业组织中技术更新速度快,员工的知识技能与学习能力是适应竞争、取得职业成功的关键要素。但在公共组织中,官僚行政文化仍占主导地位,公务员"官本位"思想比较严重,很多公务员在设定目标的过程中,无疑会把"晋升"、"当领导"作为实现人生价值的唯一目的或途径 (杜兴洋, 田进, 2011)。在影响公务员职业生涯发展和工作胜任的因素中,人际关系、政治背景、思想品德往往是极为重要的因素 (何胤, 2005; 候奕斌, 2007)。因此,除了少量专业技术性职务类型的公务员外,公共组织对员工知识和技能的要求总体上不如企业组织高,这在本研究前期访谈中也有相同的发现。因而,内容高原在公务员群体中并不凸显,也是容易理解的。

5.2 公务员职业生涯高原的现状特征

本研究发现,在中国基层 (处级以下) 公务员群体中,存在较为普遍的职业生涯高原现象。当陷入职业生涯高原状态时,最直接的表现就是职务、职级升迁调动的停滞 (升迁停滞),更严重的是公务员所拥有的权力、资源减少,个人影响力下降,不被领导和组织重视、发展无望所导致的心理上的困扰 (职位边缘化)。具体而言,我们通过 t 检验发现,基层公务员的职业生涯高原在性别、婚姻上不存在显著差异。相关分析表明,随着年龄的增长,公务员职业生涯高原状况越严重,这一特征在两个维度上具有一致性,其中以 45 岁以上的公务员的职业生涯高原最为明显。并且,由于工龄与年龄具有较高的相关 ($r = 0.92, p < 0.01$),所以在工龄上也基本呈现与年龄一致的情况。此外,本研究还发现,职业生涯高原与教育程度显著地负相关,其基本趋势是教育程度越高,职业生涯高原状况越好,这与当前干部选拔中对"年轻化"、"知识化"的要求是一致的。

中国是一个具有两千多年封建官僚传统的国家,自古以来就有"官者,管也"的文化传统。在职务上,本研究发现相对于领导职务,非领导职务公务员感知的职业生涯高原更为严重。由于领导职务的权力相对较大,地位较高,在某种程度上来讲,许多长期居于非领导职务系列的公务员觉得自己"有职无权",或"无职无权",处于边缘化状态,缺乏影响力,个人成

就感低便是必然的结果。我国政府公务员领导职务与非领导职务的划分虽然为公务员提供了"两个职业发展通道",但由于在非领导职务中,只有极少数的公务员能晋升到上一级领导职务,专业技术人员很难沿着非领导职务的系列向上攀升(宋斌,2010)。而高层次的非领导职务(如调研员)往往成了给担任领导职务的公务员以高一级职务待遇的"虚职",这就直接造成领导职务的"通道"吸引力依然大于非领导职务,这种"官本位"的制度设计思路使得公职人员都想方设法往领导职务的目标奋斗(施康,2006)。

5.3 职业生涯高原与公务员职业心理与行为的关系

本研究通过层级回归分析,在控制了重要的人口学变量后,考察了职业生涯高原与公务员的多种职业心理与行为效果变量(如组织承诺、职业倦怠、退缩行为)的关系。首先,众多研究证实企业员工的职业生涯高原与情感承诺的负向关系(Ji-hyun & Jinkook, 2008; Lemire et al., 1999; Chay, Argee, & Chew, 1995; Allen, Poteet, & Russell, 1998; 谢宝国,龙立荣,2008)。本研究同样证实,职位边缘化对公务员的组织情感承诺产生显著的消极影响。与我们的假设一致,本研究还发现,职位边缘化会在一定程度上削弱公务员对组织的规范承诺,同时也会增强公务员对组织的继续承诺。但本研究同时发现,升迁停滞与公务员的组织情感承诺、继续承诺和规范承诺的相关均不显著。这说明,对于组织承诺,职业生涯高原不同维度的预测作用存在差异,职位边缘化是更为显著的预测源。

与企业员工不同,公务员的收入稳定,有良好的社会保障和福利待遇,工作环境和条件较好,职业地位和职业声望较高,使得公务员成为一个热门职业。对公务员而言,放弃工作的损失较大,而重新就业的风险和难度也较大,这对年龄较大、工龄较长、学历较低的公务员更是如此。因此,即便遭遇职业生涯高原,但继续承诺却越发增强,这可能也是公务员队伍低离职率的一个原因。另一方面,在公共组织中,公务员群体大多直接受共产党与政府多年的培养和教育,文化素质、思想政治觉悟普遍较高,受制度的管

束也较多,随着年龄与工龄的增长,其忠诚于组织的规范承诺应该不低。但本研究证实,职业生涯高原不仅降低公务员对组织的情感承诺,增强继续承诺,也会在一定程度上削弱公务员的规范承诺。

本研究还表明,职业生涯高原不仅会导致公务员的职业倦怠,还会促发工作退缩行为。例如,升迁停滞会造成公务员在工作中表现出更多的情绪耗竭和犬儒主义的现象,这显然会对工作效率带来消极影响;而职位边缘化会导致公务员出现更多的情绪耗竭、犬儒主义、退缩行为,还会降低专业效能感。有意思的是,本研究再次发现,相对于升迁停滞,职位边缘化对各效果变量(组织承诺、职业倦怠、工作退缩行为)均有更加显著的预测作用,这一研究结果再次说明公务员职务升迁的复杂性。这也说明,对于公务员而言,职位边缘化是一种更应值得关注和重视的职业生涯高原状态,对公务员工作态度与行为的影响更为消极。

目前,国内缺乏公务员职业生涯高原及其后果的研究,本研究基于公务员群体的研究发现既有特殊性,也进一步丰富了职业生涯高原的研究。由于组织承诺、职业倦怠、工作退缩行为也可能与公务员的工作投入、工作绩效以及众多工作偏差行为相关联,那么,公务员职业生涯高原是否能引发"心理偏差"、"不平衡感"以及职业价值的扭曲,是否能导致当前中国社会普遍关注的官员"腐败行为"、"不作为"或者"过度行为"呢?这是未来值得研究的问题,本研究给我们带来一些重要的启示,未来可从不同理论视角(例如归因、公平、社会交换、社会比较等)进行深入探究。

5.4 研究贡献与展望

本研究采用层层递进的深度访谈、多阶段大样本问卷调查的方法,通过"自下而上"地探索并验证了公务员职业生涯高原的双因素理论结构,即体现"工作流动"的"升迁停滞",体现"权力影响"的"职位边缘化",这一理论结构提炼出中国公共组织中公务员职业生涯发展停滞现象的本质与特色,有助于我们深入理解中国公务员职业生涯发展规律,并做好公

共组织成员的职业生涯管理。本研究开发出公务员职业生涯高原问卷，具有良好的信度和效度，符合心理测量学标准，可应用于未来研究，以及公职人员职业生涯高原的咨询、诊断与干预。此外，本研究还考察了公务员职业生涯高原对一系列重要的职业心理与行为结果的影响，有助于认识职业生涯高原对公务员个人和公共组织所造成的负面影响。但作为一项探索性的研究，本研究对公务员职业生涯高原现象的研究还并不深入。未来的研究可在此基础上，进一步探讨公务员职业生涯高原对一些严重的工作偏差行为（如贪污腐败或行政不作为等）的影响机制，可从个人层面、组织层面、政策层面考察引致公务员职业生涯高原的前因以及公务员的心理反应，还可针对公务员职业生涯高原现象开展一些咨询、诊断、管理对策及干预研究。

参考文献

Allen, N. J., & Meyer, J. P. (1990). The measurement and antecedents of affective, continuance and normative commitment to the organization. *Journal of Occupational Psychology, 63*(1), 1–18.

Allen, T. D., Poteet, M. L., & Russell, J. E. A. (1998). Attitudes of managers who are more or less career plateaued. *The Career Development Quarterly, 47*(2), 159–172.

An, S. M., & Lin, Q. L. (2007). View of civil servants' organizational commitment system to the government. *Chinese Public Administration,* (10), 76–78.

[安世民 , 蔺全录 . (2007). 论公务员对政府的组织承诺制度 . *中国行政管理* , (10), 76–78.]

Anderson, S. E, & Williams, L. J. (1992). Assumptions about unmeasured variables with studies of reciprocal relationships: The case of employee attitudes. *Journal of Applied Psychology, 77*(5), 638–650.

Bai, G. L., Lin, W. Q., & Li, G. H. (2011). Research on the relationship among career plateau structure dimensions,

job satisfaction and resignation intention. *Science & Technology Progress and Policy, 28*(3), 144–148.

[白光林 , 凌文辁 , 李国昊 . (2011). 职业高原结构维度与工作满意度、离职倾向的关系研究 . *科技进步与对策* , *28*(3), 144–148.]

Bardwick, J. M. (1986). *The plateauing trap: How to avoid it in your career and your life*. New York: American Management Association.

Blau, P. M. (1964). Exchange and power in social life. New York: Wiley.

Chao, G. T. (1990). Exploration of the conceptualization and measurement of career plateau: A comparative analysis. *Journal of Management, 16*(1), 181–193.

Chay, Y. W., Aryee, S., & Chew, I. (1995). Career plateauing: Reactions and moderators among managerial and professional employees. *International Journal of Human Resource Management, 6*(1), 61–78.

Chen, Z. T., Jin, Y. Y., & Li, J. (2011). Empirical study on the relationship between career plateau and job burnout of knowledge worker. *Journal of Wuhan Textile University, 24*(2), 31–33.

[陈子彤 , 金元媛 , 李娟 . (2011). 知识型员工职业高原与工作倦怠关系的实证研究 . *武汉纺织大学学报* , *24*(2), 31–33.]

Dong, X. (2007). An analysis of the intension and extension of civil servant concept. *Qi Lu Journal,* (1), 148–150.

[董鑫 . (2007). 浅析公务员概念的内涵与外延 . *齐鲁学刊* , (1), 148–150.]

Du, F. J. (2009). The ceiling of official career: A survey of cadres' career development. *People,s Tribune,* (23), 16–19.

[杜凤娇 . (2009). 仕途 "天花板" : 突不破的 "围城" ——干部成长 "天花板" 调查 . *人民论坛* , (23), 16–19.]

Du, X. Y., & Tian, J. (2011). The study of civil servant's career development path based on competence. *Chinese Public Administration,* (11), 105–109.

[杜兴洋, 田进. (2011). 基于公务员胜任力的职业发展路径研究——以湖北省为例. *中国行政管理,* (11), 105–109.]

Evans, M. G., & Gilbert, E. (1984). Plateaued managers: their need gratifications and their effort–performance expectations. J*ournal of Management Studies, 21*(1), 99–108.

Feldman, D. C., & Weitz, B. A. (1988). Career plateaus in the sales force: Understanding and removing blockages to employee growth. *Journal of Personal Selling & Sales Management, 8*(3), 23–32.

Ference, T. P., Stoner, J. A. F., & Warren, E. K. (1977). Managing the career plateau. *Academy of Management Review, 2*(4), 602–612.

Gao, F. (2011). Research on the situation of the career plateau of the youth P. E. Teacher in college: Take the college P. E. Teachers in Beijing as example. *Journal of Beijing Sport University, 34*(7), 102–105.

[高峰. (2011). 高校青年体育教师职业生涯高原现状的研究——以北京市高校青年体育教师为例. *北京体育大学学报, 34*(7), 102–105.]

Greenhaus, J. H., Parasuraman, S. & Wormley, W. M. (1990). Effects of race on organizational experiences, job performance evaluations, and career outcomes. *Academy of Management Journal, 33*(1), 64–86.

Griffeth, R. W., Gaertner, S., & Sager, J. K. (1999). Taxonomic model of withdrawal behaviors: The adaptive response model. *Human Resource Management, 9*(4), 577–590.

Hanisch, K. A., & Hulin, C. L. (1990). Job attitudes and organizational withdrawal: An examination of retirement and other voluntary withdrawal behaviors. *Journal of Vocational Behavior, 37*(1), 60–78.

He, Y. (2005). *Research on the achievement motivation & attribution of primary level leader* (Unpublished master's thesis). Suzhou University, Suzhou.

[何胤. (2005). *基层党政干部成就动机及其归因的研究 (硕士学位论文).* 苏州大学, 苏州.]

Heslin, P. A. (2005). Conceptualizing and evaluating career success. *Journal of Organizational Behavior, 26*(2), 113–136.

Hobfoil, S. E. (1989). Conservation of resources: A new attempt at conceptualizing stress. *American Psychologist, 44*(3), 513–524.

Hou, Y. B. (2007). *The study on the civil servants of section chief´s competency and relative factors* Unpublished doctorial dissertation. Jinan University, Guangzhou.

[候奕斌. (2007). *科级公务员胜任特征及相关因素研究 (博士学位论文).* 暨南大学, 广州.]

Hui, S. K., & Cao, J. (2010). Analysis on factors affecting the shaping of teacher's career plateau in elementary and middle school. *Psychological Development and Education, 26*(5), 527–533.

[惠善康, 曹健. (2010). 中小学教师职业生涯高原现象的特征及相关因素. *心理发展与教育, 26*(5), 527–533.]

Ji–hyun, J., & Jinkook, T. (2008). The effects of perceived career plateau on employees' attitudes: Moderating effects of career motivation and perceived supervisor support with Korean employees. *Journal of Career Development, 35*(2), 187–201.

Joseph, J. (1996). An exploratory look at the plateauism construct. The Journal of Psychology: *Interdisciplinary and Applied, 130*(3), 237–244.

Judge, T. A., Scott, B. A., & Ilies, R. (2006). Hostility, job attitudes, and workplace deviance: Test of a multilevel model. *Journal of Applied Psychology, 91*(1), 126–138.

Kou, D. Q. (2007). *A study on the structure and characteristics of teacher´s career plateau and its relationship to the job effect* (Unpublished doctorial dissertation). Southwest University, Chongqing.

[寇冬泉. (2007). *教师职业生涯高原的结构、特点及*

其与工作效果的关系 (博士学位论文). 西南大学 ,
重庆 .]

Lee, B. C. P. (2003). Going beyond career plateau: Using
professional plateau to account for work outcomes.
Journal of Management Development, 22(6), 538 –
551.

Lehman, W. E., & Simpson, D. D. (1992). Employee
substance use and on–the–job behaviors. *Journal of
Applied Psychology, 77*(3), 309 – 321.

Lemire, L., Saba, T., & Gagnon, Y. C. (1999). Managing
career plateauing in the Quebec Public Sector. *Public
Personnel Management, 28*(3), 375 – 391.

Li, H. (2006). *A study of relationship between career
plateau, job satisfaction, organizational commitment
and turnover intention of business managers*
(Unpublished doctorial dissertation). Chongqing
University, Chongqing.

[李华 . (2006). *企业管理人员职业高原与工作满意度、
组织承诺及离职倾向关系研究* (博士学位论文).
重庆大学 , 重庆 .]

Liang, L. Z., & Zheng, F. J. (2007). The development of
China's professional technical civil servants career
path. *Seeker,* (2), 73 – 75, 78.

[梁丽芝 , 郑凤娇 . (2007). 中国专业技术类公务员职
业的发展路径 . *求索 ,* (2), 73 – 75, 78.]

Liang, W. M., & Yang, L. X. (2006). Opinions on the
construction of civil servants' career planning support
system in China. *Jiangxi Social Sciences,* (8), 129 –
132.

[梁文懋 , 杨龙兴 . (2006). 我国公务员职业生涯规划
支持体系建设刍议 . *江西社会科学 ,* (8), 129 – 132.]

Liu, X. (2014). *Research on the horizontal flowing of
civil servants of leading posts in China* (Unpublished
master's thesis). South China University of
Technology, Guangzhou.

[刘欣 . (2014). *我国领导职务公务员横向流动问题研
究* (硕士学位论文). 华南理工大学 , 广州 .]

Maslach, C., & Jackson, S. E. (1981). The measurement
of experienced burnout. *Journal of Occupational
Behavior, 2*(2), 99 – 113.

Maslach, C., Jackson, S. E., & Leiter, M. P. (1996).
Maslach Burnout Inventory: Manual (3rd ed.). Palo
Alto, CA: Consulting Psychologist Press.

McCleese, C. S., Eby, L. T., Scharlau, E. A., & Hoffman,
B. H. (2007). Hierarchical, job content, and double
plateaus: A mixed–method study of stress, depression
and coping responses. *Journal of Vocational Behavior,
71*(2), 282 – 299.

Meyer, J. P., & Allen, N. J. (1984). Testing the "side–
bet theory" of organizational commitment: Some
methodological considerations. *Journal of Applied
Psychology, 69*(3), 372 – 378.

Miles, S., Gordon, J., & Storlie, C. (2013). Job satisfaction,
perceived career plateau, and the perception of
promotability: A correlational study. *The Journal of
International Management Studies, 8*(1), 1 – 9.

Milliman, J. F. (1992). *Causes, consequences and
moderating factors of career plateauing* (Unpublished
doctorial dissertation). University of Southern
California.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. –Y., &
Podsakoff, N. P. (2003). Common method biases in
behavioral research: A critical review of the literature
and recommended remedies. *Journal of Applied
Psychology, 88*(5), 879 – 903.

Savery, L. K. (1989). Comparing plateaued and non–
plateaued employees. *Journal of Managerial
Psychology, 4*(1), 12 – 16.

Schaufeli, W. B., & Taris, T. W. (2014). A critical review
of the job demands–resources model: Implications for
improving work and health. In W. B. Schaufeli, & T. W.
Taris (Eds.), *Bridging occupational, organizational
and public health* (pp. 43 – 68). Netherlands: Springer.

Seibert, S. E., Grant, J. M., & Kraimer, M. L. (1999).

Proactive personality and career success. *Journal of Applied Psychology, 84*(3), 416–427.

Shi, K. (2006). *Research on relationship and integration among civil servant´s recruitment, management and exit mechanism in China* (Unpublished doctorial dissertation). Nanjing Agricultural University, Nanjing.

[施康 . (2006). *我国公务员录用、管理与退出机制的关系及整合研究* (博士学位论文). 南京农业大学 , 南京 .]

Slocum, J. W., Jr., Cron, W. L., Hanson, R. W., & Rawlings, S. (1985). Business strategy and the management of plateaued employees. *Academy of Management Journal, 28*(1), 133–154.

Smith-Ruig, T. (2009). Exploring career plateau as a multi-faceted phenomenon: Understanding the types of career plateaux experienced by accounting professionals. *British Journal of Management, 20*(4), 610–622.

Song, B. (2010). Research on career development channels of public officials and SWOT model. *Chinese Public Administration,* (12), 69–72.

[宋斌 . (2010). 公职人员职业发展通道和 SWOT 模型初探 . *中国行政管理，* (12), 69–72.]

Taris, T. W., Schreurs, P. J. G., & van Iersel-van Silfhout, I. J. (2001). Job stress, job strain, and psychological withdrawal among Dutch university staff: Towards a dual process model for the effects of occupational stress. Work & Stress: An International Journal of Work, *Health & Organisations, 15*(4), 283–296.

Tremblay, M., & Roger, A. (1993). Individual, familial, and organizational determinants of career plateau: An empirical study of the determinants of objective and subjective career plateau in a population of Canadian managers. *Group & Organization Management, 18*(4), 411–435.

Tremblay, M., & Roger, A. (2004). Career plateauing reactions: The moderating role of job scope, role

ambiguity and participation among Canadian managers. *The International Journal of Human Resource Management, 15*(6), 996–1017.

Veiga, J. F. (1981). Plateaued versus non-plateaued managers: Career patterns, attitudes, and path potential. *Academy of Management Journal, 24*(3), 566–578.

Wang, S. (2011). Problems with civil servant performance appraisal and counter measures. *Journal of Shandong University* (Philosophy and Social Science), (1), 25–31.

[王骚 . (2011). 公务员绩效考核中的问题与对策分析 . *山东大学学报* (哲学社会科学版), (1), 25–31.]

Wang, Z. Q., & Zhang, H. (2007). Research into relationship between career plateau and job satisfaction of enterprise employees. *Journal of Chongqing Technology and Business University (Social Sciences Edition), 24*(6), 56–59.

[王竹青 , 张慧 . (2007). 企业员工的"职业高原"和工作满意度的关系研究 . *重庆工商大学学报 (社会科学版), 24*(6), 56–59.]

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92*(4), 548–573.

Xie, B. G., & Long, L. R. (2008). The effects of career plateau on job satisfaction, organizational commitment and turnover intentions. *Acta Psychologica Sinica, 40*(8), 927–938.

[谢宝国 , 龙立荣 . (2008). 职业生涯高原对员工工作满意度、组织承诺、离职意愿的影响 . *心理学报，40*(8), 927–938.]

Xie, B. G., Long, L. R., & Zhao, Y. J. (2008). Development of career plateau questionnaire and research on its validity and reliability. *Chinese Journal of Clinical Psychology, 16*(4), 344–347.

[谢宝国 , 龙立荣 , 赵一君 . (2008). 职业高原问卷的编制及信效度研究 . *中国临床心理学杂志，16*(4), 344–347.]

Yu, C. (2006). Research on construct of career plateau and the relationship between it and job performance of knowledge employees. *Studies in Science of Science, 24*(6), 929–933.

[余琛 . (2006). 知识型员工的职业高原与工作绩效的关系研究 . *科学学研究* , 24(6), 929–933.]

Zhang, W. Q. (2006). *On the career management of common civil servants in China government* (Unpublished master's thesis). Northeastern University, Shenyang.

[张文勤 . (2006). *论我国普通公务员职业发展管理* (硕士学位论文). 东北大学 , 沈阳 .]

Zhang, Z. S. (2005). International comparison and reference research of civil servant's career development management. *Journal of US-China Public Administration, 2(5)*, 1–6.

[张再生 . (2005). 公务员职业生涯发展管理的国际比较与借鉴研究 . *美中公共管理* , 2(5), 1–6.]

Zhao, Y. R. (2012). *The relationship of middle school teachers´ career plateau, job burnout and turnover intention* (Unpublished master's thesis). Shandong Normal University, Jinan.

[赵寅汝 . (2012). *中学教师职业生涯高原、职业倦怠和离职倾向的关系* (硕士学位论文). 山东师范大学 , 济南 .]

The career plateau of Chinese public servants: Construct, measurement and its psychological and behavioral influence

WANG Zhongjun[1]; LONG Lirong[2]; LIU Lidan[3]; HUANG Xiaohua[4]; JIA Wenwen[1]; LI Lu[1]; MA Hongyu[1]

([1] Key Laboratory of Adolescent Cyberpsychology and Behavior, Ministry of Education; School of Psychology, Central China Normal University;

Key Laboratory of Human Development and Mental Health of Hubei Province, Wuhan 430079, China)

([2] School of Management, Huazhong University of Science and Technology, Wuhan 430074, China)

([3] School of Humanities, Hubei University of Chinese Medicine, Wuhan 430065, China)

([4] United Front Work Department, CPC Jiangxi Provincial Committee, Nanchang 330006, China)

Abstract　The career plateau is a serious problem for Chinese public servants, especially for those who at the lower level of public organizations, which occurs when an individual has limited vertical and horizontal career mobility. Since career plateau is of great influence on individual career development as well as organizational efficiency, this study investigated the career plateau phenomenon of Chinese public servants. Firstly, the typical features of Chinese public servants' career plateau were found with the content analysis of data collected from 16 public servants, using structured interviews and open−ended questionnaires for other 43 public servants from various public organizations. Meanwhile, the initial items for Public Servants Career Plateau Questionnaire (PSCPQ) were also established. Secondly, data collected from a sample of 279 public servants of diverse organizations was used to explore the CSCPQ's conceptual construct, reliability and validity by exploratory factor analysis (EFA). Thirdly, conducting Confirmatory Factor Analysis (CFA) for data from another larger sample of 517 public servants, the validity and reliability of the PSCPQ was confirmed. Eventually, this study investigated the relationship between career plateau and civil servants' occupational psychological and behavioral outcomes, such as organizational commitment, occupational burnout and withdrawal behavior in workplace by hierarchical multiple regression analysis for data of 520 public servants and another sample (n=230) from three public organizations. The research results indicated the career plateau of public servants was a two−dimension structure in Chinese public organizations context. The dimensions include the stagnation of promotion and the marginalization of position, which are completely different from the constructs of employees' career plateau of enterprises found in previous studies. The Cronbach α coefficients of the two dimensions are all above 0.80 and show a steady and acceptable status in different samples. The Cronbach α coefficients of the entire PSCPQ is also above 0.87 in different sample. These results showed that PSCPQ developed in this study had good psychometric reliability and high validity. Furthermore, the results showed that after controlling demographic variables, the stagnation of promotion of public servants was significantly positive related to emotional burnout and cynicism, and was not related to organizational commitment, professional efficiency as well as withdrawal behavior at work. After controlling demographic variables, the positional marginalization of public servants was significantly negative related

to organizational affective commitment, normative commitment and professional efficiency, and was positive related to organizational continuance commitment, emotional burnout, cynicism as well as withdrawal behavior. These results would be of enlightenment to the management practices and career development for public servants in Chinese public organizations. Finally, implications and directions for future research are discussed.

组织中主管－下属关系的运作机制与效果 *

王忠军[1]，龙立荣[2]，刘丽丹[3]

（[1] 华中师范大学心理学院暨湖北省人的发展与心理健康重点实验室，武汉 430079）

（[2] 华中科技大学管理学院，武汉 430074）

（[3] 湖北中医药大学社会科学部，武汉 430065）

摘　要　基于社会交换的理论视角，以下属关系投入－主管资源回报的概念架构来展现组织中主管与下属关系互动的实质，对主管－下属关系的运作效果与机制进行跨层次的实证研究。通过问卷法获得 54 个工作群体的 426 名下属与主管的对偶数据，基于 HLM 分析的结果表明：下属在工作之余对主管的私人关系投入不仅能直接获得主管的工具性资源回报与情感性资源回报，还能通过领导－成员交换 (LMX) 间接地获得主管的工具性与情感性资源回报，而在工作群体内基于私人关系进行人力资源管理决策的特征对主管与下属之间的关系互动与关系质量也存在一定程度的影响。

关键词　主管－下属关系；关系投入；资源回报；领导－成员交换；关系导向人力资源管理

分类号　B849：C93

1　问题的提出

在中国社会，关系现象充斥于人们的日常生活、经济活动以及组织行为之中。关系管理作为中国式管理的核心，备受企业实践者与组织研究者的关注。随着现代化的发展，许多组织在管理中力图淡化关系的影响，却无法动摇传统儒家文化和价值观的社会基础，正是这种文化与价值塑造了中国的组织管理行为。许多研究者认为关系的作用在未来中国以及东亚国家及其组织中将一直持续下去 (Lovett, Simmons, & Kali, 1999; Millington, Eberhardt, & Wilkinson, 2005; Yang, 2002)。在组织的各类关系中，最重要和吸引人的是上下级之间的关系，如主管－下属关系。与西方不同，中国员工普遍重视与领导、同事建立并维持良好的私人关系，而处理并维护好与下属的关系也是管理者有效管理下属的关键要素 (Law, Wong, Wang, & Wang, 2000)。因此，基于中国社会文化情境，探究组织中主管－下属关系运作的机制与效能，对于理解组织中的关系现象，丰富关系管理有重要的理论与实践意义。

1.1　文献回顾

由中国传统社会文化所孕育的关系 (guanxi) 概念一直是华人学者进行本土心理学建构，并循此了解中国人心理与行为的核心概念，学界从概念层面对关系的文化意涵、定义、类型特征、互动法则等进行了广泛探讨（周丽芳，2002）。在管理学领域，研究者发

现中国人的关系展现方式及其结果与西方的人际关系 (interpersonal relationship) 有很大差异，关系在中国人的商业活动、企业管理与组织行为中扮演不可言喻的重要角色 (Xin & Pearce, 1996)。因此，由关系概念来透视组织管理成为中国式管理研究的最佳进路。

不过，对组织内部的关系研究存在多元层次与视角，如对偶层次 (dyad)、三方关系 (triads)、关系网络 (networks)。就对偶关系而言，大体有两类：一是上下级之间，如下属与主管、员工与领导，属于垂直型对偶关系；二是同级之间，如同事关系，高管团队中两两关系，属于水平型对偶关系。目前组织行为学的研究大多集中于前者。

对于主管 – 下属关系，早期研究集中于关系基础 (guanxi basis) 及其效能。即探讨主管与下属之间既定的特殊性关系连带（如血缘关系、九同关系等）对主管与下属的关系品质 (如亲信、友谊、认知性信任与情感性信任) 以及主管对下属的绩效评估的影响 (Farh, Tsui, Xin, & Cheng, 1998; Tsui & Farh, 1997; Xin, Farh, Cheng, & Tsui, 1998)。在上述研究基础上，Tsui、Farh 和 Xin(2000) 根据关系分类与互动原则的架构，提出了一个华人组织中关系与效能的概念模型，即"关系基础（家人、熟人、生人）→关系形式（义务、友谊、认同）→直接效果（人际信任、人际喜好、忠诚、偏私）→间接效果（职业生涯与事业成功）"。显然，关系基础的研究视角承袭的是费孝通"差序格局"的思想，关系大多被定义为一种"特殊性社会连带"(King, 1985; Yang, 1986)。在此定义下，关系经常被操作为一种二分变量，即要么存在某种类型的关系基础，要么不存在，并且不同的关系基础具有不同义务规范，会受到不同对待。但总体来看，关系基础与相关效果变量间的关系不是很稳定，其原因有二：其一，由关系基础到关系成分（情感、义务、工具等）的推论不够清晰，例如，即使过去有同宗或同学的关系基础，并不必然会出现某种形式的关系成分（周丽芳，2002）。其二，更重要的是，在现代组织中具有关系基础的对偶双方或网络成员的比例相当低，即组织成员之间"沾亲带故"的现象并不普遍，一定

程度上影响了研究结果的可靠性。例如在 Farh 等人 (1998) 的研究中，在全部 560 组主管与下属配对样本中没有发现师生关系，过去曾是上下级关系的仅占 0.04%，而同学、亲戚、同姓、同乡、同事、邻里关系出现的频次均在 2.10% 至 3.40% 之间。

因此，近期对于主管 – 下属关系的研究则逐渐转向下属与主管的私人关系质量 (guanxi quality) 及其对个体的积极影响。在该类研究中，主管 – 下属关系被定义为"主管与下属在工作时间之外通过非工作相关的行为活动而建立的私人关系质量" (Law et al., 2000)，"关系"是在互相满足关系双方个人目标的过程中所建立起来的，并为工具性目的服务。此类研究结果表明，主管与下属的私人关系质量能影响主管的管理决策，如对下属的晋升、奖酬分配、工作安排等 (Law et al., 2000)；能预测下属对主管的满意度以及对组织的情感承诺 (Wong, Tinsley, Law, & Mobley, 2003)；能让下属从主管处获得更多的关系性报酬 (guanxi payoff)，如奖酬分配、晋升机会和任务安排等，并且提升下属知觉的程序公平感 (Chen, Friedman, Yu, & Sun, 2008)；对员工的职业生涯发展也有着积极的影响（刘军，宋继文，吴隆增，2008）。

从关系质量视角的研究引出的思考是：下属要想与主管拥有良好的私人关系质量，就必然要有建立、维持和运作关系的行为活动。换言之，有了建立私人关系行为的投入，才可能具备良好的关系质量。而在以往研究中，对主管与下属关系质量的操作与测量也大多是基于双方的关系互动行为。事实上，关系概念的复杂性和丰富性也正是体现在关系的互动方面，即不具备特殊性关系连带的个体之间如何建立和发展关系，并会因此带来什么样的后果。因此，从关系行为(guanxi behavior)的视角来探讨主管 – 下属关系的运作机制及其效果是一个新的、重要的研究视角，而社会交换理论能为该视角提供理论解释。

1.2 社会交换视角的主管 – 下属关系

中国人的关系行为本质上是一种社会交换行为的观点早已得到学界的认同 (King, 1985; Hwang, 1987)，对于组织中主管 – 下属关系也不例外。根

据社会交换理论 (Blau, 1964; Foa & Foa, 1980)，本研究将组织中主管 – 下属关系 (supervisor–subordinate guanxi, SSG) 定义为："组织中下属通过工作范围之外的互动行为与其主管建立的非正式、特殊性社会交换关系"。由此定义引出的问题是：下属与主管关系交换的内容是什么？换言之，下属投入什么，主管相应地回报什么？

从下属方面来看，要建立、维持或经营与主管的特殊性私人关系，必然要付出一些成本，比如时间、金钱、情绪乃至机会成本，即关系投入的行为。比如 Law 等人 (2000) 通过问卷调查，搜集下属与其主管建立良好私人关系的各种行为活动，最后确立了六种最具代表性和有效性的行为。Wong 等人 (2003) 研究发现主管与下属关系互动行为主要表现在如下五个方面：社会活动、经济支持、优先照顾、节日庆祝和情绪支持。本研究提出"关系投入"(guanxi input, GI) 的概念，并将其界定为"为与主管建立良好的私人关系，下属在工作范围之外对其主管进行的各种时间、经济与情感的投入行为"。

而从主管方面来看，也会相应地给予下属各种资源回报。在权力距离较大的中国组织中，不同组织层级的资源是不平衡的。研究发现，中国企业的决策权力一般更多地集中于中高层，管理人员在员工选拔、薪酬和雇佣等人事决策上有着更大的影响力 (Wang & Heller, 1993)。下属与主管所拥有的资源差距在组织的中高层与基层之间表现得尤为明显 (王忠军，龙立荣，2009)。从社会资本（ social capital ）的观点来看，对于主管所拥有的资源，下属只有通过与主管建立较强的社会连带才能得以借用。因此，下属对主管的关系投入行为也才有了一个最基本的动因和条件。

根据社会资源理论 (Lin, 2001)，社会行动者主要有两类行动：①工具性行动（寻找和获得额外有价值资源），以及②情感性行动(寻找情感和支持的行动)，并且情感性行动往往比工具性行动更重要。而社会交换理论也认为，人类的社会交换行为不仅仅是工具性资源交换，还有情感性资源交换。因此，本研究认为，主管给予下属的回报主要有两种：一是工具性资源

回报，亦简称工具性回报 (instrumental output, IO)，本研究将其界定为"主管基于私人关系给予下属直接的、客观性物质利益或好处，如晋升机会、任务安排、奖金分配、绩效考评、工作支持等"。这一概念类似于其他研究者所谓的"关系性报酬" (Friedman, Chi, & Liu, 2006; Chen et al., 2008)。二是情感性资源回报，亦简称情感性回报 (affective output, AO)，本研究将其界定为"主管基于私人关系给予下属以间接的、主观性精神利益或好处，如接纳、友善、信任、认可、鼓励、关怀、宽容等"。 显然，以往中国人关系研究常将焦点置于工具性利益与义务性情感，而较少关注自我表露、内心交流、情感性支持等真实情感或情绪层面（周丽芳，2002）。从下属方面来看，下属与主管建立和维持关系，不仅仅是想获取主管的工具性资源，更想博取情感性回报。而从主管方面来看，工具性资源具有客观性和有限性，并且需要在不同下属之间来平衡，而情感性资源具有主观性和丰富性，在组织环境中，管理者对下属往往一手运用工具性资源，一手运用情感性资源，交互运作，以更好地管理和驾驭下属。因此，探究情感性回报是极为重要的，但以往的研究很少涉及。

根据以上分析，在差序格局、关系取向以及特殊主义的中国社会文化背景下，组织中主管 – 下属关系的实质表现为下属在工作之余对其主管进行关系投入，主管相应给予下属不同程度的工具性、情感性资源回报的互动过程。

1.3　研究假设

1.3.1　下属关系投入与主管资源回报

中国人常言，投之以桃，报之以李，来而不往非礼也。根据社会交换理论，若主管能感知到下属对其关系行为的投入，基于以下三个原因：①回报的要求、②信任、③互惠原则，下属的投入终会有回报。譬如 Zhang 和 Yang(1998) 研究发现中国企业管理者对于奖金分配的决策不仅受公平原则的影响，还受到关系的影响。Law 等人 (2000) 在对中国大陆的189 对主管 – 下属的对偶关系研究发现，关系会影响主管的管理决策，与主管拥有良好的私人关系的下属

能获得更多的晋升机会、更多的奖金分配和更好的工作安排。Zhou 和 Martocchio(2001) 的研究报告中国管理者会给予那些与其拥有良好关系的下属更多的非货币性报酬。Chen 和 Tjosvold(2007) 的研究也发现中国企业员工与管理者的个人关系能带来更好的工作安排以及晋升机会。Chen 等人 (2008) 的研究也发现，良好的主管–下属关系能换来更好的关系性报酬。不过正如前文的分析，以往研究大多关注关系所带来的工具性资源回报，而很少关注情感性资源回报。而由社会交换理论不难推出，下属对主管的私人关系投入行为也能获得主管的情感性资源回报，因为下属与其主管建立良好私人关系的各种行为中存在许多情感性投入的成分，这在 Law 等人 (2000) 的研究中已有所展现。对于中国人而言，直接的、赤裸裸的物质利益交换往往让人难以接受，而最有效的方式则是在利益交换的过程中渗透情感的投入与交换。比如在 Wong 等人 (2003) 研究中所发现的主管与下属关系互动中也无一不展现了情感互动的成分（如社会活动、优先照顾、节日庆祝和情绪支持）。因此，本研究提出如下假设：

假设 1-1：下属的关系投入正向影响主管的工具性回报。

假设 1-2：下属的关系投入正向影响主管的情感性回报。

1.3.2 领导–成员交换的中介作用

现有文献大多认为主管–下属关系与领导–成员交换关系 (Leader-member exchange, LMX) 是彼此独立的概念 (Law et al., 2000)。其主要的区别是：LMX 是建立在工作职责上的正式的工作关系，被定义为领导与成员彼此之间在工作上展现出信任、忠诚、情感、贡献与责任的行为 (Graen & Uhl-Bien, 1995); 而主管–下属关系反映的是工作范围之外的、非正式的私人关系 (Wong et al., 2003)。Hui 和 Graen(1997) 曾深入比较过关系概念与 LMX 概念的区别。基于此，以往的研究大多单独或并行地考察主管–下属关系和 LMX 各自的作用机制，并加以比较（比如 Law et al., 2000；Chen & Tjosvold, 2007），却很少有研究探讨二者之

间的关系。那么，LMX 在主管与下属的私人关系交换过程中起着什么样的作用呢？以往的研究发现，下属影响主管的行为，如与主管结盟、相互交换、逢迎主管等会影响 LMX 的质量 (Deluga, 1994)。而 LMX 也会进一步影响员工的晋升、工作安排和薪酬 (Wakabayashi, 1988; Chen & Tjosvold, 2007)。根据中国社会背景来看，关系在一定程度上体现了中国人所谓"做人"的一面。与西方不同，中国人是极其重视"做人"的，这直接来源于"会做人"的好处以及"不会做人"的坏处。因此，下属与主管在工作范畴之外发展出的具有强烈的"组织规定外"、"私人情感"色彩的关系会渗透到正常工作中，从而在组织制度内发挥作用（刘军等, 2008）。由此推论，下属在工作之余对主管的私人关系投入行为，可能会对彼此在工作场所中的 LMX 关系质量产生一定程度的积极影响，而良好的 LMX 关系质量也会进一步为下属带来各种情感性和工具性资源回报。基于以上分析，本研究提出如下假设：

假设 2-1：下属的关系投入通过 LMX 间接地影响主管的工具性回报。

假设 2-2：下属的关系投入通过 LMX 间接地影响主管的情感性回报。

1.3.3 关系导向人力资源管理的调节作用

主管与下属的关系交换与运作虽然是个体间的互动行为，但却是嵌入在群体或组织的背景中，受到群体或组织特征的约束，而这种嵌入性在个体层面的研究中往往被忽视。在中国组织中，关系的作用会渗透进组织的各项管理决策之中，成为"正式法制支持的替代品" (Xin & Pearce, 1996)。Chen, Chen 和 Xin(2004) 因此提出"关系导向人力资源管理"（Guanxi-based human resources management practice, GHRM）概念，指的是人力资源管理决策中以私人关系为基础的总体状况。需要说明的是，该概念既适用于组织层面，也适用于群体层面，本研究中采取的是群体层面的概念。在注重制度规范和公平正义的组织环境下，关系导向人力资源管理被认为具有众多负面性，比如破坏程序公平 (Chen et al., 2008)，降低

员工对组织管理的信任 (Chen, Chen & Xin, 2004)，损害员工的角色内和角色外绩效 (Hsu & Wang, 2007)。总之，人力资源管理决策的关系导向越强，说明 "人治" 气氛越浓厚，基于私人关系的弹性操作空间越大，同时也意味着制度规范性越差，在这样的环境下，主管根据私人关系而给予下属差别化的特殊对待的现象会拥有 "制度合法性" 的背景，并得以强化。因此，本研究提出如下假设：

假设 3-1：工作群体的关系导向人力资源管理对下属关系投入与主管工具性回报的关系具有正向调节作用。

假设 3-2：工作群体的关系导向人力资源管理对下属关系投入与主管情感性回报的关系具有正向调节作用。

2　研究方法

2.1　被试与调查程序

由于本研究属于跨层次的研究，涉及个体与群体两个层面的数据搜集，因此，问卷调查均以工作群体 (work group) 为抽样单位，并在每个工作群体（部门）内，采取了主管与下属的二元对偶研究设计。调查包含两份问卷，分别由部门内的员工及其直接主管填写。在内容上，下属问卷包括自评的任务绩效、主管资源回报、LMX、关系导向人力资源管理；部门主管则仅需填答不同下属的关系投入。为了保证问卷的隐匿性以及数据的主管–下属配对，采用了一个编码系统，以匹配主管评定与下属回答。之所以采用上述研究设计，主要出于以下考虑：

第一，主管对于下属的关系投入行为一般会有感知，并会据此相应地给予下属各种资源回报；同样，下属对于主管给予的资源回报也会有直接的感知，其感知的结果会进一步影响双方的后续互动。第二，一名主管需要评价多名下属，主管的负担会较重，因此仅要求主管填答项目数量相对较少的下属关系投入问卷。出于同样的考虑，任务绩效也由下属自评。第三，主管与下属 "错位式" 的互评可在一定

程度上降低问卷项目敏感性带来的心理压力。总之，上述研究设计既考虑了研究需要，又兼顾了可行性，从不同来源获取数据。

被试来自湖北、江西、北京、上海、广东地区 8 家企业中的不同工作群体（工作部门）。在调查程序上，首先，研究人员与企业人力资源部门一起确定了调查的部门，主要是企业中层部门。判定群体的依据有：①不同员工属于同一工作部门；②不同员工拥有一个共同的直接主管；③他们长时期在一起工作。然后，研究人员在企业助手的带领和协助下，进行现场调查。最后，由研究人员当场收回问卷，回收问卷后进行主管与下属问卷的配对组合。以现场调查方式所获得的样本占全部有效样本的 81.50%，这在一定程度上能保证数据的质量，而少量委托调查则给受托者及其单位的人力资源部门提供了指导语。

研究者对有效数据进行了筛选：①首先剔除了空白过多、反应倾向过于明显的问卷；②然后剔除了下属人数过少的群体样本（少于 5 人）。最终回收了 54 个有效群体样本，总共包含 426 份有效个体问卷，平均每个群体包含 8 人，人数最少的群体有 6 人，人数最多的群体有 13 人。主管–下属匹配后的有效填答率为 82%，其中有 83.10% 的被试与其直接主管保持的上下级关系年限在 1 年以上。在有效样本中，国有企业 40.80%，民营企业 55.50%，外资企业 3.70%；男性占 55.70%，女性占 44.30%；25 岁及以下占 20.50%，26~30 岁占 13.90%，31~35 岁占 14.90%，36~40 岁占 20.80%，40 岁以上者占 30%；管理岗位占 34.70%，生产岗位占 16.70%，技术岗位占 22.90%，销售岗位占 8.50%，行政后勤占 17.20%。

2.2　研究工具

2.2.1　下属关系投入

下属关系投入采用经本研究修订过的单维度问卷。修订过程如下：采用与 Law 等人 (2000) 编制关系问卷相同的方法，首先通过对来自不同企业的 27 名员工进行开放式问卷调查，搜集员工在工作之余与其直接主管建立并保持良好私人关系的行为，结果

发现大部分行为与 Law 等人 (2000) 的研究相似，但也有少部分行为包含在 Chen 等人 (2008) 和 Wong 等人 (2003) 的问卷项目之中。因此，本研究在综合以上相关问卷项目的基础上，初步编制了包含 9 个项目的下属关系投入问卷，为避免被试填答问卷时的"趋中性"，采取 Likert 6 点计分，1 表示"非常不符合"，6 表示"非常符合"。接下来，对来自江西 5 家企业的共 260 名员工进行初试，获得 211 名员工的有效数据，其中 99.50% 的员工与其主管保持了 1 年以上的上下级关系。进行探索性因素分析 (EFA) 后发现下属关系投入问卷为单维度，解释的变异量为 65.15%。

由于正式施测中的关系投入问卷是由主管来填答，所以我们从主管的角度对问卷项目的人称进行了相应修改，比如："该职工在平时会打电话或上门拜访我"、"该职工总是主动地与我交流他（她）的想法、问题、需要和感受"。对正式施测的样本（n=426）数据进行验证性因素分析 (CFA)，下属关系投入问卷的单维度模型的各项拟合指数均达到或接近临界值，具体如下：x^2=109.89，df=27，RMSEA=0.07，SRMR=0.05，IFI=0.91，CFI=0.91，NFI=0.90，NNFI=0.90，这表明问卷具有较好的结构效度。下属关系投入问卷的 Cronbach α 信度系数为 0.93，符合测量学的标准。此外，本研究还表明该问卷具有较好的效标效度。

2.2.2 主管资源回报

主管资源回报问卷采用本研究自编的问卷。编制过程如下：首先对来自不同企业的 27 名员工进行开放式问卷调查，请被试列举"下属与直接主管建立良好私人关系后，主管会给下属带来哪些利益或好处"。对开放式问卷调查的资料进行内容分析后，初步编制了包含 15 个项目的主管资源回报问卷，为避免被试填答问卷时的"趋中性"，采取 Likert 6 点计分，1 表示"完全不同意"，6 表示"完全同意"。

接下来，通过对初试样本（n=211）的有效数据进行探索性因素分析 (EFA)，结果表明主管资源回报问卷具有十分清晰的两维度结构：一是工具性回报，包含 6 个项目，比如："他（她）会尽量给我安排

我期望的工作岗位"、"他（她）会想方设法提拔我"；二是情感性回报，包含 7 个项目，比如："在生活中，他（她）很关心照顾我"、"他（她）会与我分享他（她）的经验、想法和感受"。两维度解释的变异量为 68.09%。

最后，利用正式施测的样本（n=426）数据进行验证性因素分析 (CFA)，主管资源回报的两维度模型的各项拟合指数均达到临界值，具体如下：x^2=231.68，df=64，RMSEA=0.06，SRMR=0.43，IFI=0.92，CFI=0.92，NFI=0.91，NNFI=0.90，这表明问卷具有较好的结构效度。信度分析表明，工具性回报和情感性回报的 Cronbach α 系数分别为 0.90 和 0.92，总问卷的 α 系数为 0.94，说明问卷的信度质量较好。此外，本研究还表明问卷具有较好的效标效度。

2.2.3 领导 – 成员交换

对于如何测量 LMX，学界还存在争议，主要源于其结构是单维的还是多维的差异。但 Liden 和 Maslyn(1998) 提出 LMX 的维度不一定需要得到一个确定的模式，而是需要与不同的考察目的和结果变量挂钩。Schriesheim，Castro 和 Coglister(1998) 运用元分析 (meta-analysis) 技术，检验了各种量表的内部一致性，结果表明 Graen 和 Uhl-bien(1995) 研制的 7 个项目量表具有最高的信度和效度，简称 LMX-7。由于本研究主要关心组织中领导 – 成员交换关系的质量，而非不同方面的交换内容。因此，本研究采用被广泛应用的 LMX-7 量表，并采用 Likert 6 级计分，1 表示"完全不同意"，6 表示"完全同意"。在本研究中，该问卷的 α 系数为 0.92。

2.2.4 关系导向人力资源管理

本研究中关系导向人力资源管理是一个群体层次的变量，而对其测量是通过对群体中的个体的测量来完成的，这里面就有一个指称迁移问题。根据 Chan（1998）所提出的"指称迁移共识模型"（referent-shift consensus model），在测量时，所用的项目不是群体中单个成员的行为描述，而必须把所有成员作为一个整体来看待，以整体为出发点来

描述群体成员的行为。在本研究中，关系导向人力资源管理采用 Chen，Chen 和 Xin(2004) 开发的量表，原量表共 5 个项目，α 系数为 0.93。所有项目均以"在我所工作的部门内"为指称。此量表在 Chen 等人 (2008) 的一项研究中被修订为 4 个项目，α 系数为 0.88。上述两个版本的量表均为英文，本研究对其进行翻译和回译后，综合了两个版本的项目，获得了 7 个项目，其中的一个项目"在我所工作的部门内，培训发展机会的获得依靠与主管的关系"为本研究新加入的一个项目。采用 Likert 6 点计分，1 表示"非常不符合"，6 表示"非常符合"。

对关系导向人力资源管理问卷 7 个项目进行了探索性因素分析，结果得到单一因素，解释的变异量为 70.90%，问卷的 Cronbach α 信度系数为 0.93，这一结果同样表明关系导向人力资源管理问卷具有较好的信度和结构效度。此外，我们运用方差分析检验了关系导向人力资源管理在企业性质上的差异，结果发现不同性质的企业之间存在显著地差异，其中关系导向人力资源管理在国有企业的表现程度最高，民营企业次之，外资企业的关系导向最低（$M_{国有企业}$=3.53，$M_{民营企业}$=3.03，$M_{外资企业}$=2.54，F=11.18，$p < 0.001$），该结果与实际情况基本相符，也说明关系导向人力资源管理问卷具有较好的同时效度。

本研究中，关系导向人力资源管理用于在群体层次上代表群体的人力资源管理决策特征与氛围。对应于每一个工作群体，其关系导向应当是唯一的，所以有必要将群体中个体提供的评估数据汇聚到群体层次。$ICC(1)$、$ICC(2)$ 和 R_{wg} 是三个最常用的用于判断个体数据汇聚是否可靠的指标，本研究同时考察三者。为了判断的一致性，我们先通过方差分析

(ANOVA) 进行了组间差异性检验，结果组内相似性高于组间相似性，即不同工作群体之间存在显著的差异 $F(53,372) = 4.81, p < 0.001$。计算得到 54 个工作群体的 R_{wg} 值在 0.27~0.96 之间，尽管少数群体的 Rwg 值较低，但均值为 0.75，高于 0.70 的标准 (Dixon & Cunningham, 2006)。同时，本研究计算所得 $ICC(1)$ 为 0.32，在 James (1982) 推荐的 0 到 0.5 的临界值范围之内，这表明变量在各群体中有充足的内部同质性；ICC(2) 为 0.79，大于 Klein 等人 (2000) 推荐的临界值 0.70，这表明采用个体的平均数作为群体变量的指标的可信度较高。总之，以上结果均一致表明，可以用群体中个体知觉到的关系导向人力资源管理数据的平均数作为群体层面变量的观察值。

除了以上关键变量外，本研究还控制了可能会影响主管给予下属资源回报的一个重要变量，即员工的任务绩效。任务绩效采用员工自评式问卷，包含 4 个项目，来源于 Williams 和 Anderson(1991) 编制的任务绩效量表，这 4 个项目分别是："和同事相比，我的工作成绩比较优秀"、"我的领导对我的工作成绩比较满意"、"同事对我的工作成绩评价比较高"、"我的工作成绩经常受到单位的表扬"。本研究对以上 4 个项目进行探索性因素分析，结果得到单一因素，解释的变异量为 67.98%，问卷的 α 系数为 0.84。

2.3 统计方法

采用 SPSS 11.5 进行描述统计、相关分析、探索性因素分析和信度分析，采用 LISREL 8.30 进行验证性因素分析，采用多层线性模型 HLM 6.02 对研究假设进行检验。

表 1　各研究变量的平均数、标准差与相关矩阵

变量（n=426）	M	SD	1	2	3	4	5
1. 任务绩效	4.15	0.82					
2. 关系投入	3.09	1.10	0.36**				
3. 工具性回报	3.47	1.05	0.40**	0.59**			
4. 情感性回报	3.96	0.99	0.36**	0.50**	0.69**		
5. LMX	3.75	0.98	0.43**	0.58**	0.74**	0.80**	
6. 关系导向人力资源管理	3.34	1.19	0.07	−0.06	−0.04	−0.24**	−0.18**

注：关系导向人力资源管理为个体层面的数据；**$p < 0.01$,*$p < 0.05$。

表 2 下属关系投入对主管资源回报的影响

因变量	模型	参数估计							
		γ_{00}	γ_{10}	γ_{20}	σ^2	τ_{00}	τ_{11}	τ_{22}	作用
工具性回报	**M**1：零模型 L1：$IO_{ij}=B_{0j}+r_{ij}$ L2：$B_{0j}=\gamma_{00}+\mu_{0j}$	3.47***			0.88	0.24***			
	M2：任务绩效→工具性回报 L1：$IO_{ij}=B_{0j}+B_{1j}(TP_{1ij})+r_{ij}$ L2：$B_{0j}=\gamma_{00}+\mu_{0j}$ $B_{1j}=\gamma_{10}+\mu_{1j}$	3.47***	0.47***		0.69	0.26***	0.14**		0.22
	M3：关系投入、任务绩效→工具性回报 L1：$IO_{ij}=B_{0j}+B_{1j}(GI_{1ij})+B_{2j}(TP_{2ij})+r_{ij}$ L2：$B_{0j}=\gamma_{00}+\mu_{0j}$ $B_{1j}=\gamma_{10}+\mu_{1j}$ $B_{2j}=\gamma_{20}+\mu_{2j}$	3.47***	0.43***	0.27**	0.48	0.29***	0.06**	0.10**	0.30
情感性回报	**M**1：零模型 L1：$AO_{ij}=B_{0j}+r_{ij}$ L2：$B_{0j}=\gamma_{00}+\mu_{0j}$	3.95***			0.69	0.30***			
	M2：任务绩效→情感性回报 L1：$AO_{ij}=B_{0j}+B_{1j}(TP_{1ij})+r_{ij}$ L2：$B_{0j}=\gamma_{00}+\mu_{0j}$ $B_{1j}=\gamma_{10}+\mu_{1j}$	3.95***	0.34***		0.51	0.32***	0.22**		0.26
	M3：关系投入、任务绩效→情感性回报 L1：$AO_{ij}=B_{0j}+B_{1j}(GI_{1ij})+B_{2j}(TP_{2ij})+r_{ij}$ L2：$B_{0j}=\gamma_{00}+\mu_{0j}$ $B1j=\gamma_{10}+\mu_{1j}$ $B2j=\gamma_{20}+\mu_{2j}$	3.95***	0.37***	0.18*	0.39	0.34***	0.02*	0.17***	0.24

注：① ***$p < 0.001$,**$p < 0.01$,*$p < 0.05$。② σ^2 为水平 1 的残差；τ_{00} 为截矩残差，即 μ_{0j}；τ_{11} 和 τ_{22} 为斜率残差，即 μ_{1j} 和 μ_{2j}；③作用 =（原始残差－条件残差）/原始残差；④ IO 为工具性回报，AO 为情感性回报，TP 为任务绩效，GI 为关系投入。

3 研究结果

3.1 描述性统计及相关分析结果

表 1 呈现的是本研究中涉及的关键变量的描述性统计和相关分析结果。

3.2 下属关系投入对主管资源回报的影响

在运用多层线性模型 (HLM) 对假设进行验证时，本研究将下属的任务绩效作为一个关键的控制变量纳入 HLM 分析之中，表 2 为分析的结果。由表 2 可知，任务绩效对工具性回报具有显著的正向预测作用 (模型 M2，$\gamma_{10}=0.47$，$p < 0.001$)，控制变量解释的方差为 0.22。在模型 M3 中，当同时纳入关系投入和任务绩效时，任务绩效对工具性回报仍具有显著的正向预测作用 ($\gamma_{20}=0.27$，$p < 0.01$)，而关系投入对工具性回报同样具有更为显著的正向影响 ($\gamma_{10}=0.43$，$p < 0.001$)，关系投入解释的方差为 0.30，这表明控制了任务绩效后，关系投入对工具性回报具有显著的正向作用，假设 1–1 得到验证。另外，由表 2 可知，任务绩效对情感性回报具有显著的正向预测作用 (模型 M2，$\gamma_{10}=0.34$，$p < 0.001$)，控制变量解释的方差为 0.26。在模型 M3 中，当纳入关系投入和任务绩效一起分析时，任务绩效对情感性回报仍具有显著的正向预测作用 ($\gamma_{20}=0.18$，$p < 0.05$)，而关系投入对情感性回报同样具有更为显著的正向影响 ($\gamma_{10}=0.37$，$p < 0.001$)，关系投入解释的方差为 0.24，这表明控制了任务绩效后，关系投入对情感性回报具有显著的正向作用，假设 1–2 得到验证。

3.3 LMX 的中介作用

在表 3 中，在模型 M1 中，关系投入对领导–成员交换 (LMX) 具有显著的正向作用 ($\gamma_{10}=0.46$，$p < 0.001$)。在模型 M3 中，LMX 对工具性回报具有显著的正向作用 ($\gamma_{10}=0.83$，$p < 0.001$)。而在模型 M4 中，当将关系投入和 LMX 作为预测变量一起纳入模型中时，LMX 对工具性回报具有显著的正向作用 ($\gamma_{10}=0.69$，$p < 0.001$)，而关系投入虽然对工具性回报也具有显著的正向预测作用 ($\gamma_{20}=0.21$，$p < 0.01$)，但是其影响系数要比模型 M2 中的系数 ($\gamma_{10}=0.50$，$p < 0.001$) 明显降低。综合以上结果，据此可以推论，LMX 在关系投入与工具性回报之间起着部分中介的作用，假设 2–1 得到验证。

表 3   LMX 在下属关系投入对主管资源回报影响中的中介作用

模型	参数估计						
	γ_{00}	γ_{10}	γ_{20}	τ_{00}	τ_{11}	τ_{22}	σ^2
M1：关系投入→LMX	3.74[a]	0.46[a]		0.34[a]	0.04[b]		0.40
M2：关系投入→工具性回报	3.47[a]	0.50[a]		0.28[a]	0.06[b]		0.54
M3：LMX→工具性回报	3.47[a]	0.83[a]		0.30[a]	0.06[b]		0.40
M4：LMX、关系投入→工具性回报	3.47[a]	0.69[a]	0.21[a]	0.31[a]	0.14[a]	0.06[a]	0.33
M5：关系投入→情感性回报	3.95[a]	0.42[a]		0.33[a]	0.04[b]		0.45
M6：LMX→情感性回报	3.95[a]	0.73[a]		0.36[a]	0.09[a]		0.25
M7：LMX、关系投入→情感性回报	3.95[a]	0.69[a]	0.10[c]	0.36[a]	0.15[a]	0.06[a]	0.22

注：① a 为 $p < 0.001$，b 为 $p < 0.01$，c 为 $p < 0.05$。② σ^2 为水平 1 的残差；τ_{00} 为截矩残差，即 μ_{0j}；τ_{11} 和 τ_{22} 为斜率残差，即 μ_{1j} 和 μ_{2j}；③ IO 为工具性回报，AO 为情感性回报，GI 为关系投入，LMX 为领导 – 成员交换。
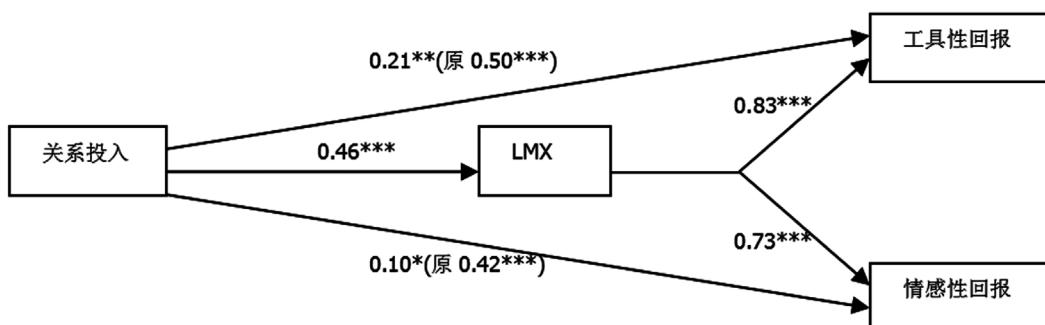
图 1   主管 – 下属关系的社会交换模型的实证研究结果（基于 HLM 分析的路径图）

注：*** $p < 0.001$，** $p < 0.01$，* $p < 0.05$；原：代表未加入中介变量时的影响系数

另外在模型 M6 中，LMX 对情感性回报具有显著的正向作用（$\gamma_{10}=0.73$，$p < 0.001$）。而在模型 M7 中，将关系投入和 LMX 作为预测变量一起纳入模型中时，LMX 对情感性回报具有显著的正向作用（$\gamma_{10}=0.69$，$p < 0.001$），而关系投入虽然对情感性回报也具有显著的正向预测作用（$\gamma_{20}=0.10$，$p < 0.01$），但是其影响系数要比模型 M5 中的系数（$\gamma_{10}=0.42$，$p < 0.001$）明显降低。综合以上结果，据此可以推论，LMX 在关系投入与情感性回报之间仍起着部分中介的作用。因此，本研究的假设 2–2 也得到验证。图 1 为基于上述研究结果而呈现的综合模型图。

3.4 关系导向人力资源管理的调节作用

由表 3 的结果可知，关系投入和 LMX 对工具性回报均具有显著的正向预测作用，但从表 4 的结果表明，关系导向人力资源管理对关系投入与工具性回报之间的关系系数（即斜率）不具有显著的调节作用（$\gamma_{21}=-0.00$，$p > 0.05$），同样地，关系导向人力

资源管理对 LMX 与工具性回报之间的关系系数的调节作用也不显著（$\gamma_{11}=0.13$，$p > 0.05$）。本研究的假设 3–1 没有得到验证。此外，在表 3 的结果中，关系投入和 LMX 对情感性回报均具有显著的正向预测作用，但在表 4 的结果中，关系导向人力资源管理对关系投入与情感性回报之间的关系系数（即斜率）不具有显著的调节作用（$\gamma_{21}=-0.01$，$p > 0.05$），同样，关系导向人力资源管理对 LMX 与情感性回报之间的关系系数的调节作用也不显著（$\gamma_{11}=0.05$，$p > 0.05$）。本研究的假设 3–2 也没有得到验证。

4 讨论

4.1 关系运作的个体效能

在现代企业组织中，工作绩效往往是极其重要的资源分配标准，比如很多企业实行绩效薪酬制度。本研究的结果也证明，下属的任务绩效对主管的工具

表 4 关系导向人力资源管理的调节作用模型

模型	参数估计									
	γ_{00}	γ_{01}	γ_{10}	γ_{11}	γ_{20}	γ_{21}	σ^2	τ_{00}	τ_{11}	τ_{22}
M1：对工具性回报的调节作用模型	3.32^{a}	0.05	0.26	0.13	0.22	−0.00	0.33	0.32^{a}	0.13^{a}	0.06^{a}
L1：$IO_{ij}=B_{0j}+B_{1j}(LMX_{1ij})+B_{2j}(GI_{2ij})+r_{ij}$										
L2：$B_{0j}=\gamma_{00}+\gamma_{01}(GXHRM_{ij})+\mu_{0j}$										
$\quad B_{1j}=\gamma_{10}+\gamma_{11}(GXHRM_{ij})+\mu_{1j}$										
$\quad B_{2j}=\gamma_{20}+\gamma_{21}(GXHRMij)+\mu_{2j}$										
M2：对情感性回报的调节作用模型	4.77^{a}	−0.25	0.51^{c}	0.05	0.13	−0.01	0.22	0.33^{a}	0.15^{a}	0.06^{a}
L1：$AO_{ij}=B_{0j}+B_{1j}(LMX_{1ij})+B_{2j}(GI_{2ij})+r_{ij}$										
L2：$B_{0j}=\gamma_{00}+\gamma_{01}(GXHRM_{ij})+\mu_{0j}$										
$\quad B_{1j}=\gamma_{10}+\gamma_{11}(GXHRM_{ij})+\mu_{1j}$										
$\quad B_{2j}=\gamma_{20}+\gamma_{21}(GXHRM_{ij})+\mu_{2j}$										

注：①零模型见表 2，直接作用模型见表 3，IO 为工具性回报，GI 为关系投入，LMX 为领导 – 成员交换，GXHRM 为关系导向人力资源管理。② σ^2 为水平 1 的残差；τ_{00} 为截矩残差，即 μ_{0j}；τ_{11} 和 τ_{22} 为斜率残差，即 μ_{1j} 和 μ_{2j}；③ a 为 $p < 0.001$，b 为 $p < 0.01$，c 为 $p < 0.05$。

性资源回报和情感性资源回报均有显著的正向影响。但是当我们将下属对其主管的私人关系投入与任务绩效一起去预测主管的资源回报时，让人意外的是，关系投入的解释力明显大于任务绩效，说明下属在工作之余的关系投入行为对主管资源回报有着重要的影响力。这一研究结果与 Law 等人 (2000)、Chen 等人 (2008) 以及刘军等人 (2008) 的研究结论一致。正如 Warner(1993) 所指，尽管技术和规范在中国组织中已经变得更为必要，但关系的重要性在中国社会仍然占据主导地位。该研究结果也凸现了关系在当代中国组织资源分配中仍占据重要地位，也给本研究从关系行为和社会交换的理论视角来审视组织中主管 – 下属关系提供了实证支持。

从社会交换的观点来看，组织中的主管与下属之间的关系投入与资源回报应是一种基于人情的社会交换行为，并具有长期互动的性质，而非一次性的交换。由于主管与下属之间身份、位阶、职权以及占有资源上的差异，主管与下属之间发展的私人关系很难像生活中单纯的朋友关系（以情感支持与寄托为主）那样简单。在组织中，下属对主管的关系投入既需要满足情感性支持的需要，更隐含着工具性回报的期待，这一点在本研究中也能得到反映，比如工具性回报与情感性回报的相关为 $0.69(p < 0.01)$，下属关系投入对主管工具性资源回报的影响系数为 $0.50(p$

$< 0.01)$，对主管情感性资源回报的影响系数为 $0.42(p < 0.01)$。总之，从本研究中可得到的启示与刘军等人 (2008) 的研究相同，即在中国组织中，除了工作上的努力与付出以外，发展与上级在生活上更为密切的私人关系更是下属不可忽视的。

4.2 关系运作的内在机制

主管 – 下属之间关系投入与资源回报的交换行为有着怎样的内在机制？本研究的结果表明：下属的关系投入除了对主管的工具性回报与情感性回报有着直接的影响外，还可以通过提升领导 – 成员交换（LMX）关系质量间接地影响主管的资源回报，即 LMX 起着部分中介的作用。在以往的研究中，往往将 LMX 看作是与关系（Guanxi）相平行的概念，前者代表正式的工作关系，后者代表着非正式的私人关系，并分别对主管的工具性回报有着积极影响（Law et al., 2000；Chen & Tjosvold, 2007），而很少有研究考察非正式的私人关系对正式的工作关系质量的影响。在国外的文献中，这一问题也许不太重要，因为在西方的组织中，二者之间往往是"泾渭分明"的。但在中国文化背景下，二者之间却可能有着剪不断的"千丝万缕"的关系。本研究发现，下属对主管的私人关系投入对 LMX 具有显著的正向作用，并且关系投入经由 LMX 的中介而获取主管的工具性回报、情感性回报的假设也得到证实。这一结果说明，

主管与下属之间在私底下建立与发展起来的、具有强烈"组织规定外"及"私人情感"色彩的关系会影响和渗透到正式的工作场所中，从而在组织制度的范畴内发挥作用（刘军等，2008）。正所谓："功夫尽在诗外"。根据社会资本 (social capital) 理论，下属在工作之余，对其主管的关系投入行为具有社会资本投资的性质，等同于个人社会资本的积累与运作。在本研究的结果也表明，在中国文化背景下，主管－下属之间的私人关系对工作范围内的领导－成员交换关系有着重要的影响，那些与领导搞好私人关系的下属更可能被领导视为工作领域中的"圈内人"，反之则有可能成为"圈外人"。这一研究结果对于 LMX 的研究也具有启示意义。

4.3 关系运作的制度强化

最后，本研究通过跨层次的研究设计，搜集不同层面（个体与群体）的变量数据，运用多层线性模型（HLM）的统计方法，检验了群体的关系导向人力资源管理对下属与主管的关系交换行为的调节作用。但结果发现关系导向人力资源管理的调节效应并不显著，本研究的假设没有得到验证。究其原因，可能有以下方面：其一，关系导向人力资源管理构念反映的是群体或组织中的人力资源管理决策依"私人关系"而论的程度，但由其测量项目可知，其中的各项人力资源管理决策更多地反映工具性资源分配，比如工作任务安排、奖金分配、薪水、晋升、考核、培训机会等，没有涉及情感性资源分配。换言之，关系导向人力资源管理的概念内涵存在局限性。其二，本研究定位于群体层次，但所属的组织样本仅 8 家企业，由于群体内又进行主管－下属的二元对偶设计，难度较大，最后仅筛选出 54 个有效群体，即群体与组织样本偏少，变异不大，今后的研究可适当增加组织或群体样本量。其三，由于中国企业组织中的各项决策权力更多地集中于中高层 (Wang & Heller, 1993)，而研究设计的难度，本研究的群体样本大部分为中层，也包含部分基层，部门主管能对其拥有的组织资源（主要是工具性资源）进行有效分配的职权有限，导致关系导向人力资源管理的调节效应在中

层与基层工作群体中难以展现，因此本研究的启示，今后有必要对更多国有企业和民营企业的高层群体进行探究。不过，本研究通过相关分析发现，关系导向人力资源管理与主管情感性资源回报 (r=–0.24, $p < 0.01$)、LMX(r=–0.18, $p < 0.01$) 均显著地负相关，这说明注重私人关系的人力资源决策可能会抑制上下级之间的情感性交换，并削弱上下级之间正式的工作关系质量，不过其原因和机制也还需进一步地探究。

4.4 贡献与局限

从理论贡献上来看，本研究基于社会交换的理论视角，以"下属关系投入－主管资源回报"的概念架构来展现组织中主管与下属的关系互动实质，将组织中关系的研究引向关系运作的层面，并在模型之中同时纳入工具性资源回报与情感性资源回报，拓展了主管－下属关系的研究空间，探究了主管－下属关系的运作效能与机制，验证了 LMX 的中介作用，对组织管理实践具有深刻的启示意义。

由于本研究属于横向研究设计，因此难以确证下属关系投入与主管资源回报之间的因果关系，并且在跨层次的研究中，也难于搜集到较大的群体样本。此外，本研究虽然采取主管－下属的对偶设计，试图克服共同方法变异的影响，但可能仍然无法有效解决测量的敏感性和社会称许性问题。未来还需要进一步探究主管与下属的不同方面的关系互动与交换及其效果差异，并考量对其他员工以及群体或组织可能产生的影响，包括积极的、消极的影响。

5　结论

本研究表明，尽管下属对主管的私人关系投入行为发生在组织规定的主管与下属工作交往范畴之外，但却能发展出一种带有特殊性、私人情谊的关系连带，不仅为下属获取主管的各种正式的和非正式的资源回报（工具性资源回报、情感性资源回报）带来积极影响，还能在组织制度的范畴之内发挥作用，即通过促进和提升作为正式的工作关系的领导－

成员交换关系质量（LMX），来间接地获取主管的各类工具性与情感性的资源回报。最后，本研究也发现工作群体的人力资源管理决策特征（如关系导向人力资源管理）在一定程度也能制约主管与下属之间的关系互动与关系质量，但其影响机制与效果还需进一步地研究。

参考文献

Blau, P. M. (1964). *Exchange and power in social life*. Wiley, New York.

Chan, D.(1998). Functional relations among constructs in the same content domain at different levels of analysis: a typology of composition models. *Journal of Applied Psychology, 83*, 234–246.

Chen, C., Chen, Y., & Xin, K.(2004). Guanxi practices and trust in management: a procedural justice perspective. *Organization Science, 15*, 200–209.

Chen, N. Y.–f., & Tjosvold, D.(2007). Guanxi and leader member relationships between American managers and Chinese employees: open–minded dialogue as mediator. *Asia Pacific Journal of Management, 24*, 171–189.

Chen, Y., Friedman, R., Yu, E., & Sun, F.(2008, August). *Examining the positive and negative effects of Guanxi: a multi–level analysis of Guanxi and procedural justice*. Paper presented at Meeting of 3th International Association for Chinese Management Research, Guangzhou.

Chou, L. F.(2002). Guanxi and social network in Chinese organization, *Indigenous Psychological Research in Chinese Societies, 18*, 175–228.

[周丽芳 (2002). 华人组织中的关系与社会网络 . *本土心理学研究 , 18,* 175－228.]

Deluga, R. J.(1994). Supervisor trust building, leader– member exchange and organizational citizenship behavior. *Journal of occupational and Organizational Psychology, 67*, 315－327.

Dixon, M. A., & Cunningham, G. B.(2006). Data aggregation in multilevel analysis: a review of conceptual and statistical issues. *Measurement in Physical Education and Exercise Science. 10*, 85－107.

Farh, J. L., Tsui, A. S., Xin, K., & Cheng, B. S.(1998). The influence of relational demography and guanxi：the Chinese case. *Organization Science,9*, :471－487.

Foa, E. B., & Foa, U. G. (1980). Resource theory of social exchange. In K. J. Gergen, M. S. R. Greenberg, & H. Willis(Eds.), *Social exchange: advances in theory and research*(pp.99－131). Plenum, New York.

Friedman, R., Chi, S., & Liu, L. A. (2006). An expectancy model of Chinese–American differences in conflict avoiding. *Journal of International Business Studies, 37*, 76－91.

Graen, G. B., & Uhl–Bien, M.(1995). Relationship–based approach to leadership: Development of leader–member exchange(LMX) theory of leadership over 25 years: applying a multi–level multi–domain perspective. *Leadership Quarterly, 6*, 219－247.

Hsu, W., & Wang, An.(2007). *Downsides of guanxi practices in Chinese organizations*. Paper presented at 68th Academy of Management meeting, Philadelphia.

Hui, C., & Graen, G.(1997). Guanxi and professional leadership in contemporary Sino–American joint ventures in mainland China. *Leadership Quarterly, 8*, 451－465.

Hwang, K. K.(1987). Face and favor: the Chinese power game. *American Journal of Sociology, 92*, 944–974.

James, L. R.(1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology, 67*, 219–229.

King, Y. C. (1985). The individual and group in Confucianism: a relational perspective. In D. J. Munro(Ed.), *Individualism and holism: studies in Confucian and Taoist values.* Ann Arbor, MI: Center for Chinese Studies, University of Michigan.

Klein, K. J., Bliese, P. D., Kozlowski, S. W. J., Dansereau, F., Gavin, M. B., Griffin, M. A., et al.(2000). Multilevel analytical techniques. In K. J. Klein, & W. J. Kozlowski(Eds.), *Multilevel theory, research, and methods in organizations: foundations, extensions, and new directions*(pp. 512－553). San Francisco: Jossey-Bass.

Law, K. S., Wong, C. S., Wang, D. X., & Wang, L. H.(2000). Effect of supervisor-subordinate Guanxi on supervisory decisions in China: An empirical investigation. *International Journal of Human Resource Management, 11*, 751－765.

Liden, R. C., & Maslyn, J. M.(1998). Multidimensionality of leader-member exchange: an empirical assessment through scale development. *Journal of Management, 24*, 43－72.

Lin, N.(2001). Social capital: a theory of social structure and action. Cambridge: Cambridge University Press.

Liu, J., Song, J. W., & Wu, L. Z.(2008). Antecedents of employee career development: an examination of politics and Guanxi. *Acta Psychologica Sinica, 40*, 201－209.

[刘军 , 宋继文 , 吴隆增 (2008). 政治与关系视角的员工职业发展影响因素探讨 . *心理学报，40,* 201－209.]

Lovett, S., Simmons, L. C., & Kali, R.(1999). Guanxi versus the market: ethics and efficiency. *Journal of International Business Studies, 30*, 231－248.

Millington, A., Eberhardt, M., & Wilkinson, B.(2005). Gift giving, Guanxi and illicit payments in buyer-supplier relationships in China: analyzing the experience of UK companies. *Journal of Business Ethics, 57*, 255－268.

Schriesheim, C., Castro, S., & Coglister, C.(1998). Leader-member exchange(LMX) research : a comprehensive review of theory, measurement, and data-analytic praticices. *Leadership Quartly, 10*, 63－113.

Tsui, A. S., & Farh, J. L.(1997). Where Guanxi matters：relational demography and Guanxi in the Chinese

context . *Work and Occupations, 24*, 56－79.

Tsui, A. S., Farh, J. L., & Xin, K.(2000). Guanxi in the Chinese context. In J. T. Li, A. S. Tsui, & E.Weldon(Eds.), *Management and organizations in the Chinese context.* London: MacMillan.

Wang, Z. M., & Heller, F. A.(1993). Patterns of power distribution in managerial decision making in Chinese and British industrial organizations. *International Journal of Human Resource Management, 4*, 113－128.

Wang, Z. J., & Long, L. R.(2009). Mechanism of social capital on Chinese employees'career success. *Management Review, 21*, 30－39.

[王忠军 , 龙立荣 (2009). 员工的职业成功 : 社会资本的影响机制与解释效力 . *管理评论 , 21,* 30－39.]

Wakabayashi, M.(1988). Japanese management progress: mobility into middle management. *Journal of Applied Psychology, 73*, 217－227.

Warner, M.(1993). Human resource management "with Chinese characteristics". *International Journal of Human Resource Management, 4*, 45－65.

Williams, L. J., & Anderson, S. E.(1991). Job satisfaction and organizational commitment as predictors of organizational citizenship and in-role behaviors. *Journal of Management, 17*, 601－618.

Wong, C., Tinsley, C., Law, K., & Mobley, W. H.(2003). Development and validation of a multidimensional measure of Guanxi. *Journal of Psychology Chinese Societies, 4*, 43－69.

Xin, K. R., & Pearce, J. L.(1996). Guanxi: connections as substitutes for structural support. *Academy of Management Journal, 36*, 1641－1658.

Xin, K. R., Farh, J. L., Cheng, B. S., & Tsui, A. S.(1998). *Guanxi in vertical dyads: evidence from Taiwan and the PRC*. Paper presented at the research conference Management and Organization in the Chinese Context. Hong Kong University of Science and Technology,

Hong Kong.

Yang, M. M.(2002). The resilience of Guanxi and its new deployments: a critique of some new Guanxi scholarship. *The China Quarterly, 170*, 459‐476.

Yang, K. S.(1986). Chinese personality and its changes. In M. H. Bond(Ed.), *The Psychology of the Chinese People. Hong Kong*: Oxford University Press.

Zhang, Z. X., & Yang, C. F.(1998). Beyond distributive justice: the reasonableness norm in Chinese reward allocation. *Asian Journal of Social Psychology, 1*, 253‐269.

Zhou, J., & Martocchio, J. J.(2001). Chinese and American managers' compensation award decisions: a comparative policy‐capturing study. *Personnel Psychology, 54*, 115‐145.

Operation mechanism and effects of supervisor-subordinate guanxi in Chinese organizations

WANG Zhongjun[1]; LONG Lirong[2]; LIU Lidan[3]

([1] School of Psychology, Central China Normal University, Hubei Province Key Lab for Human Development and Mental Health, Wuhan 430079, China)

([2] School of Management, Huazhong University of Science and Technology, Wuhan 430074, China)

([3] Department of Social Science, Hubei University of Chinese Medicine, Wuhan 430065, China)

Abstract　Different from western society, Chinese employees attached much importance to developing good personal relationship with their leaders. So, the concept of "guanxi" and "guanxi management" were the most important aspects in Chinese management. In perspective of social exchange theory, this study enriched the concept of supervisor–subordinate guanxi, developed a social exchange model of supervisor–subordinate guanxi, and then investigated the mechanism of supervisor–subordinate guanxi operation and its effects on subordinate in Chinese organizations. By using questionnaire survey, the exploratory factor analysis (EFA) for data of 211 employees and confirmatory factor analysis (CFA) for data of 426 employees were implemented. The results showed that subordinate's guanxi input had only one dimension, and supervisor's resources output had two dimensions, including instrumental resources output and affective resources output. The study also showed that the subordinate guanxi input and supervisor resources output questionnaires had good reliability and high validity. By using questionnaire, Data was from a total of 426 matched supervisor–subordinate dyads in 54 work groups from different organizations. Hierarchical liner modeling (HLM) analysis was implemented. the results showed that after controlling task performance, subordinate's guanxi input had a positive effect on supervisor's instrumental resources output and affective resources output. Subordinate's guanxi input had a positive effect on leader–member exchange (LMX). The results also indicated that LMX partially mediated the relationship between subordinate's guanxi input and supervisor's resources output. Although our hypothesis that guanxi–based human resources management practice in work group had a positive moderating effect on the relationship between subordinate's guanxi input and supervisor's resources output was not tested, the results indicated that guanxi–based human resources management practices of work group were significantly relative to supervisor's affective resources output and LMX. The present study contributes to our understanding of the private guanxi operation behavior happened outside of work and its mechanism involved between supervisor and subordinate, as well as LMX in Chinese organizations. The results of this study will be of benefit to the guanxi management practices in organizations. Finally, the limitations in this study were discussed, and the future directions were also presented.

Keywords　supervisor–subordinate guanxi; guanxi input; resources output; leader–member exchange (LMX); guanxi–based human resources management practice

青少年健康相关危险行为的类型及与心理健康的关系:
基于潜在剖面分析的结果

谢　员 [1,2]，江光荣 [1,2]，邱礼林 [1]，江　丽 [1]，朱　旭 [1,2]，张　微 [1,2]

（[1] 青少年网络心理与行为教育部重点实验室，武汉 430079）

（[2] 湖北省人的发展与心理健康重点实验室，华中师范大学心理学院，武汉 430079）

摘　要　目的：探讨青少年健康相关危险行为的类型，在性别上的差异，及其与心理健康水平的关系。方法：采用潜在剖面分析方法对 2734 名中学生的健康危险行为进行分析。结果：①青少年的健康危险行为分为三个潜在类别：高危组、边缘组和低危组，比例分别为：2.9%、16.4%、80.7%。②与低危组相比，女生更不容易进入高危组，且达到统计显著（OR=–0.95, $p<0.01$）；与低危组相比，女生更不容易进入边缘组，且达到统计显著 (OR=–0.56, $p<0.01$)。③除了人际和谐维度，生活幸福、乐于学习、情绪稳定、考试镇定和心理健康总分在低危组上均显著高于边缘组和高危组；同时，边缘组得分也显著高于高危组。结论：青少年健康危险行为分为三类；高危组、边缘组和低危组，且在性别和心理健康水平上表现出差异。

关键词　青少年；健康危险行为；心理健康；潜在剖面分析

　　青少年健康危险行为 (adolescent health risk behavior) 是指那些对青少年的健康、幸福甚至生命造成直接或间接危害的行为。国内外对之的分类略有不同，美国青少年危险行为监测系统（YRBSS）将青少年健康危险行为细分为六类：①导致非故意伤害和暴力的行为；②吸烟行为；③饮酒和其他物质滥用行为；④导致非意愿妊娠和性传播疾病的性行为，包括 HIV 感染；⑤不健康的饮食行为；⑥缺乏体育锻炼／体力活动行为。我国学者则将之分为自杀自残、不健康饮食（缺乏体力活动／健康妥协）、攻击／暴力、破坏纪律、吸烟／饮酒、无保护性行为等六类。

　　目前对健康危险行为的研究中，部分对该行为进行了流行性和描述性研究，例如：孙江平等（2001）调查了中国 5 省市中学生对艾滋病及性行为的知识态度行为、易导致意外伤害的危险行为以及吸烟饮酒和成瘾类药物滥用状况；另一些学者探索了健康危险行为的影响因素，如 Wang 等发现自我效能感、自尊和情绪管理对健康危险行为具有影响；还有学者对某一类健康危险行为进行了分类，如黎亚军等采用潜在剖面分析（latent profile analysis）对中国儿童青少年身体攻击进行分类,发现三种类型: 高发组、边缘组和低发组。

　　心理健康状况和健康危险行为有一定的关联性。有研究表明青少年的抑郁与焦虑是多种危害健康行为的危险因素；心理健康与健康危险行为有一定的聚集性，心理健康问题组的健康危险行为发生率均

高于心理健康状况正常组。

本研究主要是基于以下考虑：以往很少有研究去进一步分析青少年健康危险行为类型下的亚群体，因此本研究将青少年健康危险行为作为一个总变量（包括五个维度），通过潜在剖面分析的方法探讨由五个维度组成的健康危险行为的潜在类别，给出分类的相对标准，探讨不同类别在性别上是否存在差异，并对其不同类型上的个体心理健康水平差异进行探讨。

本研究使用的潜在剖面分析法广泛应用于社会学、心理学、管理学等领域的类型划分。潜在剖面分析能够做出所有个体属于某一青少年健康危险行为类型群体的可能性估计，借助拟合信息等统计指标帮助确定最合理的分类，从而保证分类结果的准确性。在潜在剖面分析中常用的拟合信息有艾凯克信息准则 (AIC) 和贝叶斯信息准则 (BIC)、罗梦戴尔鲁本校似然比 (LMRT)、平均信息量 (Entropy) 等。一般说来一个模型如果有更高的 Entropy、更低的 AIC 和 BIC，LMRT 达到显著性，则说明这个模型的拟合程度高。

1 对象与方法

1.1 研究对象

研究数据取自国家科技部支撑计划课题《我国青少年危险行为的评估、预警与干预示范研究》中一级干预的数据库。该数据库采取分层取样的策略，从湖北省武汉市 4 所学校的初一、初二、高一和高二的 4 个年级中整班取样，每个年级随机抽取 4 个班，共 64 个班级，最后收集 2846 份数据。删除在研究变量上有缺失值的数据，最后有效样本包括 2734 个中学生，平均年龄 14.8 岁，最小 11 岁；最大 20 岁。其中男生 1411 名，占比 51.6%；女生 1323 名，占比48.4%。

1.2 工具

1.2.1 青少年危险行为倾向性问卷

王孟成等的《青少年健康相关行为调查问卷》共有 6 个分量表，分别测量了 6 种健康危险行为：暴力 / 攻击、违纪 / 违法、自杀 / 自伤、吸烟 / 饮酒、不健康饮食 / 缺乏体力活动、无保护性行为，除不健康饮食 / 缺乏体力活动的 Cronbach α 系数为 0.40 外，其他分量表的 Cronbach α 系数均在 0.75 以上。全量表及各分量表间隔 4 周和 8 周的重测系数分别在 0.55–0.80 和 0.53–0.76 之间。本次研究对除无保护性行为外的其它 5 种健康危险行为进行了测量，共 43 个项目，被试自评。保留原量表的题干，将选项描述改为出现该行为的可能性，5 点计分，分别为 1= 绝不可能；2= 不太可能；3= 难以确定；4= 有些可能；5= 非常可能，部分项目反向计分，得分越高，说明青少年健康危险行为倾向性越高。

1.2.2 中学生心理健康量表

采用苏丹、黄希庭的《中学生心理健康量表》。该量表采用自评，5 点计分， 25 个项目，包括生活幸福、乐于学习、人际和谐、考试镇静、情绪稳定 5 个维度。量表具有良好的信效度，总量表的内部一致性系数为 0.83，分半信度系数为 0.85，各维度的内部一致性系数和分半信度系数均在 0.56 以上；量表总分及各维度与 SCL–90 总分及各因子分均呈负相关。

1.3 数据分析

使用 Mplus7.0 软件分析，首先分析青少年健康危险行为的潜在类别数，在此基础上探讨其在性别和心理健康上的差异。

2 结果

2.1 青少年健康危险行为的潜在剖面分析

以健康危险行为的五个维度为指标，将健康危险行为的类型依次分成 1 类、2 类、3 类、4 类进行潜在剖面分析的模型拟合性估计。

表 1 的模型适配估计结果表明，当类别数目为 4 类时，虽然 AIC、BIC 和 Entropy 指标略小于 3 类的 AIC、BIC 和 Entropy，但是 LMRT 没有达到显著性水平（LMRT=875.80，$p>0.05$），而该指标在 3 类上达到了统计性显著，且 3 类和 4 类在另外的三个指标上相差很小，说明模型分为三类时拟合度最好。

三种类型中，五个维度在第一类健康危险行为

表 1 不同健康危险行为类型的潜在剖面分析拟合信息

剖面数	AIC	BIC	Entropy	LMRT
1 类	19947.91	20007.04		
2 类	15186.53	15281.14	0.96	4674.92**
3 类	13598.17	13728.26	0.95	1567.35*
4 类	12715.92	12881.50	0.96	875.80

注：*P<0.05;**P<0.01,下同。AIC: 艾凯克信息准则; BIC: 贝叶斯信息准则; Entropy: 平均信息量; LMRT: 罗梦戴尔鲁本校似然比。

表 2 心理健康各维度在不同健康危险行为类型上的差异

参数（BETA）	心理健康（总）	生活幸福	乐于学习	情绪稳定	人际和谐
低危组与高危组相比 (BETA1)	0.71**	1.24**	1.09**	0.99**	−0.06
低危组与边缘组相比 (BETA2)	0.23**	0.81**	0.72**	0.60**	0.08*
高危组与边缘组相比 (BETA3)	−0.48**	−0.43**	−0.38**	−0.39**	0.15

注：model constraint: 用来做参数间统计检验；beta1=m1−m2; beta1 是第一类和第二类被试心理健康均值的差值；beta2=m1−m3; beta2 是第一类和第三类被试心理健康均值的差值；beta3=m2−m3; beta3 是第二类和第三类被试心理健康均值的差值。

上的均值最小，说明健康危险行为最低，命名为低危组，它占整个群体的比例最高，为 80.7%；第二类健康危险行为五个维度上的均值最大，其中均值最大的为吸烟喝酒维度，说明健康危险行为表现最高，命名为高危组，它占的比例最低，为 2.9%；第三类健康危险行为中，五个维度的均值处于中间位置，命名为边缘组，比例为 16.4%。从三类健康危险行为比例来看，青少年大多处在低危组（80.7%），高危组人数最少（2.9%）。

2.2 健康危险行为类型在性别上的差异

以性别为预测变量 (男 =1，女 =2)，以 3 类型健康危险行为的类型（高危组、边缘组、低危组）为因变量，用软件 MPLUS7.0 进行 logistic 回归分析来探讨不同类型在性别上的差异。结果表明，与低危组相比，女生更不容易进入高危组，且达到统计显著（OR=−0.95, p<0.01）；与低危组相比，女生更不容易进入边缘组，且达到统计显著 (OR=−0.56, p<0.01)，说明女生的健康危险行为发生率要低于男生健康危险行为发生率。

2.3 不同健康危险行为类型的心理健康水平比较

利用 MPLUS7.0 软件中的模型限制（model constraint）法来分析心理健康水平在不同健康危险行为类型上的差异。结果表明：低危组的心理健康总分（M=3.59）显著高于边缘组 (M=3.10) 和高危组 (M=2.88)，边缘组的心理健康总分显著高于高危组；

低危组的生活幸福得分（M=3.91）显著高于高危组 (M=2.68) 和边缘组 (M=3.10)，边缘组的生活幸福得分显著高于高危组；低危组的乐于学习得分 (M=3.54) 显著高于高危组 (M=2.44) 和边缘组 (M=2.82)，边缘组的乐于学习得分显著高于高危组；低危组的情绪稳定得分 (M=3.66) 显著高于高危组 (M=2.67) 和边缘组 (M=3.06)，边缘组的情绪稳定得分显著高于高危组；但是人际和谐维度在三个健康危险行为类型上的水平和上述结果有所不同，除了低危组的人际和谐得分 (M=3.37) 显著高于边缘组 (M=3.29) 外，其他两组差异不显著，而且高危组的人际和谐得分最高（M=3.43），如表 2 所示。

3 讨论

根据个体在暴力 / 攻击、违纪 / 违法、自杀 / 自伤、吸烟 / 饮酒、不健康饮食 / 缺乏体力活动五个维度上的表现，利用潜在剖面分析方法对青少年健康危险行为进行分类，结果表明，青少年健康危险行为可以分为高危组、边缘组、低危组三类。青少年大多处在低危组，较少处在高危组，这符合实际观察。在实践应用中，高危组群体是心理健康重点干预的对象，边缘组是重点教育和引导的对象。此外，青少年在五个维度上的得分并不在同一类别上，个体的吸烟酗酒得分也许处于低危组，但是他的饮食健康维

度得分也许就处于高危组，而违纪违法维度得分处于边缘组；所以个体的健康危险行为是一个剖面体。在教育实践中应该重视对个体健康危险行为中的高危维度的评估与干预。

研究结果还表明，三类健康危险行为类型在性别上存在差异。与低危组相比，女生更不容易进入高危组和边缘组，且达到统计显著。说明青少年群体中女生的健康危险行为发生概率要低于男生，这与以往的研究结果一致。在健康危险行为的预防和干预中，应根据性别特点，针对性地开展预防和干预措施。

而对心理健康水平在健康危险行为类型上的差异比较发现：除了人际和谐，在生活幸福、乐于学习、情绪稳定、考试镇定和心理健康总分上，低危组均显著高于边缘组和高危组，边缘组显著高于高危组。说明健康危险行为越低，心理健康水平越高。在学校教学工作中，应该加大中学生心理健康教育课的开展，提高其心理素质，从而预防和降低青少年健康危险行为的发生。人际关系是否和谐受到个体人格、自尊水平、父母教养方式、价值观等因素的综合影响；人际关系受阻，主要是由个体的自卑、自负、猜疑、胆小、害羞、嫉妒等主观体验造成的。此外，在高危组中最突出的健康危险行为是吸烟喝酒，这种行为往往受到同伴关系的影响，同伴的吸烟人数与青少年的吸烟行为之间存在紧密的联系；这也许导致了人际和谐维度在三个健康危险行为类型上看似矛盾的关系。

参考文献

黄希庭，苏丹 . (2007). 中学生适应取向的心理健康结构初探 . *心理科学*, *6*(30), 1290–1294.

季成叶 . (2007). 青少年健康危险行为 . *中国学校卫生*, *4*, 289–291.

李金，王宏，许红，等 . (2010). 重庆库区农村中学生心理健康与健危险行为关系研究 . *中国全科医学*, *5*, 2859–2861.

黎亚军，卢富荣，骆方，王耘 . (2012). 中国儿童青少年身体攻击的发生状况—基于潜在类别分析的结果 . *中国临床心理学杂志*, *20*(4), 559–561

孙江平，马迎华，陈虹 . (2001). 中国 5 省市中学生危险行为现状 (一). *中国学校卫生*, *22*(3), 202–204.

陶芳标，张金霞，毛琛 . (2004). 抑郁，焦虑症状与中学生多种危害健康行为 . *中国学校卫生*, *25*(2), 131–133.

王孟成，蚁金瑶，蔡琳，等 . (2012). 青少年健康相关危险行为问卷的编制及信效度检验 . *中国心理卫生杂志*, *4*, 287–292.

王瑶 . (2004). 小组辅导提高大学生人际交往水平的实验研究 . *心理科学*, *5*(27), 983–985.

赵崇莲，郑涌 . (2009). 大学生人际关系质量的影响因素研 . *心理科学*, *4*(32) , 1254–1256.

张洁婷，焦璨，张敏强 . (2010). 潜在类别分析技术在心理学研究中的应用 . *心理科学进展*, *18*(12) , 1991–1998.

张琳，星一，季成叶 . (2005). 中国北方两城市青少年健康危险行为聚集模式分析 . *中国行为医学科学*, *14*(8), 741–743.

郑宇，方晓义 . (2000). 同伴与青少年吸烟行为关系的研究进展 . *心理科学*, *23*(6), 746–747

Brener, N. C. (1998). Co-occurrence of health-risk behaviors among adolescents in the united states. *Journal of Adolescent Health, 22*, 209–213.

Centers for Disease Control and Prevention. Youth risk behavior surveillance united states, 2011. Morbidity and Mortality Weekly Report, Surveillance Summaries, 2012, 61(4), 1–45.

Farhat, T., Iannotti, & R. J. (2010). Simons-Morton BG. Overweight, obesity, youth, and health-risk behaviors. *American Journal of Preventive medicine, 38*(3), 258–267.

Hallfors, D. D., Waller, M. W., Ford, C. A, et al. (2004). Adolescent depression and suicide risk: Association with sex and drug behavior. *American journal of preventive medicine*, 224–231.

Jessor R. (1998). New perspectives on adolescent risk behavior.Cambridge University Press.

Moreno, M. A., Parks, M. R., Zimmerman, F. J., et al. (2009). Display of health risk behaviors on my space by adolescents: Prevalence and associations. *Archives of Pediatrics and Adolescent Medicine, 163*(1), 27–34.

Muth é n, L. K., Muth é n, B., O. (2012). Mplus: Statistical analysis with latent variables; user's guide; [Version 7]. Muth é n and Muth é n, 159–257.

Rabaglietti, E., Kliewer, W., & Ciairano, S. (2009). Adolescent risk behavior in Italy and the netherlands. *European Psychologist, 14*(3), 180–192.

Walsh, J. L., Senn, T. E., & Carey, M.P. (2012). Exposure to different types of violence and subsequent sexual risk behavior among female sexually transmitted disease clinic patients: A latent class analysis. *Psychology of Violence ,2*(4), 339–354.

Wang, M., & Hanges, P. J. (2010). Latent class procedures: Applications to organizational research. *Organizational Research Methods, 14*(1), 24–31.

Wang, R. H., Hsu, H. Y., Lin, S. Y., et al. (2010). Risk behaviours among early adolescents: risk and protective factors. *J Adv Nurs, 66*(2), 313–323.

Wolff, U. (2010). Subgrouping of readers based on performance measures: A latent profile analysis. *Reading and Writing, 23*(2), 209–238.

Adolescent health risky behavior: Types and the relationship with mental health—based on latent profile analysis

XIE Yuan[1,2] , JIANG Guangrong[1,2] , QIU Lilin[1] , JIANG Li[1] , ZHU Xu[1,2] , ZHANG Wei[1,2]

(Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, ,Wuhan 430079, China)

Abstract　Objective: To investigate types of adolescent health risk behaviors and the relationship with mental health, the gender difference in frequency of them as well. Methods: Latent profile analysis (LPA) was used to explore types of health risk behaviors of 2734 middle school and high school students. Results: ① Adolescent health risk behaviors could be divided into three types, namely, high−risk group, marginal−risk group and low−risk group, and the prevalence were 2.9%, 16.4%, 80.7%, respectively. ② In comparison to male students, female students are 0.89 times less likely to be in high−risk group rather than in low−risk group，and 0.56 times less in marginal−risk group rather than in the low−risk group; both showing significant differences(P<0.01). ③ The scores on the total scale of mental health and the subscales (including life well−being, learning satisfaction, exam calmness and emotional stability, except for interpersonal harmony) are significantly higher in low−risk group than that in other groups. Furthermore, the results of marginal−risk group are significantly higher than high−risk group. Conclusion: Adolescent health risk behaviors could be divided into three types with a gender difference as well as the different mental health levels.

Keywords　adolescent; health risk behaviors; mental health; latent profile analysis

模糊规避的形成机制：基于可评价性假设的视角

徐富明 [1,2,*]，张军伟 [1]，刘腾飞 [1]，蒋 多 [1]，文桂婵 [1]

([1] 华中师范大学心理学院暨湖北省人的发展与心理健康重点实验室，武汉，430079)

([2] 中国科学院心理研究所，北京，100101)

摘 要 模糊规避是指在相同奖赏的情况下，决策者会力图规避从主观上判断具有模糊概率的事件而偏好具有相同精确概率的事件。本研究使用同时评价、单独评价的研究范式从随机事件和自然事件两个领域来探讨模糊规避的形成机制。研究结果表明，当风险事件和模糊事件同时评价时，个体倾向于模糊规避；当风险事件和模糊事件单独评价时，模糊规避会消失。

关键词 模糊规避；同时评价；单独评价

1 问题提出

Savage(1954) 在期望效用 (expected utility，EU) 理论的基础上提出了主观期望效用 (subjective ex-pected utility，SEU) 理论。Savage 认为个体可以通过决策者的偏好推断出模糊事件发生的主观概率，只要决策者赋予事件的主观概率相等，事件发生概率的模糊性对决策者的选择不会有影响。这个看似合理的定论遭到了 Ellsberg(1961) 的反对。他通过精巧的实验展示了有名的 Ellsberg 悖论：被试偏好罐子Ⅱ (有 100 个彩球，红球和黑球各 50 个) 中的红球而不是罐子Ⅰ (有 100 个彩球，但红球和黑球的比例未知) 中的红球，同时被试也偏好罐子Ⅱ中的黑球而不是罐子Ⅰ中的黑球。决策者会规避罐子Ⅰ中红球和黑球混合比例的不确定性，这种现象就称之为模糊规避 (Ellsberg，1961; 张军伟，徐富明，刘腾飞，陈雪玲，蒋多，2009) 。

自从模糊规避的概念提出以后，有研究者试图探讨这种现象的形成机制，理论解释主要包括：(1) 他人评价假设 (other evaluation hypothesis) 。他人评价假设认为个体之所以会模糊规避是由于个体会预期到他人 (即同伴) 将对决策结果进行评价，所以他做出的选择应该是最能自我辩解的，以免受到责备 (Curley，Yates， & Abrams，1986) 。(2) 能力假设 (competence hypothesis) 。能力假设认为模糊规避是由能力所引起的称赞 (credit) 和责备 (blame) 的不对等所导致的 (Heath & Tversky，1991) 。(3) 相比无知假设 (comparative ignorance hypothesis) 。相比无知假设认为当与不模糊的事件或高能力的人相比时，决策者会产生无能力感，因而将倾向于模糊规避 (Fox & Tversky，1995) 。

在行为决策领域，Hsee(1996) 提出了可评价性假设 (evaluability hypothesis) 来解释偏好反转 (pref-erence reversal) 现象。即当单独评价时，容易评价的属性对决策影响较大，但在同时评价时，难评价的属性提供了额外的信息，增加了难评价属性的可评价性。例如，

---

* 通讯作者：徐富明。E– mail: fumingxu@126. com。

他提出的"字典问题"：

字典 A: 1993 年出版,有 10000 页,看起来是新的。

字典 B: 1993 年出版,有 20000 页,看起来很破旧。

字典的新旧是容易评价的属性,而字典的页数是难评价的属性。所以在单独评价时,大多数被试会选择字典 A,而在同时评价时,大多数被试会选择字典 B。按照同样的道理,在同时评价时,难评价的属性(事件的模糊性)更容易评价,个体会倾向于模糊规避,而在单独评价时,事件的模糊性不容易评价,模糊规避会消失。

本研究将从随机事件和自然事件两个领域来探讨模糊规避的形成机制。在随机事件领域,本研究选择了模糊规避研究中经常使用的摸球情境。在自然事件领域,本研究选择了大家都熟悉的气温作为研究材料。我们之所以选择这两种实验情境是为了能交叉验证,提高研究的效度。

2　研究方法

2.1　实验 1: 摸球情境中模糊规避的形成机制

2.1.1　被试

被试是 253 名心理学院大一的本科生,有效回收问卷 234 份,其中男性 98 名,女性 136 名,被试的平均年龄是 19. 03(SD=2. 48) 。

2.1.2　实验设计与材料

实验采用 3 种实验版本,第一种版本是单独呈现模糊事件的实验版本,第二种版本是单独呈现风险事件的实验版本,最后一种版本是模糊事件和风险事件同时呈现的实验版本。这 3 种实验版本随机分发给 3 组被试完成。本实验采用定价(price)的研究范式,

表 1　同时评价、单独评价对模糊规避的影响

		Mean	*Standard errors*	*N*
单独评价				
	模糊事件	28. 19	2. 39	84
	风险事件	29. 31	2. 28	84
同时评价				
	模糊事件	21. 33	2. 48	66
	风险事件	33. 89	3. 07	66

即研究者要求参与者说出玩这个博彩游戏所愿意支付的最高价格 (willingness to pay，WTP) 。

2.1.3　数据分析结果

对数据进行分析表明 (如表 1) : 在单独评价时,对模糊事件和风险事件的定价不存在显著差异, t(166)=0.34， $p > 0.05$; 在同时评价时,对模糊事件和风险事件的定价存在显著差异, t(65)=4.04， $p < 0.01$。

2.2　实验 2: 现实情境中模糊规避的形成机制

2.2.1　被试

被试是 180 名心理学院的在职硕士班学员、部分本科生和研究生,有效回收问卷 168 份,其中男性 73 名,女性 95 名,被试的平均年龄是 30.29 (SD=5.41) 。

2.2.2　实验设计与材料

本实验改编了 Fox 和 Tversky (1995) 的实验情境, 每个实验情境均由"低于"与"不低于"两个子问题构成。因为本实验选用武汉的大学生作为研究被试,所以这些大学生对武汉未来 7 天的气温比较熟悉,而对哈尔滨未来 7 天的气温相对不熟悉。为了确保这一操作的准确性,我们在实验中设计了对武汉和哈尔滨的气温熟悉度的评价。统计结果表明 : 在单独评价时,武汉与哈尔滨的气温熟悉度差异显著, $M_{武汉}$=4.83(SD=0.85)， $M_{哈尔滨}$=2.45(SD=1.02)， t(82)=11.67， $p < 0.01$; 在同时评价时,武汉与哈尔滨的气温熟悉度差异显著, $M_{武汉}$=5.50(SD=0.67)， $M_{哈尔滨}$=2.40(SD=1.04)， t(83)=24.08， $p < 0.01$。这一统计结果说明我们的操作是可行的。另外,武汉和哈尔滨未来一周的最高气温是通过天气预报设计的,我们把一周中最高气温的中值作为未来一周的最高气温。

本实验的数据统计方法使用的是 Fox 和 Tversky(1995) 提出的互补赌注 (complementary bets) 范式。具体来说就是, 每个实验情境均包括"低于"与"不低于"两个子问题,所以被试对"低于"与"不低于"两个子问题的定价相加就是对每个问题的定价。如果被试对模糊事件 (哈尔滨气温问题) 的"低于"与"不低于"两个子问题的定价相加小于其对风险事件 (武汉气温问题) 的"低于"与"不低于"两个

表 2 同时评价、单独评价对模糊规避的影响

	Mean	*Standard errors*	*N*
单独评价			
武汉	50. 15	3. 56	42
哈尔滨	47. 42	3. 34	42
同时评价			
武汉	54. 26	2. 66	84
哈尔滨	36. 09	2. 77	84

子问题的定价相加，表明被试倾向于模糊规避，反之被试倾向于模糊寻求。为了方便起见，本研究把被试对低于"与"不低于"两个子问题的定价相加，取其平均数进行统计分析。

2.2.3 数据分析结果

对数据进行分析表明 (如表 2) : 在单独评价时，被试对武汉和哈尔滨不低于和低于定价的平均数不存在显著差异，$t(82)=0.561$，$p > 0.05$; 在同时评价时，被试对武汉和哈尔滨不低于和低于定价的平均数存在显著差异，$t(83)=8.340$，$p < 0.01$。

3 总的讨论

本研究从随机事件和自然事件两个领域探讨了模糊规避的形成机制，研究结果表明，模糊规避是模糊事件和风险事件相比较的结果。具体来说，当风险事件和模糊事件同时评价时，由于个体意识到模糊事件的模糊性，因而个体倾向于模糊规避；而当风险事件和模糊事件单独评价时，由于个体并未意识到模糊事件的模糊性，因而个体没有模糊规避的倾向。另外，本研究的结果也与 Hsee(1996) 的可评价性假设一致。可评价性假设是指当单独评价时，容易评价的属性对决策影响较大，但在同时评价时，难评价的属性提供了额外的信息，增加了难评价属性的可评价性。在我们的研究中，当单独评价时，个体没有意识到难评价的属性 (模糊事件的模糊性)，因而个体没有模糊规避的倾向；而当同时评价时，难评价的属性 (模糊事件的模糊性) 提供了额外的信息，增加了难评价属性的可评价性，因而个体倾向于模糊规避。新近有些研究者采用同时评价和单独评价

研究范式的研究在一定程度上支持了本研究的结论。例如，来自心理账户 (mental accounting) 的研究发现，同时评价能使被试意识到决策偏差，从而刺激了被试的理性行为，进而削弱了心理账户原则的解释力 (Chatterjee，Heath， & Min，2009)。van Putten，Zeelenberg 和 van Dijk(2008) 的研究表明，当错过的机会与另一个没有吸引力的机会同时呈现时会减弱不作为惯性 (inaction inertia)。

4 结论

本研究使用同时评价、单独评价的研究范式从随机事件和自然事件两个领域来探讨模糊规避的形成机制。研究结果表明，当风险事件和模糊事件同评价时，个体倾向于模糊规避；当风险事件和模糊事件单独评价时，模糊规避会消失。

参考文献

张军伟，徐富明，刘腾飞，陈雪玲，蒋多 . (2009) . 行为决策中模糊规避研究的回顾与展望 . *应用心理学*，*15*(3)，245-250.

Chatterjee，S.，Heath，T. B.，& Min，J. (2009) . The susceptibility of mental accounting principles to evaluation mode effects. *Journal of Behavioral Decision Making, 22*, 120-137.

Curley，S. P.，Yates，J. F.，& Abrams，R. A. (1986). Psychological sources of ambiguity avoidance. *Organizational Behavior and Human Decision Processes, 38*, 230-256.

Ellsberg，D. (1961) . Risk，ambiguity，and the Savage axioms. *Quarterly Journal of Economics,75*，643-669.

Fox，C. R.，& Tversky，A. (1995) . Ambiguity aversion and compara-tive ignorance. *Quarterly Journal of Economics,110*，585 -603.

Heath，C., & Tversky. A. (1991) . Preference and belief:

Ambiguity and competence in choice under uncertainty. *Journal of Rise and Uncertainty,4*,5–28.

Hsee,C. K.（1996）. The evaluability hypothesis: An explanation for preference reversals between joint and separate evaluations of alterna–tives. *Organizational Behavior and Human Decision Processes,67*，247–257.

Savage，L. J.（1954）. *The foundations of statistics*. New York: John Wiley. Trautmann，S. T.

Van Putten，M.，Zeelenberg，M.，& van Dijk，E.（2008）. Multiple options in the past and the present: The impact on inaction inertia.*Journal of Behauioral Decision Making,21*,519–531.

The formation mechanism of ambiguity aversion: A perspective based on the evaluability hypothesis

Xu Fuming[1,2], Zhang Junwei[1], Liu Tengfei[1], Jiang Duo[1], Wen Guichan[1]

([1] School of Psychology，Central China Normal University，and Hubei Human Development and Mental Health Key Laboratory，Wuhan，
430079)

([2] Institute of Psychology，Chinese Academy of Sciences，Beijing，100101)

Abstract　Ambiguity aversion refers to the phenomenon that people prefer events with known probabilities to similar ambiguous events in which the decision maker does not know the values of the probabilities. Since ambiguity aversion was proposed by Ellsberg in 1961， there have been three main models accounting for ambiguity aversion: the other evaluation hypothesis，the competence hypothesis and the comparative ignorance hypothesis. The other evaluation hypothesis suggests that increasing the number of people watching a decision enhances ambiguity aversion，and enhances it more than other factors that researchers manipulate. The competence hypothesis suggests that people prefer betting on their own judgment to an equiprobable chance event only when they consider themselves knowledgeable，but not otherwise. The comparative ignorance hypothesis suggests that ambiguity aversion increases with the perception that others are more competent and more knowledgeable. To investigate the influence of joint evaluation and separate evaluation on individuals＇ambiguity aversion，a 2 (ambiguous event vs. risky event) × 2 (joint evaluation vs. separate evaluation) mixed experimental design was adopted. In the 1first experiment，the subjects were required to indicate their WTP of ambiguous events and risky events. In the 2nd experiment，every piece of experimental episode consisted of "below" and "not below" conditions under which the subjects were required to indicate their WTP. The research revealed that there was no significant difference in WTP between ambiguous events and risky events when evaluated simultaneously，when they were separately evaluated，there was a significant difference in WTP between ambiguous events and risky events. The results indicate that，when risky events and ambiguous events are evaluated simultaneously，people opt for ambiguity aversion; in case of a separate evaluation，however，ambiguity aversion will disappear.

Keywords　ambiguity aversion; joint evaluation; separate evaluation

Learning process and learning outcomes of video podcasts including the instructor and PPT slides: A Chinese case

Zhongling Pi , Jianzhong Hong*

(Key Laboratory of Adolescent Cyberpsychology and Behavior（Ministry of Education）,

School of Psychology, Central China Normal University, Wuhan China)

Abstract Video podcasts have become one of the fastest developing trends in learning and teaching. The study explored the effect of the presenting mode of educational video podcasts on the learning process and learning outcomes. Prior to viewing a video podcast, the 94 Chinese undergraduates participating in the study completed a demographic questionnaire and prior knowledge test. The learning process was investigated by eye–tracking and the learning outcome by a learning test. The results revealed that the participants using the video podcast with both the instructor and PPT slides gained the best learning outcomes. It was noted that they allocated much more visual attention to the instructor than to the PPT slides. It was additionally found that the 22 min was the time at which the participants reached the peak of mental fatigue. The results of our study imply that the use of educational technology is culture bound.

Keywords video podcasts; learning process; learning outcomes; eye movements; visual attention; mental fatigue; culture

1 Introduction

With the rapid development and growing ubiquity of the Internet, video podcasts have become one of the fastest developing trends in learning and teaching. Video podcasts refer to video files that are distributed in a digital format through the Internet and accessed using personal computers or mobile devices (McGarr, 2009). They are generally used to record and transmit live lectures (Li, 2013).

Video podcasts are considered as having great promise for education, as they can present information through vivid visual and audio forms simultaneously (Homer, Plass, & Blake, 2008; Zhang, Zhou, Briggs, & Nunamaker, 2006), and the learning outcome of video podcasts has become a subject of considerable research interest (Copley, 2007；Homer et al., 2008; Zhang et al., 2006).

Answering the question whether video podcasts are effective for learning is far from straightforward. Previous studies have shown that the presenting mode of video podcasts influences the learning outcome (Chen & Wu, in press; Wang, Hao, & Lu, 2014). For example, Wang et al. (2014) compared the learning outcomes of two types of video podcasts in an investigation involving 60 Chinese undergraduates and graduates. The first type of video podcast included the instructor and PPT slides, whereas the second type of video podcast only included PPT slides. It was found that the learners viewing the first type of video podcast gained significantly higher scores. The findings of the study thus implied that the presenting mode of the video podcast affects the learning outcome and that the presence of the instructor in a video podcast enhances learning.

Some western studies, however, have produced different results (Homer et al, 2008; Kizilcec, Papadopoulos, & Sritanyaratana, 2014; Lyons, Reysen, & Pierce, 2012). For example, Homer, Plass, and Blake (2008) examined the effect on the learning outcomes of the presence of the instructor in video podcasts. Similar to Wang et al.'s study (2014), there were two different types of video podcast: the first included the instructor and PPT slides, and the second included only PPT slides. Twenty-six American undergraduates were assigned to view one of the two versions of the video podcasts. No significant differences were found between the two groups as regards the learning outcome. Similar results were obtained by Kizilcec et al. (2014).

A possible reason for the discrepancy between the studies is the fact that the participants in the reviewed studies were different. The participants in Wang et al.'s study (2014) were Chinese undergraduates and graduates, whereas in Homer et al.'s study (2008) and Kizilcec et al.'s study (2014) they were American undergraduates. The students lived in different cultures, which could

have affected the cognitive and learning process and the learning outcome. A growing number of studies have suggested that people from different cultures exhibit different cognitive processing styles (Nisbett & Miyamoto, 2005; Varnum, Grossmann, Kitayama, & Nisbett, 2010). Does this imply that culture may play an important role in video podcast learning?

Although previous studies on the topic have produced inconsistent results, the findings collectively imply that some presenting modes of video podcasts could be more appropriate and effective than others. Therefore, for researchers and educators, the key question to be investigated becomes not whether video podcasts are effective tools for learning, but rather to identify a presenting mode which optimizes the learning process and outcomes (Homer et al, 2008).

Understanding the learning process of video podcasts, in particular, viewing behaviors and mental fatigue of learners, is crucial to improving learning outcomes. Recently, Kizilcec et al. (2014) used an eye-tracking technique to explore viewing behavior while watching video podcasts. In the study, video overlays of the instructor were used such that the slides were presented alternately with and without the instructor overlay, with the form alternating approximately every 2 min. The study found that when the instructor was present, learners spent 41% of their time looking at the instructor and switched between the instructor and slides every 3.7 s.

Previous studies have suggested that mental fatigue seriously affects learning outcomes (Gonzalez, Best, Healy, Kole, & Bourne, 2011). Controlling the length of video podcasts thus becomes an important element in avoiding mental fatigue and thus ensuring a good learning outcome. Danforth, Schumacher, Cullen, and Ma (2012) explored the effect of video podcast length on nutrition knowledge learning. They presented nutrition knowledge through three formats. The first format focused on the ingredients of

one dish, including nutrition and preparation information. The second was a quick recipe demonstration. The third was a long recipe demonstration that included all the steps involved in completion of the finished recipe. The three formats were included in video podcasts ranging from 2 to 11.5 min in length. After viewing the video podcasts, participants filled out a learning test. It was found that the participants viewing a 4–6 min video podcast gained the best learning scores. The results of the study thus implied that the learners experienced mental fatigue after 4–6 min of watching a learning video podcast.

Previous studies have compared the learning outcomes of different presentation modes, for example (Wang et al., 2014), but few studies have explored the viewing behavior and mental fatigue associated with video podcasts based on eye movements and within the specific context of China. Based on the above discussion, it becomes clear that we know relatively little about which presenting mode of video podcast is the most effective for learning, how learners process the learning content, how cultural context affects video podcast learning and when learners experience mental fatigue.

Our study focused on investigating which presenting mode of video podcast optimally enhances learning by comparing the learning outcomes of four frequently used modes of video podcast. The learning process of the video podcast was examined using eye–tracking, which is a promising tool for tracking the learning process because eye movements can be considered as reflecting in real time the attention and encoding processes and mental fatigue of learners during the entire learning period (Hyona, 2010; Kizilcec et al., 2014). Specifically, fixation data can be considered as an indicator representing learners' visual attention, saccade eye movement as reflecting visual attention transition (Kizilcec et al., 2014), and blink data, a classical indicator, can be used to assess learners' mental fatigue (Zhu, Wu, Wang, & Qi, 2008).

Based on the above literature review and analysis, the following hypotheses were formulated:

(1)Video podcasts including both the instructor and PPT slides best facilitate learning outcomes.

(2)Chinese learners using video podcasts including the instructor and PPT slides spend more than half the learning time visually fixated on the instructor.

(3)Learners' mental fatigue increases with time and reaches a peak at a certain point.

2 Method

2.1 Participants

Ninety–six undergraduate students (23 male and 73 female) were recruited from Central China Normal University (CCNU) via advertisement. Two of them were excluded from the study because of poor eye calibration. The study participants were aged 17.17 to 25.25 $years$ (M=20.25, SD=1.32). Using informal interviews, it was ensured that none of the participants was tired or knowledgeable on the topic of the video podcasts. The participants were from 10 study majors and no participant was a psychology major. All participants completed a prior knowledge test to assess their prior knowledge. The information gained was used to balance the groups and avoid the possibility of differences in prior knowledge introducing bias into the results. Participants all had normal or corrected–to–normal vision. They all gave written informed consent. On completion of the study, every participant received a small gift; they could choose a small item such as a delicate spoon and a pair of chopsticks, cleansing tissue, washing powder and candy.

2.2 Apparatus

Eye movements were registered by Eyelink 1000 eye tracker (SR Research Ltd., Canada). Prior to viewing the video podcasts, a 9–point calibration and validation procedure was applied to determine periods of fixation

and eye reaction to stimuli presentation. The stimuli were presented on a 21-in. CRT monitor (NESOJXC FS210A) at a viewing distance of 60 cm. The resolution of the monitor was 1024 × 768 pixels and the refresh rate was 75 Hz. The participants listened to the auditory explanation via Philips headphones connected to the PC. Participants used a chin rest with a head-stop to stabilize head position during the experiment. Data were monitored for a 1000 Hz monocular sampling rate.

2.3 Materials

2.3.1 Stimuli

Four modes of 25 min video podcasts were presented on the topic of attachment by Dr Huang from the School of Psychology (CCNU). The modes used were (1) *the mode of PPT slides (PPT)*, in which only synchronized PPT slides were included; (2) *the mode of instructor without PPT slides (Instructor)*, in which only the instructor was included; (3) *the mode of instructor with PPT slides (Instructor and PPT)*, in which the instructor and synchronized PPT slides were included; and (4) *the mode of classroom (Classroom)*, in which the whole learning activity in the classroom was recorded, including the instructor, students and synchronized PPT slides.

The instructor's audio explanations were the same in all four modes. The difference between the four modes was the visual material. The PPT slides were simply to reinforce the learning points. Before the video podcasts were recorded, the instructor was asked to write a script and keep the learning content identical.

2.3.1 Measures

Demographic questionnaire. Participants were asked to report their gender, age, grade, QQ number (for the QQ social media network in China), e-mail address, study major, online learning experience and experience of studying psychology courses. The demographic data were collected as controlling factors in the study.

Prior knowledge test. The prior knowledge test included 16 multiple choice questions to test participants' knowledge of developmental psychology. The test drew on a popular Chinese developmental psychology textbook, *Developmental Psychology* (Lin, 2005). There were four options on each item, and every item had two or more right options. Only if all right options were selected, did the participant get 1 point on an item. The total score was 16. The test had high discrimination (t (60) =16.94, *p*< .001).

Learning test. This test included two parts. The first part, which tested learning recall, contained 15 multiple choice questions testing participants' comprehension of key concepts in the video podcast. The scoring method was the same as with the prior knowledge test. The second part, which tested learning transfer, contained 3 short-answer questions testing participants' ability to apply their knowledge to novel situations. The total score was 25. The test had high discrimination (t (60) = 20.79, p<.001).

2.4 Design

A between-subject design was used with the presenting modes of the video podcasts as the between-subject variable. The between-subject variable included 4 levels: PPT, Instructor, Instructor and PPT and Classroom. The between-subject design meant that the participants were randomly assigned to one of the four experimental condition groups.

The process of video podcast learning was analyzed by eye movement tracking. Specifically, it could be identified what information was processed and visually focused on based on data of the learners' visual fixation, such as fixation duration, fixation counts and mean fixation duration. The eye-mind hypothesis suggests that there is a positive link between learners' fixation and visual attention, i.e., the more time fixated on an item, the more visual attention is being paid (Just & Carpenter, 1984). Saccade, which is a rapid movement of an eye, can reflect transition of visual attention (Kizilcec et al., 2014). Blink

duration, tracked by eye–tracking technique, is a highly sensitive measure for tracking fluctuating levels of mental fatigue. Mental fatigue describes the phenomenon that, over time, an individual engaged in cognitive activities gradually exhibits psychological loss. The longer the blink duration, the greater the mental fatigue; and the shorter the blink duration, the lighter the mental fatigue (Stern, Boyer, & Schroeder, 1994; Zhu et al, 2008).

2.5 Procedure

The study was carried out in a laboratory at the School of Psychology CCNU. After granting consent, all participants were first given the demographic questionnaire and prior knowledge test. They were then escorted to an eye–tracking room and randomly assigned to view, without pauses and in the absence of the experimenter, one of the four video podcasts. Before the video podcast started, participants were asked to follow a dot on the screen with their eyes for calibration and validation. Immediately after viewing the video podcast, participants filled out the learning test.

3 Results

3.1 Prior knowledge of different groups

One–way between–subjects analysis of variance was performed, with the scores of the prior knowledge test as the dependent variable and the presenting modes as the independent variable. No significant difference was found between the groups (F (3, 90) =1.36, p>.05).

3.2 Learning outcomes

The differences in learning test scores were compared between the different groups to assess the learning outcome

Table 1. Mean and standard deviations of the learning test scores.

Groups	N	M	SD
PPT	24	11.42	3.06
Instructor	22	8.52	3.80
Instructor and PPT	24	14.75	3.13
Classroom	24	11.25	2.73

of the different video podcasts by presenting mode. Table 1 shows the mean and standard deviation of the scores of the different groups. It was found that the participants in the Instructor and PPT group gained the highest scores on the learning test.

An analysis of covariance was performed with the scores of the prior knowledge test as the covariable, the presenting modes as the independent variable and the learning test scores as the dependent variable. This statistical method could exclude the effect of prior knowledge on the learning outcome. It was found that there was significant difference between the groups (F (3, 89) = 15.67, p < .001).

The results of the posthoc test (Bonferroni) are shown in Table 2. It was found that the learning test scores of the Instructor and PPT group were significantly higher than those of the other three groups. In addition, the Instructor group gained significantly lower scores than the PPT group and the Classroom group. There was no statistically significant difference between the PPT group and Classroom group. The results implied that video podcasts with Instructor and PPT optimally helped learners' learning.

3.3 Eye movements of the video podcast with Instructor and PPT

As the Instructor and PPT approach gave best results for learning, it is necessary to review the evidence from eye movements on how the participants viewed the video with this mode and how mental fatigue changed during the learning process.

3.3.1 Allocation of visual attention

In analysis of the Instruction and PPT mode, the video was divided into 2 areas of visual interest. These areas were the area of the instructor and that of the PPT slides. Table 3 shows the mean and standard deviation of the total fixation duration, fixation counts and mean fixation duration in the two areas. Looking at the percentage of

Table 2. Mean differences (*p*–value) between the learning test scores.

Groups	Instructor	Instructor and PPT	Classroom
PPT	3.24 (0.005) **	3.06 (0.007) **	0.15 (1.000)
Instructor		6.30 (0.000) ***	3.08 (0.008) **
Instructor and PPT			3.21 (0.004) **

**p<0.01; ***p<0.001.

Table 3. Mean and standard deviation of the total fixation duration, fixation counts and mean fixation duration for Instructor and PPT video podcasts.

Areas of interest	Total fixation duration (s)		Fixation counts		Mean fixation duration (ms)	
	M	SD	M	SD	M	SD
Instructor	691.04	19.00	2864.48	707.02	241.79	37.97
PPT slides	418.22	15.43	1113.62	355.36	378.09	94.20

total fixation duration on the areas of interest, it was found that the participants spent 62.30% of the time fixating on the instructor, and 37.70% fixating on the PPT slides, on average. The transition rate of visual attention between the areas, analyzed by saccade data, was every 24.59 s.

Paired samples *t*–tests were performed with the areas of interest. It was found that the total fixation duration in the area of the instructor was significantly longer than in that of the PPT slides ($t(23)=6.83$, $p< .001$). The fixation count in the area of the instructor was greater than in that of the PPT slides ($t(23)=14.92$, $p< .001$), and the mean fixation duration in the area of the instructor was shorter than in that of the PPT slides ($t(23)=7.53$, $p< .001$).

The above results indicate that the participants allocated much more visual attention to the instructor than to the PPT slides, which suggests that the instructor facilitated the enhanced learning outcome.

3.3.2 Mental fatigue in the process of learning

Mental fatigue was examined by blink duration, for which one minute was taken as the time unit and averaged blink duration per minute calculated. The data were sampled within two standard deviations.

There are three upturns in the participants' blink duration. The first is during the 10–13 min, with the participants' blink duration slowly rising from 567.20 to 1622.59 ms; the second is at the 22 min, where the participants' blink duration suddenly rises to 7881.79 ms, the peak value in this study; and the third upturn occurs at 25 min, where blink duration rises to 3984.37. The interval between the first and the second upturn was much longer than the interval between the second and the third upturn.

The results for mental fatigue therefore implied that the participants started to experience mental fatigue after 10 min of video podcast learning. The trend suggests that they were immediately aware that they felt tired and refreshed their mind; that is, they re–entered a normal learning mental status as found earlier in the learning period. However, they entered heavy mental fatigue status at the 22 min. At this point, although they were aware that they felt tired, they could no longer refresh their mind. In other words, they could no longer re–enter their normal learning mental status. The results for blink duration suggest that a video podcast with Instructor and PPT should not be longer than 21 min.

4 Discussion

The results of the study showed that of the four video podcast modes studied the Instructor and PPT mode most enhanced the learners' learning, which supported the first hypothesis that a video podcast with both the instructor and PPT slides best facilitates the learning outcome. Previous studies have shown that the presence of the instructor in

video podcasts affects learners' engagement and cognitive load, which in turn influences the learning outcome. They suggest that presence of the instructor contributes social cues while simultaneously increases the cognitive load (Gunawardena, 1995; Homer et al., 2008; Paas & van Merriënboer, 1994). The cognitive load increases because the instructor's presence provides information beyond what is taught. Furthermore, the learners' attention is drawn away from the supporting material in the PPT slides (Pass & van Merrënboer, 1994). Learning engagement and cognitive load concurrently affect learning, and they need to be well balanced (Pi, 2014). The video podcast with PPT imposed the lowest cognitive load. However, it did not include any social cues, which suggests that it created the lowest learning engagement, which hindered effective learning. The other two presenting modes, i.e. Instructor mode and Classroom mode, can be considered as not being ideal in the sense that social presence and cognitive load are not well balanced.

The results for the learning outcomes in our study confirm the results of Wang et al. (2014) and are in contrast with the results of Homer, et al. (2008) and Kizilcec et al. (2014). One possible explanation for the difference is that cultural differences have affected the learners' perception of social cues, which further influenced the learning outcomes. The participants in Homer et al. (2008)and Kizilcec et al. (2014)were American students, and the participants in our study and in the study of Wang et al. (2014) were Chinese students. Cultural studies have indicated that Chinese are relation-oriented and live in complex social networks with prescribed role relations, whereas Americans tend to be task-oriented, live in less constraining social networks and pay much more attention to the focal object rather than background information (Hong, Heikkinen, & Blomqvist, 2010; Markus & Kitayama, 1991; Nisbett & Miyamoto, 2005). Thus, Chinese students can be expected to be

particularly sensitive to social cues and their cognitive outcomes are consequently more based on relationships than Western students. In contrast, American students may pay less attention to social cues but more attention to the content of the study. Therefore, the instructor in the video podcast has had little effect on the American participants. Comments by our participants during informal interviews after they had viewed the video podcasts provided further evidence of their preference for the presence of the instructor.

Additionally, the fixation data for eye movement in our study support the above explanation. The data showed that the learners spent much more time looking at the instructor than the PPT slides, which supported our second hypothesis that learners using video podcasts with Instructor and PPT spend more than half the learning time fixated on the instructor. In contrast to the results in this study, Kizilcec et al. (2014) found that their participants spent less time looking at the instructor than the PPT slides. It would appear that compared with American students, Chinese students spend more time on social cues. The results of our study on podcast learning imply that the use of educational technology is culture bound, despite today' s global context of learning and teaching. However, this finding needs to be tested further with research focusing specifically on cross-cultural aspects.

The eye-tracking data revealed that in the podcast with Instructor and PPT the learners started to feel mental fatigue at the 10 min and reached the peak of mental fatigue at the 22 min. These results support our third hypothesis that learners' mental fatigue increases with time and reaches a peak at a certain point. The results for mental fatigue imply that if video podcasts can be divided into several sections, it would be best that the length of any section is kept below 10 min. If the video podcasts cannot easily be divided, the length should not exceed 22 min. The results, to some extent, are consistent with those of Danforth et al.

(2012), who found that 4–6 min video podcasts gave optimal learning outcomes. The length of the video podcasts in the study of Danforth et al. (2012)was, however, limited between 2 and 11.5 min, and hence nothing is known about changes in mental fatigue after 11.5 min.

To the best of our knowledge, this study is the first to systematically compare the learning outcome of four common presenting modes of video podcasts and explore via eye movement analysis the time when learners experience the onset of mental fatigue in the educational context of China. The focus of our study is the question of which presenting mode best facilitates learning. In the explanation of our major finding that the most effective presenting mode is Instructor and PPT, we assume that learners' engagement and cultural expectations play a critical role. We believe that Chinese sensitivity to social cues, which has been noted in other contexts, is equally applicable to the studied learning situation. At this stage, all conclusions on the mechanisms underlying the results of our study are inferred and thus need to be explored in more depth. Further research is also needed to investigate how social cues can enhance online teaching and learning in different cross-cultural contexts.

There are several limitations in this study. Firstly, in order to guarantee the accuracy of the eye movement data, the head of the participants was stabilized, which is a clear difference from natural learning situations and to a certain extent will affect the measured results. Secondly, mental fatigue was explored during the study of the psychological concept of *attachment*; a different learning subject might give different results. Despite these limitations, the study makes a meaningful contribution to discussion of video podcast learning by demonstrating the clear differences in learning effectiveness of the different presentation modes, by illustrating the use of quantitative visual behavior data in study of video podcast learning, and by presenting a rationale for the contradictory findings in the literature.

References

Chen, C. M., & Wu, C. H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education, 80*, 108–121.

Copley, J. (2007). Audio and video podcasts of lectures for campus-based students: Production and evaluation of student use. *Innovations in Education and Teaching International, 44*, 387–399.

Danforth, S., Schumacher, J., Cullen, R., & Ma, Y. J. (2012). Evaluating format preferences and effectiveness of video podcasts related to nutrition education and recipe demonstrations. *Journal of Nutrition and Dietetics, 112,* A19.

Gonzalez, C., Best, B., Healy, A. F., Kole, J. A., & Bourne, L. E. J. (2011). A cognitive modeling account of simultaneous learning and fatigue effects. *Cognitive Systems Research, 12*, 19–32.

Gunawardena, C. N. (1995). Social presence theory and implications for interaction and collaborative learning in computer conferences. *International Journal of Educational Telecommunications, 1,* 147–166.

Homer, B. D., Plass, J. L., & Blake, L. (2008). The effects of video on cognitive load and social presence in multimedia-learning. *Computers in Human Behavior, 24*, 786–797.

Hong, J. Z., Heikkinen, J. & Blomqvist, K. (2010). Culture and knowledge co-creation in R&D collaboration between MNCs and Chinese universities. *Knowledge and Process Management, 17*, 62–73.

Hyona, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*, 172–176.

Just, M. A., & Carpenter, P. A. (1984). Using eye fixations to study reading comprehension. In D. E. Kieras & M. A.

Just (Eds.), *New Methods in Reading Comprehension Research* (pp. 151－182). Hillsdale, NJ: Erlbaum.

Kizilcec, R. F., Papadopoulos, K., & Sritanyaratana, L. (2014). Showing face in video instruction: Effects on information retention, visual attention, and affect. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2095–2102). New York: USA.

Li, S. K. (2013). Review of application of podcasts in teaching in China. *Software Guide, 12*, 195－197 (in Chinese).

Lin, C. D. (2005). *Developmental Psychology*. Hangzhou: Zhejiang Education Publishing House (in Chinese).

Lyons, A., Reysen, S., & Pierce, L. (2012). Video lecture format, student technological efficacy, and social presence in online courses. *Computers in Human Behavior, 28*, 181－186.

Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*, 224－253.

McGarr, O. (2009). A review of podcasting in higher education: Its influence on the traditional lecture. *Australasian Journal of Educational Technology, 25*, 309－321.

Nisbett, R. E., & Miyamoto, Y. (2005). The influence of culture: Holistic versus analytic perception. *Trends in Cognitive Science, 9,* 467－473.

Paas, F., & van Merri nboer, J. J. G. (1994). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*, 51－71.

Pi, Z. L. (2014). *The presenting formats of video podcasts and their effect on learning effectiveness and learning mechanism: An eye tracking study*. Unpublished master thesis. Wuhan: Central China Normal University (in Chinese).

Stern, J. A., Boyer, D., & Schroeder, D. (1994). Blink rate: A possible measure of fatigue. *The Journal of the Human Factors and Ergonomics Society, 36*, 285–297.

Varnum, M. E. W., Grossmann, I., Kitayama, S., & Nisbett, R. E. (2010).The origin of cultural differences in cognition. The social orientation hypothesis. *Current Directions in Psychological Science, 19*, 9－13.

Wang, J., Hao, Y. H., & Lu, J. L. (2014). The effect of presenting mode of teaching video on self－directed learning effectiveness: An experimental study. *E－education Research, 251*, 93－105 (in Chinese).

Zhang, D., Zhou, L., Briggs, R. O., & Nunamaker Jr, J. F. (2006). Instructional video in e－learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management, 43*, 15－27.

Zhu, Z. H., Wu, X. J., Wang, L., & Qi, L. (2008). Detection method of driver fatigue based on blink duration. *Computer Engineering, 34*, 201－206.

在学院成立当年，湖北省教育厅批准成立"湖北省青少年心理健康教育中心"和"人的发展与心理健康湖北省重点实验室"；同年获得心理学一级学科博士学位授予权。此后几年陆续获得"发展与教育心理学"湖北省重点学科和心理学一级学科湖北省重点学科，博士后流动站，青少年网络心理与行为教育部重点实验室，湖北省高校本科品牌专业和教育部高等学校特色专业，湖北省心理学实验教学示范中心和国家级心理学实验教学示范中心，国家级虚拟仿真实验教学中心，并加入中国基础质量监测协同创新中心。这些步步提升的专业和学科建设项目的获得，以及实验室和研究中心的建立，为学院的发展搭建了广阔的平台，提供了有利的条件。

心理学院建院十年也是与国家共同成长的十年，以网络为载体的大数据、即时通讯、交互娱乐和网络购物等已经、而且还在迅速地改变着人类的生活，它们对人们心理和行为的深刻冲击也是前所未有的。心理学院的同仁看到了这场变革所带来的难得机遇，提出今后的发展要围绕"青少年网络心理与行为教育部重点实验室"这一研究平台，以"国家级心理学实验教学示范中心"和"国家级虚拟仿真实验教学中心"为主要教学平台，进一步搭建全球化和信息化的教学、科研和社会服务体系，为人才培养、科学研究和社会进步事业做出更大的贡献！

华中师范大学
SCHOOL OF PSYCHOLOGY, CENTRAL CHINA NORMAL UNIVERSITY

为天地立心，为生民立命，为往圣继绝学，为万世开太平。

——[北宋] 张载