

《心理学报》审稿意见与作者回应

题目：基于 SCM 不同热情-能力社会群体刻板印象激活效应的行为模式

作者：杨亚平；徐强；朱婷婷；郑旭涛；董晓晔；陈庆伟

第一轮

编委意见：

意见 1：预研究选择刻板特质词时，第三步量表评定高频词符合社会群体特征的典型程度，未报告选取样本情况、具体结果等，需补充。

回应：非常感谢编委的重要建议。已在正文“2.1.2 实验材料”部分补充报告了第三步量表评定高频词符合社会群体特征的典型程度的选取样本情况、具体结果。补充报告的内容在文中用黄色高亮部分标出。

意见 2：实验 1 和实验 2 中在研究设计中都平衡了被试间按键左右手，结果中未报告这一被试间变量的结果，需补充以使研究结果更加严谨

回应：非常感谢编委的建议。在本研究中实验 1 和实验 2 中在研究设计中都平衡了被试间按键左右手，每个实验中一半被试的指导语要求用左手食指按“一致”键，用右手食指按“不一致”，另一半被试用左手食指按“不一致”键，用右手食指按“一致”（其它的指导语完全一致）。在绝大多数需要用双手按键的实验中，研究者在做实验设计时必须考虑到左右手的平衡，以控制无关变量（即消除左右手按键的速度混淆了实验结果，如对“一致”靶子的反应时更快或许是因为右手按键快，而如果左右手做了平衡，就控制了这一无关变量的影响），但并未将左右按键左右手无关变量作为一个自变量来统计分析。鉴于此以及编委的建议，笔者补充分析了这一被试间变量，但是在正文中并未呈现补充分析的结果，而是将其呈现在脚注 1 和 2 中(用黄色高亮部分标出)，结果显示实验 1 和实验 2 中按键左右手这一无关变量对结果没有显著影响，因此在实验 1 和实验 2 的结果统计分析中将两组数据合在一起，重点分析本研究操纵的两个主要自变量(社会群体，一致性)的效应，亦即原稿中的结果部分。

意见 3：实验 1 和实验 2 有关正确率的研究结果在四类群体中不完全一致，部分验证了研究假设，但在之后的结论和讨论中将其放在一起说，结论不够严谨。虽正确率的结论与反应时相近，接近假设，但毕竟未达到显著水平，请分开讨论，甚至解释一下为什么正确率未达到理想结果。

回应：非常感谢编委的细致和严谨。诚如编委所说，把反应时和正确率的结果放在一起说，结

论并不严谨，我们在文章中修改了相关说法，以使文章中的结论更加符合实验数据结果，在讨论时候我们也对正确率的结果进行了分别讨论，具体修改内容在文中以黄色标亮。详细信息如下：

P8 摘要部分，修改了研究结论，以使其更符合实验结果，具体如下：

“结果发现，无论是外显层面还是内隐层面，高热情-高能力、高热情-低能力、低热情-高能力三类社会群体均遵循经典的刻板印象激活模式——一致条件下的反应更快；而低热情-低能力群体则在外显和内隐加工上均表现出刻板印象激活效应的反转模式——冲突条件下比一致条件下的反应更快更准确。”

P9 “问题提出”第二段：因为以往刻板印象激活效应研究的反应时结果比较一致，而正确率结果并没有反应时那么一致，因此在介绍刻板印象激活效应时，我们在正确率前面加了“往往”，以使行文更加准确。具体修改为：

“尽管所探讨的刻板印象不尽相同（如性别、种族、年龄、职业刻板印象等），但大量行为研究一致发现刻板印象激活效应的基本模式为：当靶子词与之前的启动所激活的刻板印象一致时，与冲突靶子词相比，被试的反应时更快、正确率往往更高”

P18 “2.3 讨论”部分，我们修改了结论部分以使其更加符合实验结果，具体修改如下：

“具体来说，当启动 HW-HC 群体时，相对于冲突靶子，对一致靶子做出一致性判别的正确率更高，反应时显著更快，当启动 HW-LC 和 LW-HC 群体时，对一致靶子比冲突靶子做出的反应更快；而当启动为 LW-LC 社会群体时，则出现了相反的趋势——相对于冲突靶子，对一致靶子词进行一致性判别的正确率显著更低，反应时显著更长。”

P21 “3.3 讨论”部分，我们修改了结论部分以使其更加符合实验结果，具体修改如下：

“具体来说，当启动 HW-HC 和 HW-LC 社会群体时，相对于冲突靶子，对一致靶子进行真假词判别的正确率更高，反应时显著更快，当启动是 LW-HC 群体时，对一致靶子的判断要显著快于冲突靶子；而当启动为 LW-LC 社会群体时，同样出现了刻板印象激活的反转效应——相对于冲突靶子，对一致靶子词进行真假词判别的正确率显著更低，反应时显著更长。”

P22 “4 总讨论”部分，我们增加了正确率的讨论，具体如下：

“与反应时结果略有不同的是，正确率的结果并不完全符合本研究的实验假设，具体表现在实验一中 HW-LC 和 LW-HC 群体和实验二中 LW-HC 群体的刻板印象效应上一致和冲突条件在正确率上没

有达到预期的统计上的显著差异,这与前人研究结果相一致(Wang et al., 2018; Wang, Yang, Tan, Chen, & Van Cantfort, 2017)。考虑到以往使用启动范式探讨刻板印象激活效应的研究在反应时指标上的结果非常一致,而在正确率指标上的结果并不十分一致(见综述 Kidder, White, Hinojos, Sandoval, & Crites, 2018)。出现这一结果也并不意外,因为在大多采用启动范式的研究中反应时是更为灵敏的指标,也正是考虑到这一点,White 等人(2018)的最新研究仅使用反应时来考察刻板印象激活效应而并未分析正确率这一指标。”

P23 “4 总讨论”部分,我们修改了结论部分以使其更加符合实验结果,具体修改如下:

“具体来说,本研究的结果从以下三个方面补充和扩展了以往文献:首先,本研究发现对于 HW-HC、HW-LC 和 LW-HC 社会群体均出现了传统上典型的刻板印象激活效应模式——相对于冲突靶子,对一致靶子的反应时显著更快。”

意见 4: 实验 2 被试描述中,提到一名被试错误率低于 50%而被删除,请问依据什么标准?是否描述有误?

回应: 非常感谢编委严谨细致的问题。一名被试错误率低于 50%而被删除是笔误,文中已经将这句改为“一名被试由于正确率(51.13%)远低于平均正确率 93.58%(见表 2)而被剔除”。

关于剔除的标准,是笔者根据实验 2 的平均正确率而主观定的。具体来说,实验 2 中的正确率很高,如正文“表 2 词汇判断任务中不同社会群体启动下对靶子进行真假词判别的正确率和反应时”所示,每种实验条件下的正确率最低的也达到 91.4%,平均正确率是 93.58%。而这名被试 8 中实验处理下的正确率普遍很低,平均正确率只有 51.13%,推测其没有认真完成实验任务,故在数据统计分析中将其剔除。

意见 5: 细节问题,文中个别地方中英标点符号有误,语句不通顺出现多余“的”字等,请仔细修改。

回应: 谢谢编委的建议。我们仔细多次通读全文,一一修改完善了上述编委剔除的问题。具体有:

(1) 将中英标点符号统一化,参照近年来学报的文章,本文中的逗号(,)、分号(;),冒号(:)和圆括弧()等都统一为英文状态的,而句号(。)、顿号(、)和引号(“”)都统一为中文状态的。

(2) 通过多次通读全文,已仔细修改了多处语句不通顺处。因是表述类的细节且较多,因此没有一一在下面列出说明。

此外,我们还修改了一些小地方,如下:

(1) 在文章修改时,以下这篇文献已经正式出版,因此我们更新了文献的相关信息。

Kidder, C. K., White, K. R., Hinojos, M. R., Sandoval, M., & Crites, S. L. (2018). Sequential stereotype priming: A meta-analysis. *Personality and Social Psychology Review*, 22(3), 199-227.

(2) 在问题提出中综述国内刻板印象内容模型相关研究时,我们也增添了《心理学报》最新发表的相关成果。

“佐斌等人(2018)探讨了评价意图与结果在群际评价时对热情和能力关系的情境性影响。”

Zuo, B., Wen, F., Wu, Y., & Dai, T. (2018). Situational evolution of the relationship between warmth and competence in intergroup evaluation: Impact of evaluating intention and behavioral outcomes. *Acta Psychologica Sinica*, 50(10), 1180-1196.

[佐斌, 温芳芳, 吴漾, & 代涛涛. (2018). 群际评价中热情与能力关系的情境演变: 评价意图与结果的作用. *心理学报*, 50(10), 1180-1196]

(3)在总讨论部分,我们也增加了最新的英文相关成果,如:

“这一结果与以往大量采用单个/类群体探讨刻板印象激活效应的研究结论相一致(e.g., Dovidio, Evans, & Tyler, 1986; Hehman, Volpert, & Simons, 2014; Wang, Yang, Tan, Chen, & Van Cantfort, 2017; Macrae, Milne, & Bodenhausen, 1994; White, Crites, Taylor, & Corral, 2009; White, Danek, Herring, Taylor, & Crites, 2018; 陈莉 & 王沛, 2015; 王沛, 杨亚平, & 赵仑, 2010; Zarate & Smith, 1990)”。

第二轮

审稿人 1 意见:

意见 1: Please indicate the power analysis you've done for the studies.

回应: 非常感谢审稿专家的重要建议。已在正文“**2.1.1 被试**”部分补充报告了统计效能分析结果。补充报告的内容在文中用黄色高亮部分标出,即“为了保证较高的统计效能,在实验实施前采用 G*power 3 分析方法对样本量进行估算(Faul, Erdfelder, Buchner, & Lang, 2009)。具体来说,根据中等的效应量($f^2 = 0.20$, Cohen, 1988), 0.01 的 α 水平以及实验 1 的实验设计来计算确定样本量, G*power 分析结果建议实验 1 需要 46 名被试可使统计效能达到 0.953。考虑到可能有无效被试,实际取样 50 名被试,有效数据 50 名。”

同样,在实验 2 “**3.1.1 被试**”分补充报告了实验 2 统计效能分析结果。补充报告的内容在文中用黄色高亮部分标出,即“为了保证较高的统计效能,在实验实施前采用 G*power 3 分析方法对样本量进行估算(Faul, Erdfelder, Buchner, & Lang, 2009)。具体来说,根据中等的效应量($f^2 = 0.20$, Cohen, 1988), 0.01 的 α 水平以及实验 1 的实验设计来计算确定样本量, G*power 分析结果建议实验 1 需要 46 名被试可使统计效能达到 0.953。考虑到可能有无效被试,实际取样 48 名被试,有两名被试数据未参与分析,其中一名被试由于正确率(51.13%)远低于平均正确率(93.58%,见表 2)而被剔除,一名被试未完成实验,最后进行统计分析的被试有 46 名。”

意见 2: In the first paragraph of the general discussion, I fail to see how White and colleagues' 2018 study applies to the context of your research. Please make it more specific in your writing.

回应：非常感谢审稿专家仔细的意见。经过我们仔细审读 White 等人(2018)的研究之后，发现他们的文章更多强调实验范式的不同对刻板印象激活效应的影响，因此放在文章此处确实不太合适，因此我们在修改稿中的这一部分删去了这条文献的相关信息。修改部分如下：

“结果部分支持了假设，即无论是外显层面抑或内隐层面，HW-HC、HW-LC 和 LW-HC 群体均出现了传统上典型的刻板印象激活效应模式——相对于冲突靶子，对一致靶子的反应时显著更快，简言之，一致条件下被试的反应更快；这一结果与以往大量采用单个/类群体探讨刻板印象激活效应的研究结论相一致(e.g., Dovidio, Evans, & Tyler, 1986; Hehman, Volpert, & Simons, 2014; Wang, Yang, Tan, Chen, & Van Cantfort, 2017; Macrae, Milne, & Bodenhausen, 1994; White, Crites, Taylor, & Corral, 2009; 陈莉 & 王沛, 2015; 王沛, 杨亚平, & 赵仑, 2010; Zarate & Smith, 1990)。”

而 White 等人(2018)的研究中汇总了他们团队的刻板印象激活效应的实验研究并进行对比，其中就包括了本研究使用的两种实验范式，并且他们的研究考虑到正确率指标有时对刻板印象激活效应并不一定敏感，因此只使用了敏感性更高的反应时指标，这也能一定程度上解释本研究正确率的实验结果。因此，我们认为 White 等人(2018)的研究与本研究还是具有直接的相关关系。为了更好说明这一部分的内容，我们增加了相关说明，并以黄色部分标出。修改部分如下：

“出现这一结果也并不意外，因为在大多采用启动范式的研究中反应时是更为灵敏的指标，也正是考虑到这一点，White 等人(2018)的最新研究仅使用反应时来考察刻板印象激活效应而并未分析正确率这一指标，这也提示未来研究中在考察刻板印象激活效应时可能需要更多考虑反应时指标的结果。”

.....

审稿人 2 意见：

本研究讨论了刻板印象内容模型中不同类别的热情与能力群体的刻板印象激活效应的行为模式，并通过外显、内隐方式加以探讨，并发现激活效应在不同群体中的启动的差异性。文章思路较为清晰，写作较为规范。提出如下问题，建议作者完善。

意见 1：实验材料中，预研究选择**刻板特质词未报告选取样本的方法和基本规模**。因为预研究所得群体是基础群体，如抽样不合理或选取有偏差会直接影响后续研究。作者在评审意见一后虽然做了补充，但是未精确，只描述了多少名大学生，如果是同质性群体的整群抽样事实上并不能完全具有较高的信效度。

回应：非常感谢审稿专家详细严谨的意见和建议。原文方法实验材料部分关于刻板特质词的前

测的确未报告选取样本的方法和样本基本规模,这可能会让读者对这一过程不是非常了解甚至质疑实验结果的有效性。诚如专家所言“如抽样不合理或选取有偏差会直接影响后续研究……如果是同质性群体的整群抽样事实上并不能完全具有较高的信效度。”在专家这一重要意见的提示下,笔者回溯当初刻板特质词的编制过程,查看前测的问卷等相关资料。所幸的是,在当初编制刻板特质词前测过程中的每个步骤都没有采取整群抽样的方式,而是在某大学的图书馆、阅览室、自习教室随机发放问卷收集数据的。

为了使读者更加了解本研究刻板特质词的编制过程,笔者已根据专家的意见在正文“2.1.2 实验材料”,“靶刺激”部分补充报告了刻板特质词前测过程中选取样本的方法和样本基本规模。补充报告的内容在文中用黄色高亮部分标出,具体来说,有如下两处做了补充说明:

“(1)首先,通过开放式问卷(见附录 2)收集每个群体的刻板特质词,指导语要求被试根据印象写出描述每个社会群体典型特征的词语。在某大学的图书馆、阅览室、自习教室随机发放问卷 56 份,回收有效问卷 48 份。”

“(3)采用 7 点量表评定前面选出的高频词描述相应社会群体特征的典型程度,即确定哪些词语是某个社会群体典型的刻板特质词。指导语要求被试根据特质词描述社会群体的贴切程度进行评分。如:根据你对“科学家”的印象,请对下列词语描述“科学家”的贴切程度在 7 点量表上进行评分,1-7 表示贴切程度依次加强:1 代表“非常不贴切”,7 代表“非常贴切”。在某大学的图书馆、阅览室、自习教室随机发放问卷 60 份,回收有效问卷 52 份”

意见 2: 实验 1 和实验 2 的词汇包括部分假词,的确假词可以做与真词的对比。但是假词中使用类似“溺”、“贫”等接近“弱”和“贫”字形的词汇,以及“残暴”与“贱泰”,是否考虑到类似词汇属于非中性词汇,容易启动被试的不同情绪,甚至激活不同反应。也就是说,使用中性词汇可能清除无关干扰,但是本身有些极为负面的词汇容易启动不必要的激活。建议作者予以缜密思考,词汇的选择需要更为严谨。

回应: 非常感谢审稿专家的重要建议。借鉴以往相关研究的制作假词的思路(e.g., Hill, Strube, Roesch-Ely, & Weisbrod, 2002),笔者及其研究者在实验 2 制作假词时仅仅考虑到:假词中的每个字是真词中每个字的形近字且组合起来不是一个词,这样真假词形近且笔画一致。也就是说,我们仅仅考虑到了物理属性的一致而未考虑到假词中个别字具有明显的积极或者消极的效价,在以后的研究中我们也会考虑这一因素,在物理属性的一致前提下尽量避免使用具有极端效价的字。

就本研究的实验 2 数据分析的对象而言,笔者推测假词少数具有极端效价的字并不会影响实验结果。这是因为,实验 2 采用词汇判断任务(LDT),被试的任务是判断靶子刺激是不是一个词,但假

词只是为了构成实验任务。尽管记录了被试对真词和假词的反应,但结果只分析真词,而不分析假词,因此理论上假词的效价并不会影响实验的结果。

Hill, H., Strube, M., Roesch-Ely, D., & Weisbrod, M. (2002). Automatic vs. controlled processes in semantic priming — differentiation by event-related potentials. *International Journal of Psychophysiology*, 44(3), 197-218.

意见 3: 文章错别字过多,建议通读并修改。如 P9 中错别字“圭臬”? P11 “冀 全面揭示”,“冀”是否为“既”? 类似问题,建议修正。

回应: 非常感谢审稿专家的细致严谨的意见和建议。笔者及其他作者通过了全文,修改了一些生僻词语和错别字等等表述不当之处。

如“圭臬”其意为标准,尺度,尽管放在“……因此在反应时上一致条件快于冲突条件已被社会认知领域研究者们视为判定刻板印象激活效应发生的圭臬”意思是通的,但该措辞太过生僻,笔者已将其改为“标准”;再如“冀”其意为“希望”,尽管放在“冀全面揭示不同热情-能力社会群体刻板印象激活效应的行为模式”意思也通,但因过于生僻影响读者对于关键信息的获得,故将其改为“力图”。

同样,笔者及其他作者在反复通读全文的过程中,也一一修改了文中其它的生僻措辞,错别字等表述不当之处,因所改之处较多,没有在此一一列举,改后的内容在文中用绿色高亮标出。

意见 4: 本研究在实验任务上选择启动与靶子一致性判断任务来探讨外显加工,使用词汇判断任务来探讨内隐加工,建议能更加深入化。

回应: 非常感谢审稿专家的重要建议。诚如专家所言,本研究采用启动-靶子一致性判断任务和词汇判断任务分别从外显和内隐加工层面对 SCM 四类群体刻板印象激活效应的行为模式进行考察。

本研究实验 1 选用“启动-靶子一致性判断任务”来探讨外显加工,因为以往研究显示通过该任务可以发现典型的刻板印象激活效应模式(White, Crites, Taylor, & Corral, 2009; 王沛, 杨亚平, & 赵仑, 2010)。其原因是在该任务中被试只有意识到(即外显层面)并对启动-靶子之间的关联(如,消防员-勇敢)作出准确判断,才能最终做出正确的反应(如,按“一致”键)。

实验 2 选用词汇判断任务来探讨内隐加工,也是因为以往诸多研究采用该范式来探讨内隐或自动的加工。其原因是在该任务中被试只需要判断靶子刺激是不是一个词,无需意识到启动-靶子之间的关联也能作出正确判断和反应。在这种任务无关的情境下,如果刻板印象一致和冲突条件在反应时或者正确率上依然出现了显著差异,则证明在内隐加工层面刻板印象也可以激活,亦即刻板印象激活不受外显/内隐实验任务的影响。

尽管从逻辑上和以往研究文献的基础上,选择启动-靶子一致性判断任务来探讨外显加工,使用

词汇判断任务来探讨内隐加工是可行的，本研究的结果也是可靠的。然而，还有其它探讨外显加工和内隐加工的实验范式。因此，今后的研究还需要采用其它探讨外显加工和内隐加工的实验范式进一步验证和拓展本研究的结果。这一点在本文的讨论部分最后一段补充说明“首先，在实验任务上，仅使用了经典的启动-靶子一致性判断任务来探讨外显加工，也仅使用了词汇判断任务来探讨内隐加工，未来研究可以采用其它探讨外显加工和内隐加工的实验范式和任务(Kidder, White, Hinojos, Sandoval, & Crites, 2018; White, Danek, Herring, Taylor, & Crites, 2018) 进一步检验该研究结论的边界;……”这也是笔者及其研究团队接下来要开展的工作。

最后，再次感谢两位审稿专家花费宝贵的时间和精力为本文提出了严谨而重要的意见和建议！

第三轮

审稿人 1 意见：

意见 1：文章 2.1.2 的实验材料中，启动刺激固然通过前测获得，但是否具有典型性和有效性需要考虑。如瑜伽教练属于 HW-HC 群体？城管属于 LW-LC 群体？一般而言，刻板印象内容模型显现群体应该为社会生活中的典型群体，而作者列举的上述群体是社会生活中的群体，但是否为高频群体和典型群体需要斟酌。尤其是，文章数据中将“城管”划分为 LW-LC 群体，即“第 3 个群组有城管、性工作者、无业游民、乞丐、罪犯、恐怖分子、吸毒者”，是否妥当？

回应：非常感谢审稿专家严谨的意见和建议。逐一回应如下：

首先，关于启动材料的典型性和有效应，本研究中启动刺激(基于 SCM 不同热情、能力的社会群体)是严格按照以往相关研究测量 SCM 的步骤和方法，通过足够样本的预研究获得，因此从理论上来说是有效的，但是也需要补充说明某些前测的过程和结果。详述如下：

①SCM 测量的第一步是通过开放式问卷收集体群，本研究参照以往研究中常用来收集社会群体的方法获得 41 个群体。但是原文对这一过程陈述的太过简略，仅仅列出了 41 个群体。笔者已在附录 1 预研究“2.2 研究材料与工具”第一段详细报告了被试、样本量以及对所收集体群的频次统计结果——“(1)研究材料：研究材料为 41 个社会群体。参照前人研究调查社会群体的方法，通过开放式问卷请被试回答以下问题(管健，程婕婷，2011;高明华，2010)：“社会生活中人们会根据不同的标准(年龄、性别、职业、地域、收入、社会地位等)将人划分成各个不同的群体，您认为当今的中国社会可以划分为哪些群体?”。在某大学发放问卷 200 份，在回收的 184 份有效问卷中，男生 84 名，女生 95 名，未报告性别者 5 名，平均年龄 21.25 岁。对 184 份问卷上所列出的社会群体进行整理和频

次统计, 结果如表 1 所示, 共收集到频次大于 10% 的群体有 41 个。”

表 1 社会群体频次统计结果(%)

序号	群体名称	%	序号	群体名称	%	序号	群体名称	%
1	公务员	72.80	15	留守儿童	50.91	29	空姐	28.86
2	农民工	70.48	16	企业家	49.28	30	残疾人	26.70
3	大学生	70.19	17	家庭主妇	48.39	31	城管	25.17
4	演艺明星	68.51	18	大学教师	46.96	32	罪犯	24.89
5	白领	66.88	18	军人	43.19	33	下岗工人	23.32
6	老年人	65.84	20	心理咨询师	41.25	34	暴发户	20.71
7	政府官员	64.11	21	科学家	40.41	35	恐怖分子	18.94
8	富人	62.71	22	瑜伽教练	39.24	36	同性恋	17.19
9	知识分子	60.49	23	乞丐	38.10	37	吸毒者	16.45
10	商人	59.88	24	工人	36.72	38	低保人员	15.05
11	农民	58.87	25	体育明星	35.14	39	私营企业主	14.15
12	年青人	56.22	26	无业游民	33.21	40	女强人	12.58
13	清洁工	54.31	27	海归	30.29	41	性工作者	10.95
14	穷人	53.93	28	消防员	29.04			

在反思审稿专家的意见“……刻板印象内容模型显现群体应该为社会生活中的典型群体, 而作者列举的上述群体是社会生活中的群体, 但是否为高频群体和典型群体需要斟酌”的基础上, 笔者将本文预研究群体收集相关数据与以往研究对照。如下表 A 所示: 本文预研究保留了频次大于 10% 的群体, 将其默认为高频群体; 而在两项引用较高的研究中, 同样将频次大于 10% 的群体视为高频群体。例如, 在管健和程婕婷(2011)的研究中“……收集整理 125 名被试的所有问卷, 根据被试分类群体进行频次分析, 发现高频群体 32 个, 包括农民 (68.3%) ……和罪犯 (12.1%)。”

表 A 社会群体收集相关研究汇总比较

研究来源	研究被引量 (CNKI)	样本 来源	有效 问卷	群体 个数	群体频次	
					最高频次	最低频次
高明华. (2010). 刻板印象内容模型的修正与发展——源于大学生群体样本的调查结果. <i>社会</i> , 30(5), 193-216.	28 次	大学生	115 份	21 个	54.57%	10.13%
管健, & 程婕婷. (2011). 刻板印象内容模型的确认、测量及卷入的影响. <i>中国临床心理学杂志</i> , 19(2), 184-188.	55 次	大学生	103 份	32 个	68.30%	12.10%
本文预研究		大学生	184 份	41 个	72.80%	10.95%

更重要的是, 所用于本研究正式实验的 24 个启动群体的频次普遍较高(如上表 1 加了浅蓝色底纹的群体), 除了“吸毒者”的频次为 16.45%, 其它 23 个群体的频次均高于 24%, 均大于 Fiske 等人制定的 15% 标准(Fiske, 2018; Fiske, Cuddy, Glick, & Xu, 2002; Wu, Bai, & Fiske, 2018)。由此可见, 本研究正式试验所采用的 24 个启动群体可以代表大学生样本中的高频群体。

参考文献:

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, 27(2), 67 – 73.

Fiske, S. T., Cuddy, A. J., Glick, P., & 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(6), 878 – 902.

Wu, S. J., Bai, X., & Fiske, S. T. (2018). Admired rich or resented rich? How two cultures vary in envy. *Journal of Cross-Cultural Psychology*, 49(7), 1114 – 1143.

可能由于调查的时间、样本不同，预研究所收集的一部分群体(如表 1 所示)与上述两项研究所收集的群体(如下表 B 所示)或者群体的高频程度不同，这也许是随着时代变化，大学生群体所直接或者间接了解的社会群体也在发生变化。

表 B 前人两项研究收集的社会群体汇总表

管健, & 程婕婷. (2011). 高频群体32 个	高明华. (2010). 高频群体32 个
农民(68.3%)、农民工(65.2%)、男性(59.1%)、女性(58.2%)、教师(56.3%)、老人(55.3%)、白领(54.9%)、公务员(52.1%)、知识分子(50.3%)、企业家(49.8%)、城市人(46.3%)、蓝领(45.1%)、商人(41.3%)、南方人(40.9%)、北方人(39.4%)、富人(36.8%)、穷人(37%)、大学生(35.6%)、科学家(35.8%)、残疾人(33.8%)、个体工商户(30.9%)、海归(30.1%)、私营企业主(24.8%)、领导干部(23.6%)、无业游民(22.3%)、低保人员(19.7%)、演艺明星(19.1%)、体育明星(18.9%)、乞丐(16.7%)、下岗人员(15.6%)、常驻中国的外国人(13.5%)和罪犯(12.1%)	老年人(54.57%)、农民(40.04%)、少数民族(34.91%)、医生(34.50%)、大学生(29.79%)、富人(28.93%)、技术工人(28.93%)、农民工(25.51%)、公司企业白领(23.80%)、人民解放军(21.04%)、人民警察(19.70%)、大学教师(17.82%)、失业人员(16.97%)、知识分子(16.97%)、个体户(15.26%)、政府官员(11.84%)、80' 后青少年(11.84%)、宗教信徒(11.13%)、残疾人(10.98%)、明星艺人(10.98%)、同性恋(10.13%)

②SCM 测量的第二步是确定热情和能力两大维度上的具体项目，并通过 Likert 量表对各个群体的每个项目进行评价，这一步骤也参照以往研究者常见的做法，因此可以确保其有效。

如附录 1 预研究“(2)研究工具”部分所述——“根据 Fiske 以及近几年关于刻板印象内容模型的测量的研究(Asbrock, 2010; Bye, Herrebrøden, Hjetland, Røyset, & Westby, 2014; Durante, Fiske, Gelfand, Crippa, Suttora, & Stillwell, et al., 2017; Fiske, Cuddy, Glick, & Xu, 2002; 管健 & 程婕婷, 2011)，确定了 SCM 中测量热情和能力两大维度的具体项目…… 接下来以测量 SCM 的研究为参考(Asbrock, 2010; Bye et al., 2014; Durante et al., 2017; Fiske et al., 2002; 管健 & 程婕婷, 2011)，编制了本研究中的问卷。”

第二，关于群体在热情，能力维度的划分。笔者在思考审稿专家的意见和疑问（……如瑜伽教练属于 HW-HC 群体？城管属于 LW-LC 群体？……尤其是，文章数据中将“城管”划分为 LW-LC 群体，即“第 3 个群组有城管、性工作者、无业游民、乞丐、罪犯、恐怖分子、吸毒者”，是否妥

当?)的基础上,重新检查了预研究 SCM 测量数据的统计分析方法和过程,笔者确定所采用的分析方法和过程是依据经典研究(Bye et al., 2014; Durante et al., 2017; Fiske et al., 2002),因此这样在统计过程上可以保证所划分的四类群体的可靠性。

此外,关于“城管”是否属于 LW-LC 群体这一疑问,我们也查阅了相关资料和文献,结果发现,随着近年来新闻媒体和社交网站上关于城管的负面新闻报道日益增多,“城管”这一群体被严重污名化(刘晓伟,2014; 汤天甜 & 李杰,2016; 谢和平 & 田晓红,2013; 杨小雨,2018; 张洋,2015; 张玉磊 & 任帆,2014),甚至有“妖魔化”的倾向(郭庆梦,2018),以至于城管人员自身都会对自己的职业身份产生认同威胁(赵静 & 杨宜音,2017)。因此,在本文预研究中“城管”被归类于 LW-LC 群体是可以理解的。

参考文献:

- 郭庆梦. (2018). 城管“妖魔化”现象的社会学分析. 上海,华东理工大学.
- 刘晓伟. (2014). 网络公共空间中城管形象的污名化与突围之路——以武汉“城管卧底”事件为例. *新闻界*(1), 7 - 10.
- 汤天甜, & 李杰. (2016). 传播偏向、群体极化与风险放大——城管污名化的路径研究. *西南交通大学学报(社会科学版)*, 17(5), 22 - 29.
- 谢和平, & 田晓红. (2013). 城管污名化的形成机制与去污名化对策研究——以武汉市为例. *理论导刊*(7), 18 - 21.
- 杨小雨. (2018). 城管污名研究. 上海,华东师范大学.
- 张洋. (2015). “城管”媒介形象的建构、污名化成因与重构路径. *宁夏大学学报(人文社会科学版)*, 37(1), 177 - 181.
- 张玉磊, & 任帆. (2014). 城管污名化的形成机理与去污名化的路径探析. *城市观察*(3), 149 - 154.
- 赵静, & 杨宜音. (2017). 城管的身份认同威胁及其身份协商策略. *学术研究*(4), 63 - 68.

综上,笔者根据审稿专家的意见,反思并检查了确定启动刺激材料的预研究,补充了相关的描述和统计结果,可以确定启动材料的有效性。另外,对正式实验数据剔除了上述有疑问的群体并重新分析了数据,结果再次验证了启动材料的可靠性(新的结果与原始结果的对比见下面回应第二条意见)。

意见 2: 附录 3 中的一致性靶子和非一致性靶子是否存在刻板印象偏差? 如何保证其信效度,如政府官员的一致靶词为“官腔、形式主义、官腔官调、喊口号、不办实事、贪污腐败、大腹便便”,这些颇具有污名化的词汇是否会对后续研究有影响?

回应: 非常感谢审稿专家严谨的意见。如本文“2.1.2 实验材料”部分对靶刺激材料甄选过程的描述,本研究基于测量刻板印象的经典方法,并参考以往实证研究测量刻板印象的常用方法,通过 3

个步骤的前测获取不同社会群体的刻板印象特质词：**第一步**，通过开放式问卷(见附录 2)收集每个群体的刻板特质词；**第二步**，对表述不同但意义相近的词语归类并进行频次统计；**第三步**，采用 7 点量表评定前面选出的高频词描述相应社会群体特征的典型程度。指导语要求被试根据特质词描述社会群体的贴切程度进行评分。如：根据你对“科学家”的印象，请对下列词语描述“科学家”的贴切程度在 7 点量表上进行评分, 1-7 表示贴切程度依次加强: 1 代表“非常不贴切”，7 代表“非常贴切”。

通过上述步骤获得 24 个启动群体的刻板印象词语，尽管有些词语的确带有一定的污名化，但这确实是前测 52 名被试的真实刻板印象。正因为是**刻板印象**，其代表了对某一群体的固化的“刻板”认知，其本质上往往具有过度概化、不准确性和以偏概全的特征(Hamilton & Sherman, 1994; Macrae & Bodenhausen, 2001)。就审稿专家所提及的**政府官员**的刻板印象词语，也是通过上述三个步骤获得，而且这些词语在第三步典型程度 7 点量表上的评分都比较高(如下表 C 所示)。因此不论其在效价上是否具有污名化，由于本研究中对刻板印象的测量依照经典方法，可以确保这些词语能够反映被试关于政府官员的刻板印象。

表 C 政府官员刻板印象词语典型程度描述统计结果

	<i>n</i>	<i>Minimum</i>	<i>Maximum</i>	<i>M</i>	<i>SD</i>
官腔	52	2	7	6.0577	1.0740
形式主义	52	2	7	6.0192	1.2444
官腔官调	52	2	7	6.0000	1.1716
喊口号	52	2	7	5.9615	1.1875
大腹便便	52	3	7	5.7500	1.0824
不办实事	52	2	7	5.6154	1.3598
贪污腐败	52	1	7	5.4808	1.3209

此外，本研究的主要探讨的科学问题是：**对于不同热情、能力的社会群体，刻板印象激活效应是否不同？**如前所述，本研究通过预研究获得不同热情、能力的启动群体——在研究方法和统计过程上都有据可依，也通过经典方法甄选每个群体的刻板印象词语，因此可以保证实验材料的有效性，从而也能确保实验结果的可靠性。

鉴于第一位审稿专家的**意见 1**（将“城管”划分为 LW-LC 群体……是否妥当？）和**意见 2**（如政府官员的一致靶词为“官腔、形式主义、官腔官调、喊口号、不办实事、贪污腐败、大腹便便”，这些颇具有污名化的词汇是否会对后续研究有影响）以及**第二位审稿专家对于两个群体的质疑**（“大学教授”“空姐”归为一类，留守儿童属于 HW-LC，实在让人难免质疑的质疑），笔者剔除了上述两位审稿专家所质疑的 4 个群体并重新分析了两个实验的数据。

表 D 启动材料比较: 原文 vs.剔除 4 个群体后

来源		启动社会群体	Trials 个数
原文	HW-HC	军人、消防员、心理咨询师、空姐、瑜伽教练、大学教师	30
	HW-LC	老年人、农民、家庭主妇、农民工、留守儿童、清洁工	30
	LW-HC	商人、海归、政府官员、公务员、富人、演艺明星	30
	LW-LC	罪犯、无业游民、乞丐、吸毒者、恐怖分子、城管	30
剔除 4 个群体后	HW-HC	军人、消防员、心理咨询师、瑜伽教练、大学教师	24
	HW-LC	老年人、农民、家庭主妇、农民工、清洁工	26
	LW-HC	商人、海归、公务员、富人、演艺明星	23
	LW-LC	罪犯、无业游民、乞丐、吸毒者、恐怖分子	26

原启动群体及其删除 4 个群体后的启动群体如表 D 所示, 原实验中每种处理有 30 个 trials, 高热情-高能力(HW-HC)社会群体剔除“空姐”后有 24 个 trials; 高热情-低能力(HW-LC) 社会群体剔除“留守儿童”后有 26 个 trials; 低热情-高能力(LW-HC) 社会群体剔除“政府官员”后有 23 个 trials; 低热情-低能力(HW-LC)社会群体剔除“城管”后有 26 个 trials。由此可见, 每个启动群体剔除一个社会群体后的 trial 数量充分, 因此可以将剔除 4 个群体后的结果与原始结果比较。

与原文中数据分析方法相同, 即对于剔除 4 个群体后的实验 1 和实验 2 的数据, 以反应时和正确率为因变量, 进行 4(社会群体: HW-HC, HW-LC, LW-HC, LW-LC)×2(一致性: 一致 vs. 冲突)的重复测量方差分析, 所有因素均为组内变量。剔除 4 个群体后的结果与原文结果对比如下。

表 E 实验 1 方差分析结果对比: 原文 vs.剔除 4 个群体后

观测指标	变异来源	df	MS	F	p	Partial η^2
原文	社会群体	3	0.136	45.035	.000	0.479
	Error(社会群体)	147	0.003			
	一致性	1	0.031	3.272	.077	0.063
	Error(一致性)	49	0.009			
	社会群体 × 一致性	3	0.048	16.505	.000	0.252
	Error(社会群体×一致性)	147	0.003			
	社会群体	3	151388.811	37.333	.000	0.432
	Error(社会群体)	147	4055.114			
	一致性	1	30800.066	3.106	.084	0.060
	Error(一致性)	49	9917.319			
	社会群体 × 一致性	3	117622.203	29.832	.000	0.378
	Error(社会群体×一致性)	147	3942.784			
剔除 4 个 群体 后	社会群体	3	0.043	14.143	.000	0.224
	Error(社会群体)	147	0.003			
	一致性	1	0.009	1.119	.295	0.022
	Error(一致性)	49	0.008			
	社会群体 × 一致性	3	0.028	8.050	.000	0.141
	Error(社会群体×一致性)	147	0.004			
	社会群体	3	48106.486	10.233	.000	0.173
	Error(社会群体)	147	4701.294			

一致性	1	51715.387	5.388	.024	0.099
Error(一致性)	49	9598.972			
社会群体×一致性	3	101823.940	21.954	.000	0.309
Error(社会群体×一致性)	147	4637.968			

①实验 1 统计分析显示剔除 4 个群体后的结果与原文实验 1 的结果完全一致。详述如下：

首先，正确率和反应时的方差分析结果与原文实验 1 的结果一致。如表 E 所示，不论在正确率还是反应时上，前后数据均显示社会群体与一致性的交互作用极其显著。进一步对交互作用分析结果进行对比，如表 F 所示，简单效应检验的结果在显著性上没有发生任何变化。由此可见，剔除 4 个社会群体后，实验 1 的基本结果趋势及其显著性没有受到影响(如下图 1 所示)。

表 F 实验 1 的正确率和反应时 (*M*; *SE*) 及其一致性的简单效应检验结果对比：原文 vs. 剔除 4 个群体后

观测指标	启动	启动-靶子的一致性		Mean Difference (I-C)	F	p	Partial η^2	
	社会群体	一致(C)	冲突(I)					
原文	正确率	HW-HC	0.971 (0.006)	0.944 (0.007)	-0.027	10.08	.003	0.171
		HW-LC	0.938 (0.007)	0.935 (0.010)	-0.003	0.07	.788	0.001
		LW-HC	0.877 (0.012)	0.903 (0.010)	0.025	2.21	.143	0.043
		LW-LC	0.843 (0.010)	0.918 (0.012)	0.075	21.3	.000	0.303
	反应时(ms)	HW-HC	774 (23)	851 (26)	77	27.73	.000	0.361
		HW-LC	839 (24)	875 (28)	36	5.64	.022	0.103
		LW-HC	872 (24)	911(26)	39	6.35	.015	0.115
		LW-LC	937 (29)	856 (26)	-81	34.75	.000	0.415
剔除 4 个群体后	正确率	HW-HC	0.968 (0.007)	0.938 (0.007)	-0.030	9.33	.004	0.16
		HW-LC	0.935 (0.008)	0.932 (0.010)	-0.004	0.12	.729	0.002
		LW-HC	0.902 (0.009)	0.926 (0.012)	0.024	2.24	.124	0.048
		LW-LC	0.882 (0.011)	0.930 (0.013)	0.048	7.86	.007	0.138
	反应时(ms)	HW-HC	796 (26)	873 (28)	77	23.34	.000	0.323
		HW-LC	838 (24)	881 (27)	43	7.96	.007	0.103
		LW-HC	848 (23)	888(25)	40	6.76	.012	0.121
		LW-LC	922 (29)	852 (27)	-69	23.17	.000	0.321

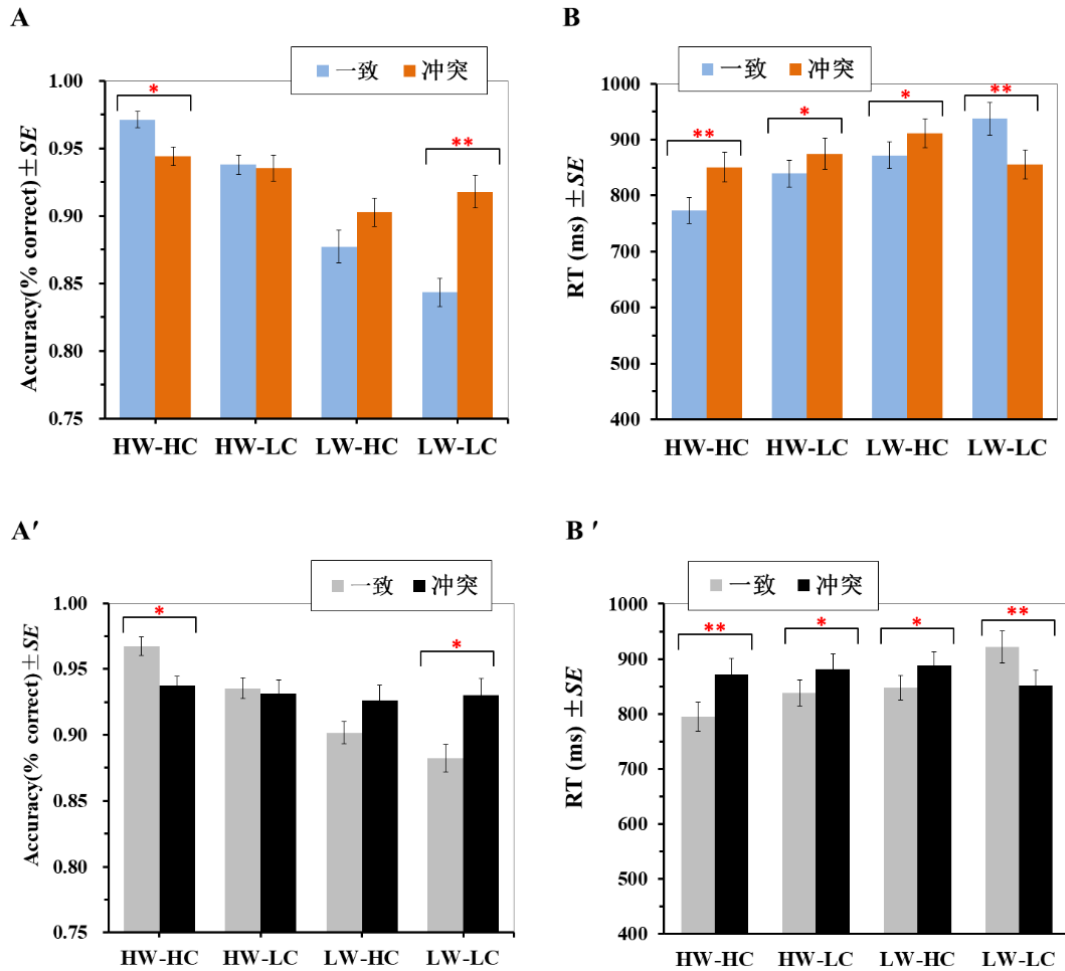


图 1 实验 1 结果对比图：原文正确率(A)和反应时(B) vs. 剔除 4 个群体后的正确率(A')和反应时(B')

②实验 2 统计分析结果显示剔除 4 个群体后的结果与原文实验 1 的结果完全一致。详述如下：

首先，正确率和反应时的方差分析结果与原文实验 1 的结果一致，如表 G 所示，不论在正确率还是反应时上，前后数据均显示社会群体与一致性的交互作用显著。

表 G 实验 2 方差分析结果对比：原文 vs. 剔除 4 个群体后

观测指标	变异来源	df	MS	F	p	Partial η^2
原文	社会群体	3	0.01	3.53	.017	0.07
	Error(社会群体)	135	0.00			
	一致性	1	0.03	5.67	.022	0.11
	Error(一致性)	45	0.01			
	社会群体 × 一致性	3	0.02	12.35	.000	0.22
	Error(社会群体 × 一致性)	135	0.00			
	社会群体	3	3445.28	2.91	.040	0.06
	Error(社会群体)	135	1184.70			
	一致性	1	13599.79	11.87	.000	0.21
	Error(一致性)	45	1146.04			
剔除 4 个群体后	社会群体 × 一致性	3	51092.61	34.77	.000	0.44
	Error(社会群体 × 一致性)	135	1469.37			
	社会群体	3	0.01	2.95	.035	0.06
	Error(社会群体)	135	0.00			
	正确率					

反应时	一致性	1	0.04	6.51	.014	0.13
	Error(一致性)	45	0.01			
	社会群体 × 一致性	3	0.02	11.31	.000	0.20
	Error(社会群体×一致性)	135	0.00			
	社会群体	3	2268.80	1.25	.300	0.03
	Error(社会群体)	135	1821.05			
	一致性	1	32171.84	22.21	.000	0.33
	Error(一致性)	45	1448.46			
	社会群体 × 一致性	3	49558.04	25.66	.000	0.36
	Error(社会群体×一致性)	135	1931.43			

表 H 实验 2 的正确率和反应时 (M ; SE) 及其一致性的简单效应检验结果对比：原文 vs. 剔除 4 个群体后

	启动 社会群体	启动-靶子的一致性		Mean Difference (I-C)	F	p	Partial η^2
		一致(C)	冲突(I)				
原文	HW-HC	0.967 (0.010)	0.914 (0.015)	-0.052	18.16	.000	0.287
	HW-LC	0.953 (0.011)	0.927 (0.014)	-0.026	5.33	.026	0.106
	LW-HC	0.931 (0.011)	0.914 (0.016)	-0.017	3.01	.090	0.063
	LW-LC	0.928 (0.009)	0.952 (0.010)	0.024	6.29	.016	0.123
	HW-HC	582 (18)	638 (19)	56	57.11	.000	0.559
	HW-LC	588 (17)	618 (18)	30	13.99	.001	0.237
	LW-HC	607 (17)	625 (18)	18	4.98	.031	0.100
	LW-LC	631 (18)	576 (16)	-55	48.47	.000	0.519
剔除 4 个群体后	HW-HC	0.967 (0.010)	0.915 (0.015)	-0.053	19.96	.000	0.307
	HW-LC	0.950 (0.011)	0.923 (0.014)	-0.027	5.45	.024	0.108
	LW-HC	0.934 (0.013)	0.909 (0.016)	-0.025	4.96	.031	0.099
	LW-LC	0.930 (0.010)	0.952 (0.011)	0.023	4.17	.047	0.085
	HW-HC	586 (20)	638 (21)	52	41.69	.000	0.481
	HW-LC	588 (16)	622 (18)	34	17.23	.000	0.277
	LW-HC	593 (18)	632 (19)	39	15.28	.000	0.254
	LW-LC	628 (17)	578 (16)	-50	30.15	.000	0.401

进一步对交互作用分析的结果进行对比，如表 H 所示，除了正确率指标上，剔除 4 个社会群体后在 LW-HC 群体水平，一致靶子的正确率显著高于冲突靶子的($p = .031$)，原文中一致靶子的正确率也高于冲突靶子的，不过未达到统计上的显著($p = .090$)，其它简单效应检验的结果在显著性上没有发生任何变化。由此可见，剔除 4 个社会群体后，实验 2 的基本结果趋势及其显著性在反应时上没有受到任何影响，在正确率上也几乎没有受到影响（如下图 2 所示）。

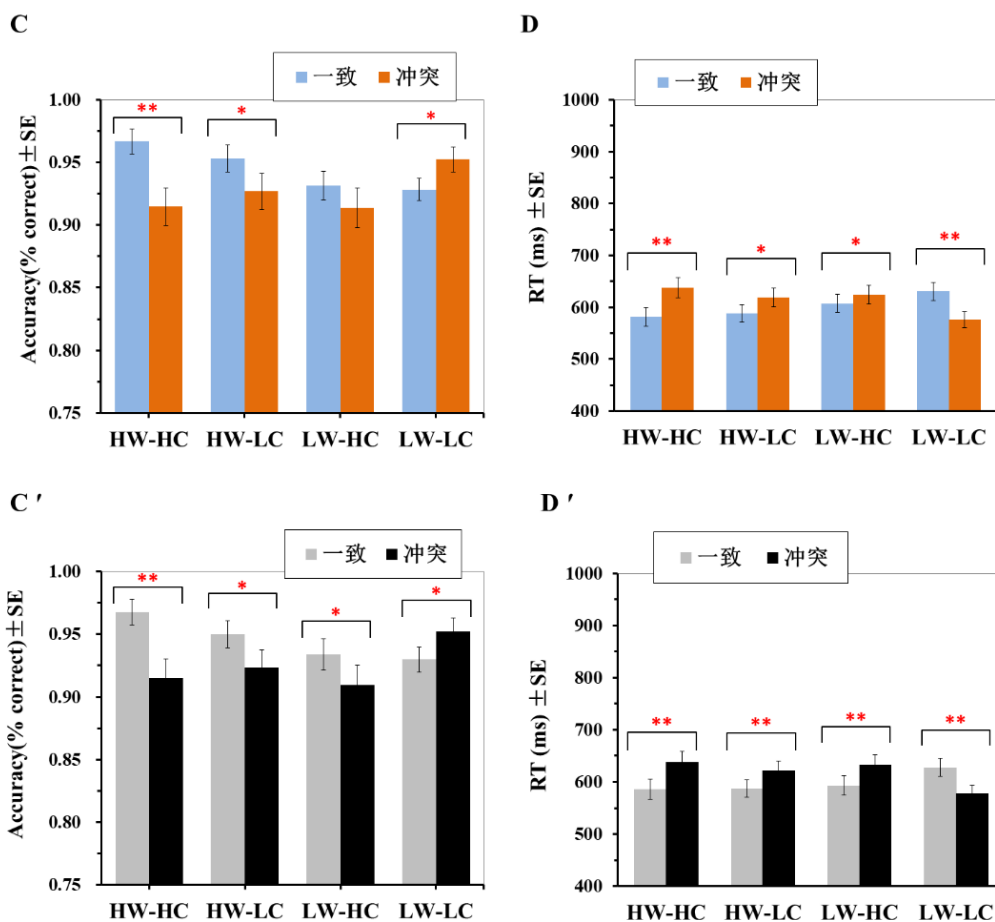


图2 实验2结果对比图：原文正确率(C)和反应时(D) vs.剔除4个群体后的正确率(C')和反应时(D')

综上所述, 通过将原始结果与剔除4个群体后的结果一一对比发现, 不论是实验1外显任务还是实验2内隐任务, 结果趋势没有受到任何影响。这反映了本研究的结果是可靠和稳定的; 从另一侧面也说明了这几个群体划分也是合适的。

意见3: 另外, 2.1.2 中低热情-低能力群体, 作者标注为 (HW-LC) 具有明显书写错误, 也建议修改。同时, 建议作者文中尽量不使用“我们”、“笔者”等词汇。

回应: 感谢审稿专家细致的建议! 我们按照您的要求修改了对应的书写错误, 并仔细对全文中的其他部分也进行了核查。文中“我们”、“笔者”这些词汇也尽数删去。

审稿人2意见:

意见1: 论文经过数次修改后, 写作比较规范, 研究的逻辑比较清晰, 但依然存在问题与作者商榷。研究基于刻板印象内容模型 (SCM) 探讨我国不同热情-能力四类社会群体刻板印象激活效应的行为模式。佐斌, 代涛涛, 温芳芳, & 索玉贤 (2015) 指出, 虽然 SCM 具有普适性, 研究者对其内涵本质的认识是一致的, 但在具体命名上存在文化和地区差异, 应当根据本土实际加以确定。中

国学者一直沿用 Fiske 提出的热情和能力命名,但事实上,中国人自有中国人的生存哲学,在对人评价时未必是这两个维度的命名和内容,实际上是道德和才能两个。因此,中国的模型必须建立在本土文化的实际上,采用更加合理的维度划分标准和命名。在实证研究中,管健,程婕婷(2011)的研究发现 SCM 对中国群体有很好的预测作用,也有学者提出在修正的 SCM 模型(高明华,2010)。可见,这一模型命名划分的在我国的适用性还在探讨阶段,如果采用这一划分得出本研究的分类,比如大学教授和空姐为一类(HW-HC),留守儿童属于 HW-LC,实在让人难免质疑是否采用了最重要的划分维度呢?最关键的问题是刻板印象特质词一般是形容词,研究中选择的特质词并未很好的控制,比如家庭主妇的特质词是做饭带孩子,罪犯的特质词是监狱,这显然都不恰当,不仅词性不同,做饭带孩子和监狱也不能用于描述家庭主妇和罪犯的特质,如果描述家庭主妇的词汇“做饭带孩子”改为贤惠的可能更加恰当,诸如此类的特质词还有不少。特质词如果没有控制好显然会影响反应时和正确率。此外,某些论述显得过于繁复冗长,可以更加精炼,比如问题提出部分的论述:

“近期有研究者采用控制更为严格的实验研究试图从实证层面对该模型进行更加多样化的探索,相关研究也蓬勃发展。例如,王沛等人(2015)探讨刻板印象内容对刻板印象抑制反弹效应的影响。张珊明等人(2015)通过内隐测量手段发现人们对高权力者抱有高能力-低热情的刻板印象,而对低权力者则正相反。王美芳等人(2017)进一步探讨了权力刻板印象的热情和能力两个维度及其对自发特质推理的影响(Wang & Yang, 2017; Yang & Wang, 2016)。江红艳等人(2016)探讨了刻板印象内容模型在营销领域中的应用。宋静静等人(2017)则从交叉分类的角度来探讨贫富和年龄对热情和能力行为的评价和归因的影响。韦庆旺等人(2018)则从维度之争的角度出发,探讨了社会阶层对热情优先效应的调节作用。佐斌等人(2018)探讨了评价意图与结果在群际评价时对热情和能力关系的情境性影响。”鉴于论文存在实质性问题,建议大修。

回应: 非常感谢审稿专家严谨的意见和建议! 我们将上述意见根据其内容分成三个方面, 逐一回应如下:

(1) 研究基于刻板印象内容模型(SCM)探讨我国不同热情-能力四类社会群体刻板印象激活效应的行为模式。佐斌, 代涛涛, 温芳芳, & 索玉贤(2015)指出, 虽然 SCM 具有普适性, 研究者对其内涵本质的认识是一致的, 但在具体命名上存在文化和地区差异, 应当根据本土实际加以确定。中国学者一直沿用 Fiske 提出的热情和能力命名, 但事实上, 中国人自有中国人的生存哲学, 在对人评价时未必是这两个维度的命名和内容, 实际上是道德和才能两个。因此, 中国的模型必须建立在本土文化的实际上, 采用更加合理的维度划分标准和命名。在实证研究中, 管健, 程婕婷(2011)的研究发现 SCM 对中国群体有很好的预测作用, 也有学者提出在修正的 SCM 模型(高明华, 2010)。可见, 这一模型命名划分的在我国的适用性还在探讨阶段, 如果采用这一划分得出本研究的分类,

比如大学教授和空姐为一类（HW-HC），留守儿童属于 HW-LC，实在让人难免质疑是否采用了最重要的划分维度呢？

回应：感谢审稿人中肯的建议。诚如您所言，SCM 在中国本土的适用性确实可能需要根据国内的本土情况进行一定的调整，然而，就目前国内对于 SCM 的研究现状而言，由于适用性的研究还在进行，根据目前已有文献来看，并不能得出相对于“热情”和“能力”更为恰当的命名方式。虽然近年来，有学者对 SCM 的维度区分有异议，认为应该把道德作为一个单独维度独立出来，然而相关的维度之争还在激烈进行，也并未得到一致认同。至于在中国的研究背景下是否可以用“道德”和“才能”来表示“warmth”和“competence”，我们认为可能还值得商榷。原因如下：

首先，在理论和实际应用上，这一命名方式并未得到研究者们的普遍认同和采用。近年来的维度之争中也只是提出要把“道德”作为一个单独维度从“warmth”维度中独立出来，以便和“warmth”维度进一步区分，如果直接把“道德”等同于“warmth”，可能与目前 SCM 的大部分研究结论都有所背离。我们整理了国内关于 SCM 的实证研究对于维度的命名方式（详见附表 5），结果发现，不同的研究团队使用的命名方式具有一定差异。大部分的实证研究还是使用“热情”和“能力”的命名方式，也有一部分研究使用“温情”/“温暖”和“能力”，还有一部分研究响应国外“道德”维度的研究，把“道德”也作为一个单独的维度进行研究。

第二，在现有的实证研究中绝大多数都是采用“热情”和“能力”这一命名方式。例如，《心理学报》2018 年发表的两篇涉及 SCM 的实证研究，采用的也是“热情”和“能力”的命名方式。虽然如专家所言，佐斌教授团队曾经提出“**虽然 SCM 具有普适性，研究者对其内涵本质的认识是一致的，但在具体命名上存在文化和地区差异，应当根据本土实际加以确定**”，然而在该团队的一系列实证研究中，他们依然采用“热情”和“能力”这一命名方式。佐斌教授团队在 SCM 理论的国内引进以及相关的实证研究方面做了大量工作，因此，他们的选择对于国内 SCM 的相关研究具有重要的指导和借鉴意义。此外，王美芳教授，钟毅平教授等团队也采用相同的命名方式也体现了这种命名方式的合理性。在梳理了国际上 SCM 跨文化的相关研究后，我们发现，除了早期在香港的研究之外，大陆方面的研究最近也有所发表(Wu, Bai, & Fiske, 2018)，他们依然采用这两个维度的命名方式。

总之，考虑到本研究的重点不是对相关理论和概念进行辨析，而是对理论的应用，同时为了使中文表达更容易理解，能够与国际上的研究进行对比，并且与国内的研究传统保持一致，本研究采用的是“热情”和“能力”这一维度命名方式。在修改版的正文中，我们也在问题提出的第四段对 SCM 进行介绍时加了相应的尾注进行了说明。具体如下：

“¹ 国内研究者也有翻译成“温情”/“温暖”和“能力”（有关国内 SCM 实证研究的维度命名总

结详见附录 5)，也有研究者建议把“道德”作为单独的维度分离出来，考虑到 SCM 维度之争目前并未有一致性的结论，以及本研究的重点不是对相关理论和概念进行辨析，而是对理论的应用，同时为了使中文表达更容易理解并且和国内大部分研究保持一致，本研究选用了热情和能力这种概念表述。对 SCM 维度之争有兴趣的读者，建议进一步阅读相关文献。”

附表 5 国内刻板印象内容模型实证研究维度命名总结

文献	维度命名	备注	文献	维度命名	备注	文献	维度命名	备注
(佐斌, 温芳芳, 吴漾, & 代涛涛, 2018)	热情 能力		(白丽娟, 2018)	热情 能力	学位论文 导师: 陈莉	(高明华, 2010)	热情 才能 道德	
(代涛涛, 佐斌, & 郭敏仪, 2018)	热情 能力		(刘志军, 2017)	热情 能力	学位论文 导师: 伍麟	(魏华, 段海岑, 周宗奎, 朱晓伟, & 刘美婷, 2018)	温暖 能力	
(宋静静, 佐斌, 温芳芳, 谭潇, & 赵苗苗, 2017)	热情 能力		(何睿立, 2017)	热情 能力	学位论文 导师: 梁斌	(魏华, 汪涛, 周宗奎, 冯文婷, & 丁倩, 2016)	温暖 能力	
(Song & Zuo, 2016)	热情 能力		(张伟平, 2017)	热情 能力	学位论文 导师: 赵德雷	(王汉瑛, 邢红卫, & 田虹, 2018)	温暖 能力	
(刘珂, 2018)	热情 能力	学位论文 导师: 佐斌	(方家琪, 2017)	热情 能力	学位论文 导师: 曲可佳	(穆雪茗, 白长虹, 吴波, & 王红玉, 2017)	温暖 能力	
(刘国琪, 2018)	热情 能力	学位论文 导师: 佐斌	(邓敏林, 2017)	热情 能力	学位论文 导师: 刘超	(徐朝娜, 赵玉芳, & 毕重增, 2017)	温暖 能力	
(李焱焱, 2018)	热情 能力	学位论文 导师: 佐斌	(郑倩倩, 2016)	热情 能力	学位论文 导师: 王海忠	(周飞 & 沙振权, 2017)	温暖 能力	
(张心一, 2018)	热情 能力	学位论文 导师: 佐斌	(杨光玉, 2015)	热情 能力	学位论文 导师: 王海忠	(赵静, 2017)	温暖 能力	学位论文 导师: 史有春
(张金良, 2017)	热情 能力	学位论文 导师: 佐斌	(刘婷婷, 2015)	热情 能力	学位论文 导师: 邢采	(张莉, 2013)	温暖 能力	学位论文 导师: 何佳讯
(蔡馥, 2016)	热情 能力	学位论文 导师: 佐斌	(刘嘉秋, 2014)	热情 能力	学位论文 导师: 刘春玲	(江红艳, 王海忠, 何云, & 朱力, 2016)	温暖 能力	
(王萌, 2016)	热情 能力	学位论文 导师: 佐斌	(翟成蹊, 2011)	热情 能力	学位论文 导师: 李岩梅	(江红艳 & 孙配贞, 2014)	温情 能力	
(吴月鹏, 2016)	热情 能力	学位论文 导师: 佐斌	(包蕾萍, 2010)	热情 能力	学位论文 导师: 桑标	(江红艳, 王海忠, & 钟科, 2014)	温情 能力	
(张喆颖, 2016)	热情 能力	学位论文 导师: 佐斌	(苑春永, 2009)	热情 能力	学位论文 导师: 张智勇	(朱力, 2016)	温情 能力	学位论文 导师: 何云
(王咏霞, 2015)	热情 能力	学位论文 导师: 佐斌	(王震, 2008)	热情 能力	学位论文 导师: 陈志霞	(钟科 & 何云, 2018)	能力 温情	
(戴月娥, 2018)	热情 能力	学位论文 导师: 温芳芳	(程婕婷, 2011)		学位论文 导师: 管健		能力	
(刘雪莹, 2018)	热情 能力	学位论文 导师: 温芳芳	(管健 & 程婕婷, 2011)	热情 能力		(冯文婷, 彭泗清, & 涂荣庭, 2018)	温情 能力	
(游之悦, 2018)	热情 能力	学位论文 导师: 温芳芳	(程婕婷, 管健, & 汪新建, 2012)	热情 能力		(张一, 张立英, & 仇荒宇, 2017)	温情 能力	
(韦庆旺, 李木子, & 陈晓晨, 2018)	热情 能力		(程婕婷, 2015)	热情 能力		(陈志婷, 2018)	温情 能力	学位论文 导师: 李峰
						(王佳, 2018)	温情	学位论文

(石长慧 & 王卓妮, 2017)	热情 能力		(程婕婷, 张斌, & 汪新建, 2015)	社交性 能力 道德		(王沛, 吴薇, 谢轶雯, & 徐琪, 2015)	能力 温情 能力	导师: 董大海
(王美芳, 史文凤, 颜丙淦, & 杨峰, 2017)	热情 能力		(程婕婷, 张斌, 汪新建, & 管健, 2016)	社交性 能力 道德		(谢轶雯, 2014)	温情 能力	学位论文 导师: 王沛
(姜新华, 杨峰, 李秀丽, & 王美芳, 2016)	热情 能力		(杨金花 & 金盛华, 2017)	热情 能力 道德		(徐钟庚, 席贻莹, 马婷, & 张锋, 2016)	热情度 胜任度	
(颜丙淦, 2016)	热情 能力	学位论文 导师: 王美芳	(杨金花, 2012)	热情 能力 道德	学位论文 导师: 金盛华	(席贻莹, 2014)	热情度 胜任度	学位论文 导师: 徐钟庚
(张庆 & 王美芳, 2014)	热情 能力		(吕小康 & 刘颖, 2018)	热情 能力 道德		(江筮豪, 2013)	亲和力 能力	学位论文 导师: 张惠真
(王美芳, 杨峰, & 杨云云, 2013)	热情 能力		(周春燕, 黄海, 刘陈陵, 郭永玉, & 贺金波, 2015)	热情 能力 道德		(陈晶晶, 2013)	亲和力 能力	学位论文 导师: 郭朝阳
(张庆, 2012)	热情 能力	学位论文 导师: 王美芳	(代涛涛, 2015)	热情 能力 道德	学位论文 导师: 佐斌	(毕圣, 2013)	温暖的 有能力的	学位论文 导师: 吴冠之
(张珊明, 钟毅平, & 罗伏生, 2015)	热情 能力		(高明华, 2010)	热情 才能 道德		(赵卫星 & 郑希付, 2016)	热情 能力 道德 / 价值观	
(张珊明, 钟毅平, 肖丽辉, & 罗伏生, 2015)	热情 能力					(苏珊, 2016)	热情 能力 企业社会责任	学位论文 导师: 李宏
(张珊明, 罗伏生, 钟毅平, & 陈芸, 2013)	热情 能力							
(陶塑, 许燕, & 袁灿灿, 2015)	热情 能力							
(周迎楠, 2018)	热情 能力	学位论文 导师: 毕重增				(郑健 & 刘力, 2012)	SCM 不具有普适性	

参考文献:

Chen, Z., & Fiske, S. (2008). Stereotype content model in China. Unpublished raw data.

Cuddy, A. J., Fiske, S. T., Kwan, V. S., Glick, P., Demoulin, S., & Leyens, J. P., et al. (2009). Stereotype content model across cultures: Towards universal similarities and some differences. *British Journal of Social Psychology*, 48(1), 1–33.

Durante, F., Fiske, S. T., Kervyn, N., Cuddy, A. J. C., Akande, A. D., Adetoun, B. E., . . . Storari, C. C.

(2013). Nations' income inequality predicts ambivalence in stereotype content: How societies mind the gap. *British Journal of Social Psychology*, 52, 726-746.

Fiske, S. T., Cuddy, A. J., Glick, P., & 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(6), 878–902.

Wu, S. J., Bai, X., & Fiske, S. T. (2018). Admired rich or resented rich? How two cultures vary in envy. *Journal of*

Fiske, S. T. (2018). Stereotype content: Warmth and competence endure. *Current Directions in Psychological Science*, 27(2), 67–73.

意见 2：最关键的问题是刻板印象特质词一般是形容词，研究中选择的特质词并未很好的控制，比如家庭主妇的特质词是做饭带孩子，罪犯的特质词是监狱，这显然都不恰当，不仅词性不同，做饭带孩子和监狱也不能用于描述家庭主妇和罪犯的特质，如果描述家庭主妇的词汇“做饭带孩子”改为贤惠的可能更加恰当，诸如此类的特质词还有不少。特质词如果没有控制好显然会影响反应时和正确率。

回应：非常感谢审稿专家严谨的意见！诚如专家所言——刻板印象特质词一般是形容词，然而，在国内外的实证研究中，对于社会群体刻板印象的测量和研究往往涉及名词或其它词性的词语。例如，在 White 等人(2009)探讨性别刻板印象激活效应的 ERP 研究中，有一半的靶子词是形容词，另一半的靶子词是名词(如与“女性”相关的刻板印象名词有：化妆品、洋娃娃、秘书；与“男性”相关的刻板印象名词有：修理工、工程师、雪茄)。参见 White 等人(2009)的论文 P.193——The target word was either a trait (e.g. nurturing, aggressive) or nontrait (e.g. makeup, mechanic) stereotypically associated with one of the gender categories. Target words included 14 female traits (e.g. caring, gossipy), 14 male traits (e.g. aggressive, powerful), 14 female non-traits (e.g. secretary, doll) and 14 male non-traits (e.g. engineer, cigars) chosen from previous gender stereotype research (Bem, 1974; Blair and Banaji, 1996; Oakhill et al., 2005)。

同样，在最近一项关于性别刻板印象激活与语义激活区别的 ERP 研究中(Pei Wang, Ya-Ping Yang, Chen-Hao Tan, Qing-Wei Chen & Thomas Cantfort. (2017). Gender Stereotype Activation versus Lexical Semantic Activation: An ERP Study. *The Journal of General Psychology*, 144(4): 283-308.)，其中一半性别刻板印象词语是形容词，一半是名词。

由此可见，尽管研究者们习惯上将某一群体的刻板印象词语称为刻板特质词，但实质上这些词语不仅仅包括形容词，往往也包括名词或其它词性的词语。其实，这不论从刻板印象的定义还是在实际生活中，都是讲的通的。从定义上来看，刻板印象是一种包含了知觉者关于某个人类群体的知识、观念与预期的认知结构；是联结某个社会群体与一系列特质或行为特征的抽象的知识结构 (Hamilton & Sherman, 1994; Macrae & Bodenhausen, 2001)。在这一认知结构中，包含着人们对某一群体的固定看法和观点，尽管其大多以形容词的形式表征，但也不排除其它词性的词语。因此，刻板印象词语在词性上是否一致并非刻板印象研究的关键，最本质的是这个词语是否能反映人们对某一

群体的固有认知。从实际生活中来看，例如对“家庭主妇”这个群体，根据本研究的前测发现(如下表 I 所示)，人们想到的特征是首先是“做饭带孩子”。具体来说，对“家庭主妇”的刻板印象词语典型性评定结果显示，“做饭带孩子”的典型程度最高，其次是“顾家”、“做家务”、“贤惠”和“勤劳”，在本研究中用了这 5 个典型的刻板印象词语作为“家庭主妇”启动后的一致靶子刺激。同样，前测结果显示对“罪犯”的刻板印象词语最典型的是“监狱”，其次是“可恶”、“凶狠”、“残暴”(如下表 I 所示)。

综上所述，本研究中的靶子词包括形容词、名词等，尽管词性不同——根据以往研究，词性并非是影响自变量效应且需要严格控制的无关变量，最关键的是这些靶子词是否典型。在本研究中，基于前人的研究方法和步骤，通过前测甄选出每个社会群体的典型的刻板印象词语。然而，原文中表述“刻板特质词”可能会让读者认为靶子词都是形容词。非常感谢专家的问题和意见，笔者与合作者讨论，把文中“刻板特质词”改为“刻板印象词语”（在修改后的文章中以蓝色字体呈现），这样可能会更加贴切和妥当。

表 I 家庭主妇、罪犯刻板印象词语典型程度描述统计结果

家庭主妇	<i>n</i>	<i>Minimum</i>	<i>Maximum</i>	<i>M</i>	<i>SD</i>	罪犯	<i>n</i>	<i>Minimum</i>	<i>Maximum</i>	<i>M</i>	<i>SD</i>
做饭带孩子	52	2	7	5.88	1.28	监狱	52	2	7	5.71	1.50
顾家	52	2	7	5.87	1.17	可恶	52	2	7	5.45	1.52
做家务	52	2	7	5.84	1.18	凶狠	52	2	7	5.42	1.53
贤惠	52	2	7	5.71	1.25	残暴	52	2	7	5.34	1.50
勤劳	52	2	7	5.53	1.18						
默默奉献	52	1	7	5.22	1.48						
生活单调	52	1	7	5.06	1.54						
生活圈狭窄	52	1	7	4.96	1.62						
容易亲近	52	1	7	4.96	1.39						
唠叨	52	1	7	4.94	1.29						
任劳任怨	52	1	7	4.90	1.33						
不注重外表	52	1	7	4.04	1.52						

意见 3：此外，某些论述显得过于繁复冗长，可以更加精炼，比如问题提出部分的论述：“近期有研究者采用控制更为严格的实验研究试图从实证层面对该模型进行更加多样化的探索，相关研究也蓬勃发展。例如，王沛等人(2015)探讨刻板印象内容对刻板印象抑制反弹效应的影响。张珊明等人(2015)通过内隐测量手段发现人们对高权力者抱有高能力-低热情的刻板印象，而对低权力者则正相反。王美芳等人(2017)进一步探讨了权力刻板印象的热情和能力两个维度及其对自发特质推理的影响(Wang & Yang, 2017; Yang & Wang, 2016)。江红艳等人(2016)探讨了刻板印象内容模型在营销领域中的应用。宋静静等人(2017)则从交叉分类的角度来探讨贫富和年龄对热情和能力行为的评价和归因的

影响。韦庆旺等人(2018)则从维度之争的角度出发,探讨了社会阶层对热情优先效应的调节作用。佐斌等人(2018)探讨了评价意图与结果在群际评价时对热情和能力关系的情境性影响。”鉴于论文存在实质性问题,建议大修。

回应: 非常感谢审稿专家的建议!我们在对全文进行了重新审视,并且对文献进行重新梳理和查新之后,对相应的段落进行了重新改写,以期更加全面精炼地反映国内研究现状。具体如下:

“国内学者对于 SCM 的理论发展、模型假设、维度之争、认知神经科学研究和营销领域中的应用等方面均进行了较为细致的总结和综述(代涛涛, 佐斌, & 温芳芳, 2014; 管健, 2009; 汪新建, 程婕婷, & 管健, 2014; 吴波 & 李东进, 2013; 佐斌, 代涛涛, 温芳芳, & 索玉贤, 2015; 佐斌, 代涛涛, 温芳芳, & 滕婷婷, 2014; 佐斌, 张阳阳, 赵菊, & 王娟, 2006), 理论验证方面也进行了一定的尝试和探索(程婕婷, 2015; 程婕婷, 管健, & 汪新建, 2012; 程婕婷, 张斌, & 汪新建, 2015; 高明华, 2010; 管健 & 程婕婷, 2011; 石长慧 & 王卓妮, 2017; 郑健 & 刘力, 2012), 然而, 这些研究仅仅从理论维度上对该模型进行一定的本土化验证和修订(汪新建 & 程婕婷, 2015)。而在国际上, Fiske 团队自 SCM 提出以来, 就对中国有所关注, 除了 2008 年未发表的研究之外(Chen & Fiske, 2008), 他们分别于 2002 年和 2009 年发表了香港地区的 SCM 调查结果(Cuddy et al., 2009; Fiske, Cuddy, Glick, & Xu, 2002), 并且新近发表了中国大陆地区的 SCM 调查结果(Wu, Bai, & Fiske, 2018)。近期国内研究者更是对 SCM 进行多样化的实证探索, 取得不少引人注目的结果。如热情与能力之间的关系(韦庆旺, 李木子, & 陈晓晨, 2018; 佐斌, 温芳芳, 吴漾, & 代涛涛, 2018), SCM 在营销领域中的应用(冯文婷, 彭泗清, & 涂荣庭, 2018; 江红艳 & 孙配贞, 2014; 江红艳, 王海忠, 何云, & 朱力, 2016; 穆雪茗, 白长虹, 吴波, & 王红玉, 2017; 王汉瑛, 邢红卫, & 田虹, 2018; 魏华, 汪涛, 周宗奎, 冯文婷, & 丁倩, 2016; 钟科 & 何云, 2018), 王美芳团队从发展角度探讨能力和热情的年龄变化(姜新华, 杨峰, 李秀丽, & 王美芳, 2016; 王美芳, 杨峰, & 杨云云, 2013; 张庆 & 王美芳, 2014), 权力刻板印象(王美芳, 史文凤, 颜丙淦, & 杨峰, 2017; 张珊明, 罗伏生, 钟毅平, & 陈芸, 2013; 张珊明, 钟毅平, 肖丽辉, & 罗伏生, 2015; 张珊明, 钟毅平, & 罗伏生, 2015)和阶层刻板印象(周春燕, 黄海, 刘陈陵, 郭永玉, & 贺金波, 2015), SCM 在特定群体中的应用, 如医患群体(吕小康 & 刘颖, 2018)、残疾人群体(徐钟庚, 席贻莹, 马婷, & 张锋, 2016)、贫富群体(Song & Zuo, 2016; 宋静静, 佐斌, 温芳芳, 谭潇, & 赵苗苗, 2017; 陶塑, 许燕, & 袁灿灿, 2015; 杨金花 & 金盛华, 2017), 刻板印象内容对刻板印象抑制反弹效应的影响(王沛, 吴薇, 谢轶雯, & 徐琪, 2015), 中介和调节变量的探索(代涛涛, 佐斌, & 郭敏仪, 2018; 周飞 & 沙振权, 2017)。”

此外, 我们还在文中问题提出的第四段添加了最新的文献, 以及时更新 SCM 的研究动态, 具体如下:

“该模型具有跨文化的一致性和通用性(Cuddy et al., 2009; Asbrock, 2010; Bye, Herrebrøden, Hjetland, Røyset, & Westby, 2014; Durante, Fiske, Gelfand, Crippa, Suttora, & Stillwell, et al., 2017; Fiske, 2018)”

最后，再次衷心感谢两位审稿专家花费宝贵的时间和精力阅读和评审本文，为本文的修改完善提出了严谨、重要的意见和建议！

第四轮

审稿人 1 意见：

意见 1：文章在之前的修改基础上进行了认真的完善工作，有了较大的进步。一些细小的地方建议作者及其团队加以关注。第一，文章问题提出部分，新修改的蓝色部分增加了近千字，其中关于国内学者对于 SCM 的理论发展、模型假设、维度之争、认知神经科学研究和营销领域中的应用等方面进行了较为细致的总结和综述，罗列了参考文献达 43 个之多，有点过于冗长和繁琐。第二，相应的参考文献整体达 97 篇之多，对于一篇论文而言过多，建议精准。第三，文章整体部分在修改和完善后达到 2 万余字，也建议更加精准讨论。

回应：非常感谢审稿专家的细致建议！文章修改部分用蓝色字体标出，回应如下：

第一，文章新修改的蓝色部分确实有点冗长和繁琐，主要是为了细致总结国内的 SCM 研究现状，为了更加精简，我们对该部分进行了重新改写，对该部分进行了精简处理，最后精简完之后的部分文字总数从原本的 891 减少为 412，参考文献从 43 个减少为 17 个。修改后的部分为：

“国内学者对于 SCM 的理论发展、模型假设、维度之争、认知神经科学研究和营销领域中的应用等方面均进行了较为细致的总结和综述(代涛涛, 佐斌, & 温芳芳, 2014; 管健, 2009; 汪新建, 程婕婷, & 管健, 2014; 吴波 & 李东进, 2013; 佐斌, 代涛涛, 温芳芳, & 索玉贤, 2015; 佐斌, 代涛涛, 温芳芳, & 滕婷婷, 2014; 佐斌, 张阳阳, 赵菊, & 王娟, 2006), 理论验证方面也进行了一定的尝试和探索(高明华, 2010; 管健 & 程婕婷, 2011; 石长慧 & 王卓妮, 2017), 然而, 这些研究仅仅从理论维度上对该模型进行一定的本土化验证和修订(汪新建 & 程婕婷, 2015)。而在国际上, Fiske 团队自 SCM 提出以来, 就对中国有所关注, 除了 2008 年未发表的研究之外(Chen & Fiske, 2008), 他们分别于 2002 年和 2009 年发表了香港地区的 SCM 调查结果(Cuddy et al., 2009; Fiske, Cuddy, Glick, & Xu, 2002), 并且最近发表了中国大陆地区的 SCM 调查结果(Wu, Bai, & Fiske, 2018)。近期国内研究者更是对 SCM 进行多样化的实证探索, 取得不少引人注目的结果。如热情与能力之间的关系(韦庆旺, 李

木子, & 陈晓晨, 2018; 佐斌, 温芳芳, 吴漾, & 代涛涛, 2018)等。”

第二, 我们对整篇文章的参考文献进行了精简处理之后, 参考文献从 97 篇减少到 64 篇。

第三, 在对整篇文章进行细致修改和完善之后, 修改完的版本最后字数从原稿的 20579 字减少为 16629 字。具体修改为 (1) 将靶子词的前测部分放入附录 2 中; (2) 对实验流程的描述进行了精简; (3) 效果量的数字统一为小数点后两位; (4) 笔者及其合作者仔细通读全文, 也删减并完善了表述冗长和不够精简之处。

.....

审稿人 2 意见:

意见 1: 作者对一审中所提问题进行了认真的回复, 但其中关键问题并未得到有效解决。论文中将刻板印象特质词改为刻板印象词汇, 词汇包含的词性可以超出形容词, 但前提都是要有控制, 予以平衡, 正如作者举例, White 等人(2009)探讨性别刻板印象激活效应的 ERP 研究中, 有一半的靶子词是形容词, 另一半的靶子词是名词, 研究中对词汇的词性有严格的控制, 而非随意。论文选择的词汇来自前测的结果, 但也不能成为不予以控制的理由。另外, 多数研究之所以用形容词而非名词, 是因为具体的名词不足以概括关于某群体刻板印象表征的核心内容(王沛等, 2010)。更何况, 作者选择的词汇词性也超过了形容词和名词的范围, 比如“做饭带孩子”、“灭火救人”等, 而各群体间采用的词汇又没有得到很好的平衡也是问题所在。

回应: 非常感谢审稿专家严谨的意见! 诚如专家所言, 笔者在甄选每个社会群体的刻板印象词语时确实没有考虑和控制词性。笔者在前测确定刻板特质词语时, 仅仅考虑到词语对于描述某个社会群体的**典型性或高频性**。如本文“2.1.2 实验材料”部分对**靶刺激**材料甄选过程的描述, 本研究严格按照测量刻板印象的经典方法, 并参考以往实证研究测量刻板印象的常用方法(王沛, 杨亚平, 赵仑, 2010), 通过 3 个步骤的前测获取不同社会群体的刻板印象特质词: **第一步**, 通过开放式问卷(见附录 2)收集每个群体的刻板特质词; **第二步**, 对表述不同但意义相近的词语归类并进行频次统计; **第三步**, 采用 7 点量表评定前面选出的高频词描述相应社会群体特征的典型程度。指导语要求被试根据特质词描述社会群体的贴切程度进行评分。如: 根据你对“科学家”的印象, 请对下列词语描述“科学家”的贴切程度在 7 点量表上进行评分, 1-7 表示贴切程度依次加强: 1 代表“非常不贴切”, 7 代表“非常贴切”。通过上述步骤获得 24 个启动群体的刻板印象词语, 如下表 A 中“刻板印象词语”一列所示。

尽管本研究严格按照测量刻板印象的经典方法并参考以往的相关实证研究, 通过逐步前测获得

每个群体的刻板印象词语，但是笔者及其合作者非常认同审稿专家提出的“**论文选择的词汇来自前测的结果，但也不能成为不予以控制的理由**”。笔者慎重反思这种事先**未控制词性**是否会影响研究结果的可靠性？为此，笔者专门请教了某高校研究汉语词汇的学者，请该学者对每个词语的词性进行分析并一一标注出其词性，结果如下表 A 中“**词语性质**”一列所示。如表 A 所标注的，每个群体的刻板印象词语的性质的确不尽相同，除了形容词、名词、动词，还包括主谓短语、动宾短语、联合式固定词组、偏正短语。尽管如此，每个群体的典型刻板印象词语中还是形容词最多，如表 B 对每个群体中词性个数的统计：在 HW-HC 社会群体中形容词有 23 个，在 HW-LC 社会群体中形容词有 20 个，在 LW-HC 社会群体中形容词有 19 个，在 LW-LC 社会群体中形容词有 28 个。

值得注意的是，笔者所请教的研究汉语词汇的学者在分析词性后发现，本研究中所采用的这些**主谓短语**和**动宾短语**都在表达和描述关于某一社会群体的某种性质或者状态，与形容词很接近——**形容词在本质上就是表示人或事物的形状、性质，或者动作、行为、变化的状态的词**，因此在描述社会群体的刻板印象这一语义背景中可视其具有**与形容词等同的语义功能**。详述如下：

(1) 首先，本研究采用的**主谓短语都具有形容词性的谓语**，且**语义焦点在谓语部分**，因此**这些短语与形容词接近**。如表 B 中“**主谓短语**”一行所列出的 7 个主谓短语——“工作危险”、“行动迅速”、“行动缓慢”、“生活艰辛”、“工作稳定”、“工作清闲”、“私生活混乱”——都具有形容词性的谓语（如，“危险”、“缓慢”、“艰辛”、“稳定”、“清闲”、“混乱”），而且语义焦点在谓语部分。具体来说，“消防员”的刻板印象词语“工作危险”中，语义焦点是**形容词“危险”**，强调消防员这一群体的工作性质是危险的，同样，其刻板印象词语“行动迅速”中语义焦点是**形容词“迅速”**，描述消防员的动作和行为特征；而“老年人”的刻板印象词语“行动缓慢”中，语义焦点是**形容词“缓慢”**，强调老年人这一群体常有的动作和行为状态；“农民工”的刻板印象词语“生活艰辛”中，语义焦点是**形容词“艰辛”**，强调农民工这一群体生活状态之艰难；“公务员”的刻板印象词语“工作稳定”和“工作清闲”中，语义焦点是**形容词“稳定”、“清闲”**，描述和强调公务员这一群体工作性质和状态；而“演艺明星”的刻板印象词语“私生活混乱”中，语义焦点是**形容词“混乱”**，强调演艺明星这一群体的私生活状态。总之，本研究前测所收集的刻板印象词语中所包含的 7 个主谓短语的语义焦点都是**形容词**，每个主谓短语都是在描述和强调某一群体的某方面的性质或状态，因此这些主谓短语与形容词非常接近，学者认为其具有**与形容词等同的语义功能**。

(2) 同样，本研究采用的**动宾短语也都是在描述某一群体的某方面的性质或状态**，因此与**形容词非常接近**，可视为其具有形容词的语义功能。如表 B 中“**动宾短语**”一行所列出的 9 个动宾短

语——“帮助他人”、“需要照顾”、“需要关爱”、“做家务”、“渴望陪伴”、“缺少关爱”、“重利益”、“喊口号”、“不办实事”——都在描述某一群体的某方面的性质或状态。具体来说，“帮助他人”用于描述“心理咨询师”的工作性质；“需要照顾”、“需要关爱”都在描述“老年人”这一群体的身心状态（是需要关爱和照顾的）；“做家务”描述了“家庭主妇”这一群体高频的行为状态；而“渴望陪伴”、“缺少关爱”在描述“留守儿童”这一群体的身心状态的；“重利益”用于描述“商人”这一群体的特质；“喊口号”、“不办实事”都在描述人们对“政府官员”这一群体的刻板印象。

综上所述，本研究所采用的**主谓短语**和**动宾短语**都在表达和描述关于某一社会群体的某种性质或者状态，与形容词非常接近，学者认为在描述某一社会群体的刻板印象这一语义背景中可视其具有**与形容词等同的语义功能**，因此可以将主谓短语和动宾短语归为形容词类中。总之，尽管本研究关于刻板印象词语的前测中没有控制词性，但上述事后分析发现前测确定的词语中大多数还是形容词。令人疑惑的是，为什么前测结果中也会出现上述 7 个主谓短语和 9 个动宾短语呢？笔者所请教的汉语词汇的学者认为，虽然主谓短语和动宾短语与常见的两音节（如“威严”）、三音节（如“爱唠叨”）或四音节（如“美丽大方”）的形容词在形式和结构上不同，但由于这些描述群体刻板印象词语的使用频率很高——一旦一个词语高频使用，就有可能形成结构稳定的短语或词组，其词性要**根据语义背景或/和词组构成**来确定。如前所述，在本研究的描述群体刻板印象这一语义背景中，主谓短语和动宾短语都在表达和描述关于某一社会群体的某种性质或者状态，在人们的日常生活中经常描述某个群体，因此前测中会出现高频的主谓短语和动宾短语的刻板印象词语。最后，笔者将每类社会群体别中的**形容词、主谓短语和动宾短语**合并后统计个数，如表 B “合并后个数” 一列所示，在每个群体中**具有形容词语义功能**的刻板印象词语占绝大多数，且 4 类社会群体的个数基本一致：在 HW-HC 社会群体中有 26 个，在 HW-LC 社会群体中有 27 个，在 LW-HC 社会群体中有 25 个、在 LW-LC 社会群体中依然是 28 个。如表 B 最后一行“总计”所示，本研究中每个热情-能力社会群体类别中的共有 30 个刻板印象词语，因此 26 个、27 个、25 个或 28 个词语在 30 个中占绝大部分，推测应该不会影响本研究的实验结果（为检验这一推测，笔者对两个实验的数据做了重新的统计分析，详细的过程及其结果见下文）。

此外，对于每个群体类别中的其它词性的词语，汉语词汇学者也一一做了分析。如表 B 所示，除了上面分析过的形容词、主谓短语和动宾短语，剩余为数不多 14 个词语中**名词**最多（有 6 个：“弱势群体”、“铁饭碗”、“官腔”、“形式主义”、“官腔官调”、“监狱”），其次是**联合式固定词组**（有 4 个：“保家卫国”、“灭火救人”、“教书育人”、“做饭带孩子”）和**动词**（有 3

个：“种田”、“留学”、“啃老”），最后还有1个偏正短语（“微笑服务”）。下面对每类词逐一分析。

首先，分析数量最多的名词。诚如专家所言，在以往关于刻板印象的研究中“……多数研究之所以用形容词而非名词，是因为具体的名词不足以概括关于某群体刻板印象表征的核心内容（王沛等，2010）”，刻板印象领域的研究者也普遍认可并通过形容词描述关于某群体的特质，这与“形容词”的定义是相符的——**形容词在本质上就是表示人或事物的形状、性质，或者动作、行为、变化的状态的词**。那么，为什么前测结果中也会出现这几个高频的名词呢？汉语词汇研究的学者认为这是一种较为普遍的名词活用为形容词的语言现象，这些名词在一定语义背景中往往隐含、伴随某种性质或者状态。例如，“女人”“哥们”，“农村”这些名词在一定的语义背景下活用为形容词，“女人”隐含一个人具有女性化的特质（在生活中我们常常会听到有人说“这个人真女人！”），在一定语义背景中“哥们”隐含一个人具有讲义气这一特质（在生活中我们常常会听到有人说“这个人真哥们。”），“农村”在一定的语义背景中隐含一个人土里土气（在生活中我们常常会听到有人说“这个人真农村”）。由此推测，本研究前测中获得的6个高频的名词（“弱势群体”、“铁饭碗”、“官腔”、“形式主义”、“官腔官调”、“监狱”）也许是被试根据指导语——“**请您根据印象写出一些经常用来描述下列每个社会群体特征的词语**”——写出他们关于某一群体的印象的词语时，不知不觉中地将名词活用为形容词。例如，关于农民工的刻板印象词语“弱势群体”隐含了人们认为农民工的社会地位低，会遭遇偏见和歧视的刻板印象；“铁饭碗”隐含了人们认为公务员这一工作非常稳定的刻板印象；“官腔”、“形式主义”、“官腔官调”隐含了人们为政府官员说话时拿腔拿调，做事常常搞形式主义的刻板印象，而“监狱”反映了人们刻板印象中认为与“罪犯”常常伴随的状态。

另外，学者也分析了4个联合式固定词组（“保家卫国”、“灭火救人”、“教书育人”、“做饭带孩子”），尽管这些词语不具有明确的词性，但是因为这些词语经常在一起高频使用且意义融合，而一旦一个词语高频使用，有可能形成结构稳定的固定词组，这些词组在本研究中描述某一群体的刻板印象的语义背景中，在描述和表达相应群体的状态。学者认为，不只是形容词表达人或事物的形状、性质，或者动作、行为、变化的状态。诸如“保家卫国”、“灭火救人”、“教书育人”和“做饭带孩子”这4个词语分别都是在描述“军人”、“消防员”、“大学教授”和“家庭主妇”的工作状态，这些词语经常一起使用，从而已经成为一种联合式固定词组。例如，一想到军人，人们就会想到“保家卫国”，一想到家庭主妇，人们就会想到“做饭带孩子”等等。同样地，最后，对于3个动词（“种田”、“留学”、“啃老”）和1个偏正短语（“微笑服务”），尽管

不是形容词，但是也在描述和表达人们印象中关于某一群体的经常从事的行为状态。

表 A 基于 SCM 不同热情-能力 4 类社会群体的刻板印象词语及其词性

HW-HC			HW-LC			LW-HC			LW-LC		
启动	刻板印象词语	词语性质	启动	刻板印象词语	词语性质	启动	刻板印象词语	词语性质	启动	刻板印象词语	词语性质
军人	保家卫国	联合式固定词组	老年人	需要照顾	动宾短语	商人	重利益	动宾短语	罪犯	监狱	名词
	威严	形容词		需要关爱	动宾短语		有钱	形容词		可恶	形容词
	坚毅	形容词		爱唠叨	形容词		精明	形容词		凶狠	形容词
	正义	形容词		白发苍苍	形容词		唯利是图	形容词		残暴	形容词
	正直	形容词	农民	行动缓慢	主谓短语	海归	留学	动词	无业游民	游手好闲	形容词
消防员	工作危险	主谓短语		慈祥	形容词		外语好	形容词		无所事事	形容词
	行动迅速	主谓短语		辛苦	形容词		见识广	形容词		无责任心	形容词
	灭火救人	联合式固定词组		种田	动词		有知识	形容词		啃老	动词
	令人尊敬	形容词		憨厚老实	形容词	公务员	工作稳定	主谓短语	乞丐	没有上进心	形容词
大学教师	敬业	形容词		勤劳	形容词		铁饭碗	名词		脏	形容词
	勇敢	形容词		收入低	形容词		工作清闲	主谓短语		衣衫褴褛	形容词
	有学问	形容词	家庭主妇	做饭带孩子	联合式固定词组		待遇好	形容词		邋遢	形容词
	知识渊博	形容词		顾家	形容词	政府官员	官腔	名词		懒惰	形容词
心理咨询师	教书育人	联合式固定词组		做家务	动宾短语		形式主义	名词	吸毒者	贫穷	形容词
	善于观察	形容词		贤惠	形容词		官腔官调	名词		颓废	形容词
	懂人心	形容词	农民工	勤劳	形容词		喊口号	动宾短语		堕落	形容词
	有耐心	形容词		辛苦	形容词	演艺明星	不办实事	动宾短语		萎靡	形容词
空姐	温和	形容词		劳累	形容词		贪污腐败	形容词	恐怖分子	精神涣散	形容词
	帮助他人	动宾短语		生活艰辛	主谓短语		大腹便便	形容词		骨瘦如柴	形容词
	礼仪好	形容词	留守儿童	弱势群体	名词		收入高	形容词		形容枯槁	形容词
	微笑服务	偏正短语		渴望陪伴	动宾短语		绯闻多	形容词		瘦弱	形容词
	有气质	形容词		缺少关爱	动宾短语		光鲜亮丽	形容词		极端	形容词
	漂亮	形容词		孤独	形容词		漂亮	形容词		残暴	形容词
	美丽大方	形容词	清洁工	可怜	形容词		私生活混乱	主谓短语		凶残	形容词
	举止优雅	形容词		辛苦	形容词		好看	形容词		心理扭曲	形容词

瑜伽教练	柔韧性强	形容词	早起	形容词	富人	有钱	形容词	城管	没人性	形容词
	体形优美	形容词		形容词		奢侈	形容词		欺软怕硬	形容词
	身材好	形容词		形容词		傲慢	形容词		蛮横	形容词
	健美	形容词		形容词		狡猾	形容词		凶残	形容词
	健康	形容词		形容词		挥霍	形容词		没有同情心	形容词

表 B 不同社会群体刻板印象词语的词性汇总统计

词性	HW-HC			HW-LC			LW-HC			LW-LC		
	个数	词语列举	合并后个数	个数	词语列举	合并后个数	个数	词语列举	合并后个数	个数	词语列举	合并后个数
形容词	23	威严 …… 健康	26	20	爱唠叨 …… 工资低	27	19	威严 …… 健康	25	28	可恶 …… 没有同情心	28
主谓短语	2	工作危险 行动迅速		2	行动缓慢 生活艰辛		3	工作稳定 工作清闲 私生活混乱		0		
动宾短语	1	帮助他人		5	需要照顾 需要关爱 做家务 渴望陪伴 缺少关爱		3	重利益 喊口号 不办实事		0		
联合式固定词组	3	保家卫国 灭火救人 教书育人		1	做饭带孩子		0			0		
名词	0		26	1	弱势群体	27	4	铁饭碗 官腔 形式主义 官腔官调	25	1	监狱	28
动词	0			2	种田 做家务		1	留学		1	啃老	
偏正短语	1	微笑服务		0			0			0		
总计	30		26	30		27	30		25	30		28

综上所述，对本研究所采用的刻板印象词语的词性分析发现，前测确定的词语中绝大多数是形容词（如表 C 第 A 行所示，在 HW-HC 社会群体中形容词有 23 个，在 HW-LC 社会群体中形容词有 20 个，在 LW-HC 社会群体中形容词有 19 个，在 LW-LC 社会群体中形容词有 28 个）。另外，除了形容词，还出现 7 个主谓短语和 9 个动宾短语，上述分析发现在描述群体刻板印象这一语义背景中，前测中所出现的这些高频的主谓短语和动宾短语都在表达和描述关于某一社会群体的某种性质或者状态，**具有与形容词等同的语义功能**，因此将其与形容词归为一类（如表 C 第 B 行所示，合并后在 HW-HC 社会群体中有具有形容词语义功能的词语有 26 个，在 HW-LC 社会群体中有 27 个，在 LW-HC 社会群体中有 25 个，在 LW-LC 社会群体中形容词有 28 个）。最后，前测也出现 6 个名词，4 个联合式固定词组，3 个动词和 1 个偏正短语，学者的分析发现尽管其结构与形容词不同，但其在语义功能上与形容词有一定程度的相似。总之，笔者所请教的汉语词汇学者对刻板印象词语词性的分析表明，这些词语在结构上的差异在理论上并不会影响研究结果的可靠性。

为了进一步检验确定学者的分析在统计上是否可靠，笔者对本研究两个实验的数据重新做了统计分析。具体来说，将表 C 中第 A 行 90 个形容词的结果与第 B 行 106 个合并形容词、主谓短语和动宾短语后的结果，以及第 C 行所有词即原文的结果对照，来检验词性是否会影响实验结果的稳定性。

表 C 不同社会群体刻板印象词语词性汇总频次比较

词性		HW-HC Trials 个数	HW-LC Trials 个数	LW-HC Trials 个数	LW-LC Trials 个数	总计 Trials 个数
A	形容词	23	20	19	28	90
B	合并：形容词、主谓短语和动宾短语	26	27	25	28	106
C	所有词：形容词、主谓短语、动宾短语、联合式固定词组、名词、动词、偏正短语	30	30	30	30	120

首先通过卡方检验（配合度检验）确定每类群体所分析的 trials 个数（实际频数）是否与原文中的 trial 个数（理论频数）一致。若只分析表 C 中第 A 行 90 个形容词，则 HW-HC 社会群体有 23 个 trials, HW-LC 社会群体有 20 个 trials, LW-HC 社会群体有 19 个 trials, HW-LC 社会群体有 28 个 trials。简言之，实际频数为（23，20，19，28），理论频数为（30，30，30，30），卡方检验结果表明 4 类社会群体的形容词 trials 个数没有显著差异， $\chi^2(3) = 2.18, p = 0.536$ 。

另外，若分析表 C 中第 B 行 106 个合并的具有明显形容词语义功能的词语，则高热情-高能力 (HW-HC) 社会群体有 26 个 trials，高热情-低能力 (HW-LC) 社会群体有 27 个 trials，低热情-高能力 (LW-HC) 社会群有 25 个 trials，低热情-低能力 (HW-LC) 有 28 个 trials。简言之，实际频数为（26，27，25，28），理论频数为（30，30，30，30），卡方检验结果表明合并后 4 类社会群体的 trial 个数没

有显著差异, $\chi^2(3) = 0.19, p = 0.979$ 。总之, 若只分析形容词, 或合并具有形容词语义功能的词, 每类社会群体的 trial 个数均比较充分, 且卡方检验表明 4 类社会群体的 trial 个数均与原文 trial 个数不存在显著差异。因此, 可以对 90 个形容词的结果, 106 个合并词语的结果与原文 120 个靶子词的结果进行对比, 来检验词性是否会影响实验结果的稳定性。统计结果详述如下。

表 D 实验 1 方差分析结果对比: 原文 120 个靶子词, 90 个形容词, 106 个合并词

数据来源	观测指标	变异来源	df	MS	F	p	Partial η^2
原文 120 个靶子词	正确率	社会群体	3	0.169	46.514	.000	0.487
		Error(社会群体)	147	0.004			
		一致性	1	0.005	0.476	.494	0.010
		Error(一致性)	49	0.011			
		社会群体 \times 一致性	3	0.051	16.389	.000	0.251
		Error(社会群体 \times 一致性)	147	0.003			
	反应时	社会群体	3	151388.811	37.333	.000	0.432
		Error(社会群体)	147	4055.114			
		一致性	1	30800.066	3.106	.084	0.060
		Error(一致性)	49	9917.319			
		社会群体 \times 一致性	3	117622.203	29.832	.000	0.378
		Error(社会群体 \times 一致性)	147	3942.784			
90 个形容词	正确率	社会群体	3	0.043	14.143	.000	0.224
		Error(社会群体)	147	0.003			
		一致性	1	0.009	1.119	.295	0.022
		Error(一致性)	49	0.008			
		社会群体 \times 一致性	3	0.028	8.050	.000	0.141
		Error(社会群体 \times 一致性)	147	0.004			
	反应时	社会群体	3	200704.275	33.493	.000	0.406
		Error(社会群体)	147	5993.460			
		一致性	1	65589.060	5.059	.029	0.094
		Error(一致性)	49	12963.962			
		社会群体 \times 一致性	3	134114.603	25.312	.000	0.341
		Error(社会群体 \times 一致性)	147	5298.448			
靶子词 为合并(形容词以及 具有形容词语义功 能)的 106 个词	正确率	社会群体	3	0.160	50.722	.000	0.509
		Error(社会群体)	147	0.003			
		一致性	1	0.025	2.525	.119	0.049
		Error(一致性)	49	0.010			
		社会群体 \times 一致性	3	0.038	11.905	.000	0.195
		Error(社会群体 \times 一致性)	147	0.003			
	反应时	社会群体	3	163640.602	32.170	.000	0.396
		Error(社会群体)	147	5086.750			
		一致性	1	24344.643	2.603	.113	0.050
		Error(一致性)	49	9352.526			
		社会群体 \times 一致性	3	115804.359	26.728	.000	0.353
		Error(社会群体 \times 一致性)	147	4332.737			

与原文中数据分析方法相同, 即对于实验 1 和实验 2 中 90 个形容词 trial 的数据, 以及 106 个合

并形容词以及具有形容词语义功能的词语的数据，以反应时和正确率为因变量，进行 4(社会群体: HW-HC, HW-LC, LW-HC, LW-LC)×2(一致性: 一致 vs. 冲突)的重复测量方差分析，所有因素均为组内变量。不同靶子的统计分析结果与原文结果对比如下。

①实验 1 统计分析显示：分析 90 个形容词以及 106 个合并（形容词以及具有明显形容词语义功能的）词语的结果与原文实验 1 的结果完全一致。详述如下：

首先，正确率和反应时的方差分析结果与原文实验 1 的结果一致。如表 D 所示，不论在正确率还是反应时上，方差分析结果均显示社会群体与一致性的交互作用极其显著。

表 E 实验 1 的正确率和反应时 (*M; SE*) 及其一致性的简单效应检验结果对比：原文, 90 个形容词, 106 个合并词

观测指标	启动	启动-靶子的一致性		Mean Difference (I-C)	F	p	Partial η^2	
	社会群体	一致(C)	冲突(I)					
原文 120 个 靶子词	正确率	HW-HC	0.971 (0.006)	0.944 (0.007)	-0.027	10.08	.003	0.171
		HW-LC	0.938 (0.007)	0.935 (0.010)	-0.003	0.07	.788	0.001
		LW-HC	0.877 (0.012)	0.903 (0.010)	0.025	2.21	.143	0.043
		LW-LC	0.843 (0.010)	0.918 (0.012)	0.075	21.30	.000	0.303
	反应时(ms)	HW-HC	774 (23)	851 (26)	77	27.73	.000	0.361
		HW-LC	839 (24)	875 (28)	36	5.64	.022	0.103
		LW-HC	872 (24)	911(26)	39	6.35	.015	0.115
		LW-LC	937 (29)	856 (26)	-81	34.75	.000	0.415
靶子词 为 90 个形 容词	正确率	HW-HC	0.975 (0.006)	0.953 (0.008)	-0.021	5.96	.018	0.108
		HW-LC	0.935 (0.009)	0.951 (0.011)	0.016	1.36	.250	0.027
		LW-HC	0.915 (0.010)	0.882 (0.011)	-0.033	3.79	.057	0.072
		LW-LC	0.851 (0.010)	0.918 (0.012)	0.067	17.63	.000	0.265
	反应时(ms)	HW-HC	760 (23)	834 (27)	74	18.56	.000	0.275
		HW-LC	828 (24)	864 (26)	36	4.46	.040	0.083
		LW-HC	847 (22)	921 (31)	74	12.90	.001	0.208
		LW-LC	937 (28)	856 (29)	-81	42.92	.000	0.467
靶子词 为合并（形 容词以及 具有明显 形容词语 义功能）的 106 个词	正确率	HW-HC	0.968 (0.007)	0.944 (0.007)	-0.024	5.96	.012	0.122
		HW-LC	0.944 (0.007)	0.942 (0.010)	-0.002	1.36	.858	0.001
		LW-HC	0.866 (0.012)	0.888 (0.011)	0.022	3.79	.224	0.030
		LW-LC	0.851 (0.010)	0.918 (0.012)	0.067	17.63	.000	0.267
	反应时(ms)	HW-HC	760 (23)	834 (27)	74	18.56	.000	0.275
		HW-LC	828 (24)	864 (26)	36	4.46	.040	0.083
		LW-HC	847 (22)	921 (31)	74	12.90	.001	0.208
		LW-LC	937 (28)	856 (29)	-81	42.92	.000	0.467

进一步对交互作用分析结果进行对比，如表 E 所示，简单效应检验的结果趋势及其显著性没有发生任何变化。由此可见，如下图 1 所示，不论是仅仅采用形容词（90 个），还是将形容词与具有与形容词等同的语义功能的主谓短语和动宾短语合并（106 个），或者是原文中包含所有词性的词

(120 个词=106 个合并具有形容词语义功能的词+6 个名词+4 个联合式固定词组+3 个动词和 1 个偏正短语)，实验 1 的基本结果趋势及其显著性是一致的，由此推测词性对实验 1 结果的稳定性没有显著影响。

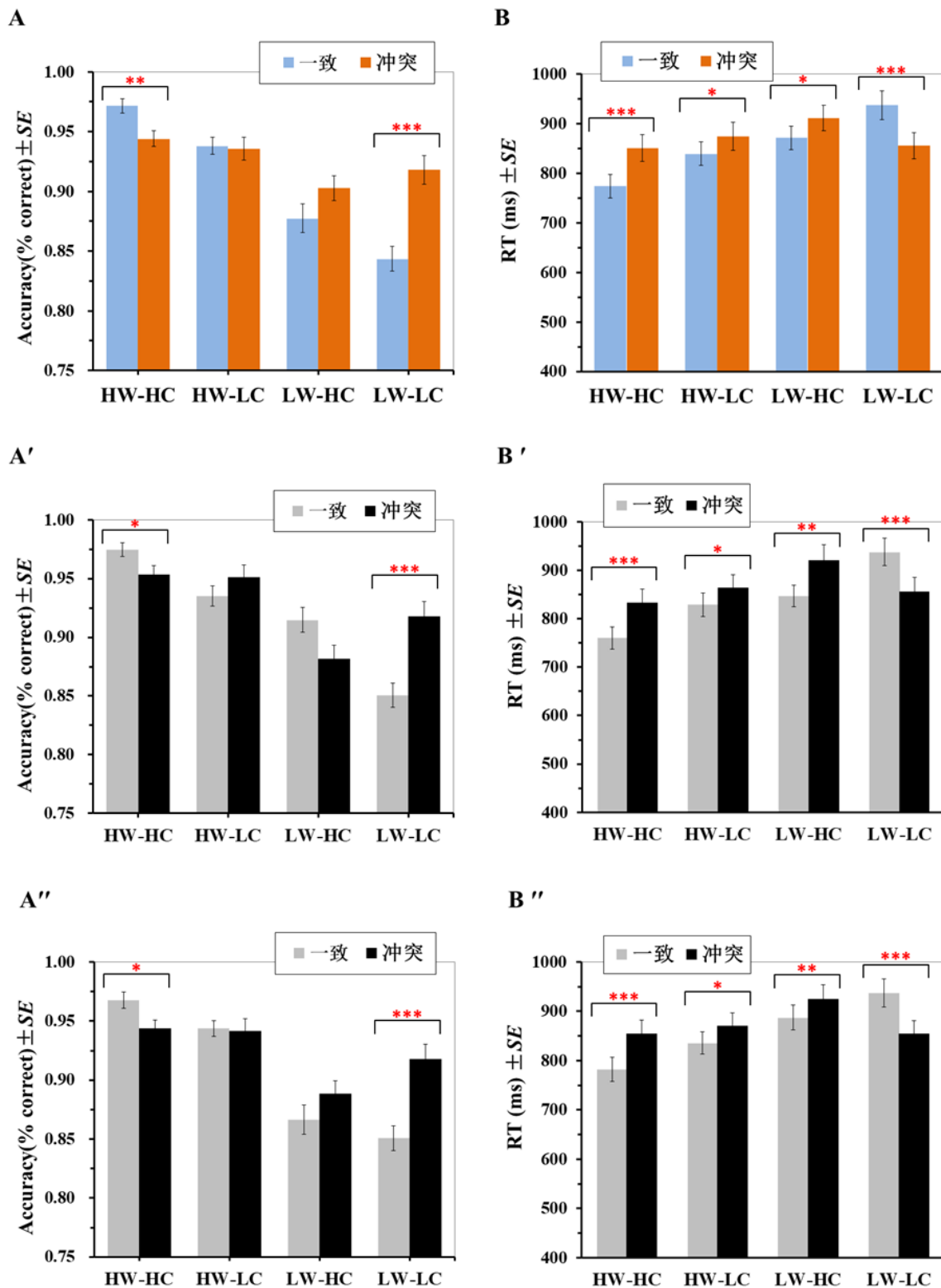


图 1 实验 1 结果对比图：原文 120 个靶子词的正确率(A)和反应时(B)，靶子词为 90 个形容词的正确率(A')和反应时(B')，靶子词为合并（形容词以及具有明显形容词语义功能）的 106 个词的正确率(A'')和反应时(B'')

②实验 2 统计分析显示：分析 90 个形容词以及 106 个合并（形容词以及具有明显形容词语义功能的）词语的结果与原文实验 2 的结果完全一致。详述如下：

首先，正确率和反应时的方差分析结果与原文实验 2 的结果一致。如表 F 所示，不论在正确率还是反应时上，方差分析结果均显示社会群体与一致性的交互作用极其显著。

表 F 实验 2 方差分析结果对比：原文 120 个靶子词, 90 个形容词, 106 个合并词

观测指标		变异来源	df	MS	F	p	Partial η^2
原文 120 个靶子词	正确率	社会群体	3	0.01	3.53	.017	0.07
		Error(社会群体)	135	0.00			
		一致性	1	0.03	5.67	.022	0.11
		Error(一致性)	45	0.01			
		社会群体 × 一致性	3	0.02	12.35	.000	0.22
		Error(社会群体×一致性)	135	0.00			
	反应时	社会群体	3	3445.28	2.91	.040	0.06
		Error(社会群体)	135	1184.70			
		一致性	1	13599.79	11.87	.000	0.21
		Error(一致性)	45	1146.04			
		社会群体×一致性	3	51092.61	34.77	.000	0.44
		Error(社会群体×一致性)	135	1469.37			
90 个形容词	正确率	社会群体	3	0.004	1.82	.147	0.04
		Error(社会群体)	135	0.002			
		一致性	1	0.032	5.39	.025	0.11
		Error(一致性)	45	0.006			
		社会群体 × 一致性	3	0.017	7.31	.000	0.14
		Error(社会群体×一致性)	135	0.002			
	反应时	社会群体	3	3128.56	2.08	.106	0.04
		Error(社会群体)	135	1504.98			
		一致性	1	10751.68	5.93	.019	0.12
		Error(一致性)	45	1814.23			
		社会群体×一致性	3	49499.55	29.23	.000	0.39
		Error(社会群体×一致性)	135	1693.44			
靶子词 为合并（形容词以 及具有形容词语义 功能）的 106 个词	正确率	社会群体	3	0.008	3.31	.022	0.07
		Error(社会群体)	135	0.002			
		一致性	1	0.020	4.22	.046	0.09
		Error(一致性)	45	0.005			
		社会群体 × 一致性	3	0.020	11.26	.000	0.20
		Error(社会群体×一致性)	135	0.002			
	反应时	社会群体	3	3015.07	1.97	.122	0.04
		Error(社会群体)	135	1533.01			
		一致性	1	15483.59	9.56	.003	0.18
		Error(一致性)	45	1618.88			
		社会群体×一致性	3	52745.86	35.02	.000	0.44
		Error(社会群体×一致性)	135	1506.16			

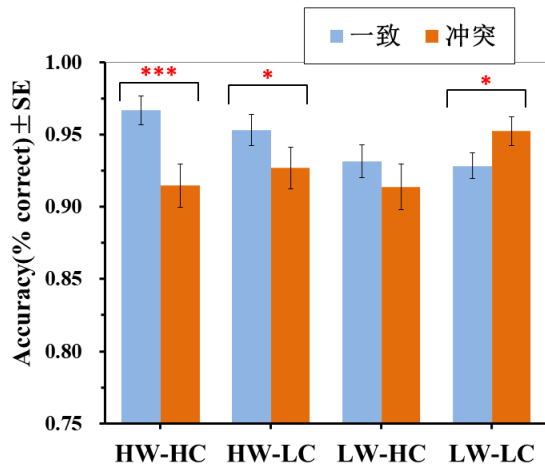
进一步对交互作用分析结果进行对比，如表 G 所示，简单效应检验的结果趋势及其显著性没有发生任何变化。由此可见，如下图 2 所示，不论是仅仅采用形容词（90 个），还是将形容词与具有与形容词等同的语义功能的主谓短语和动宾短语合并（106 个），或者是原文中包含所有词性的词

（120 个词=106 个合并具有形容词语义功能的词+6 个名词+4 个联合式固定词组+3 个动词和 1 个偏正短语），实验 2 的基本结果趋势及其显著性是一致的，由此推测词性对实验 2 结果的稳定性未产生显著影响。

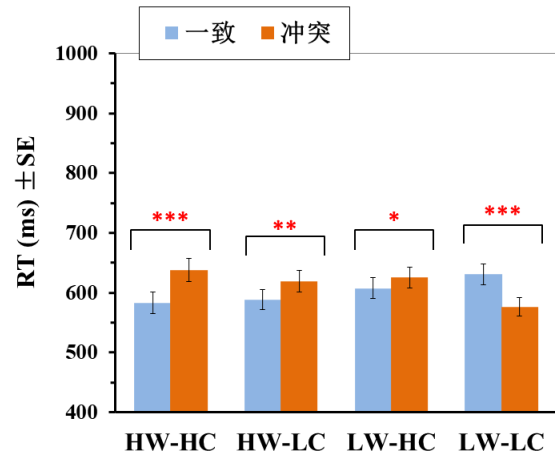
表 G 实验 2 的正确率和反应时 (M ; SE) 及其一致性的简单效应检验结果对比：原文, 90 个形容词, 106 个合并词

	启动	启动-靶子的一致性		Mean Difference (I-C)	F	p	Partial η ²
	社会群体	一致(C)	冲突(I)				
原文 120 个靶子词	HW-HC	0.967 (0.010)	0.914 (0.015)	-0.052	18.16	.000	0.287
	HW-LC	0.953 (0.011)	0.927 (0.014)	-0.026	5.33	.026	0.106
	LW-HC	0.931 (0.011)	0.914 (0.016)	-0.017	3.01	.090	0.063
	LW-LC	0.928 (0.009)	0.952 (0.010)	0.024	6.29	.016	0.123
	HW-HC	582 (18)	638 (19)	56	57.11	.000	0.559
	HW-LC	588 (17)	618 (18)	30	13.99	.001	0.237
	LW-HC	607 (17)	625 (18)	18	4.98	.031	0.100
	LW-LC	631 (18)	576 (16)	-55	48.47	.000	0.519
靶子词 为 90 个形容词	HW-HC	0.964 (0.010)	0.917 (0.015)	-0.047	12.10	.001	0.212
	HW-LC	0.960 (0.008)	0.937 (0.014)	-0.023	4.25	.045	0.086
	LW-HC	0.945 (0.010)	0.920 (0.015)	-0.025	3.30	.076	0.068
	LW-LC	0.933 (0.009)	0.952 (0.009)	0.019	4.41	.041	0.089
	HW-HC	575 (19)	630 (19)	55	69.73	.000	0.608
	HW-LC	579 (16)	602 (18)	23	6.35	.015	0.124
	LW-HC	592 (18)	612 (18)	20	4.29	.044	0.087
	LW-LC	629 (18)	574 (16)	-55	37.76	.000	0.456
靶子词 为合并(形容词以 及具有明显形容 词语义功能) 的 106 个词	HW-HC	0.966 (0.010)	0.914 (0.015)	-0.051	16.89	.000	0.273
	HW-LC	0.958 (0.010)	0.936 (0.014)	-0.022	4.79	.034	0.096
	LW-HC	0.928 (0.013)	0.924 (0.015)	-0.004	0.15	.699	0.003
	LW-LC	0.933 (0.009)	0.952 (0.009)	0.019	4.49	.040	0.091
	HW-HC	576 (19)	637 (19)	61	89.01	.000	0.664
	HW-LC	584 (17)	612 (18)	28	9.02	.004	0.167
	LW-HC	603 (18)	620 (19)	17	4.39	.042	0.089
	LW-LC	629 (18)	574 (16)	-53	36.75	.000	0.450

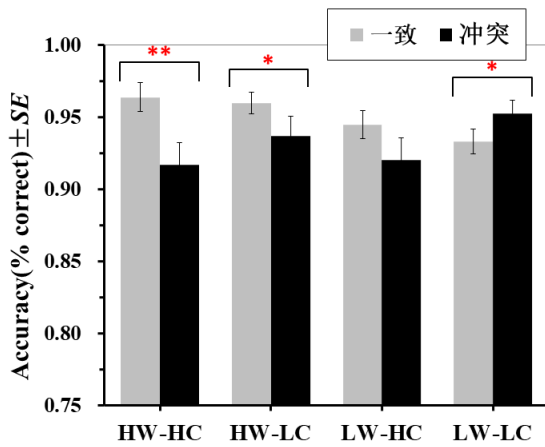
C



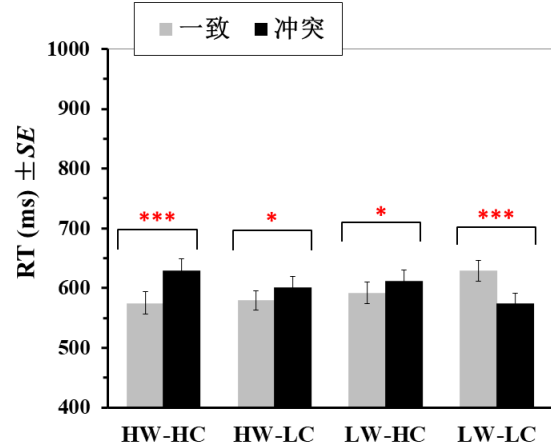
D



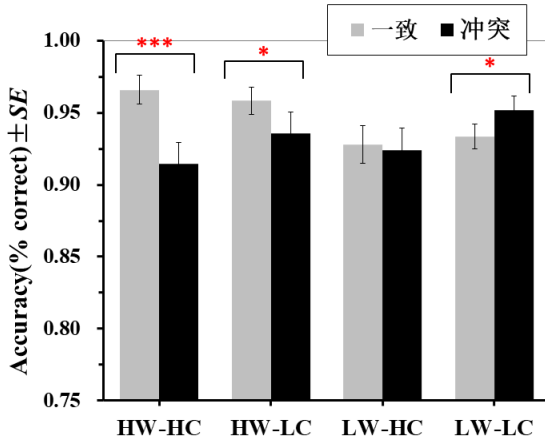
C'



D'



C''



D''

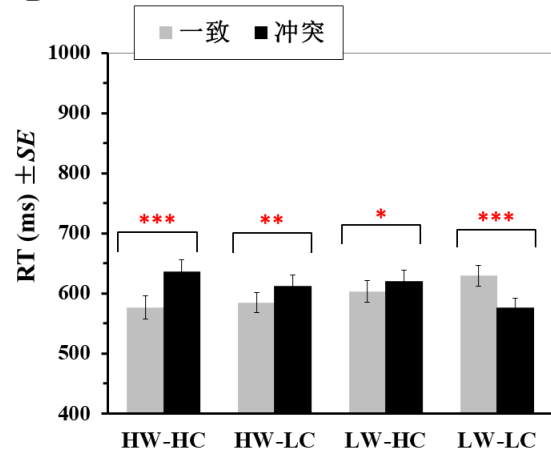


图2 实验2结果对比图: 原文120个靶子词的正确率(C)和反应时(D), 靶子词为90个形容词的正确率(C')和反应时(D'), 靶子词为合并(形容词以及具有明显形容词语义功能)的106个词的正确率(C'')和反应时(D'')

综上所述, 通过将原120个靶子词(包括106个合并具有形容词语义功能的词、6个名词、4个

联合式固定词组、3 个动词和 1 个偏正短语)的结果与 90 个形容词的结果, 以及 106 个合并形容词以及具有形容词语义功能的词语的结果一一对比发现, 不论是实验 1 外显任务还是实验 2 内隐任务, 结果的趋势和显著性没有受到任何影响。这反映了本研究的结果是可靠和稳定的。

虽然数据重新分析的结果表明, 词性并没有对本实验结果产生显著影响, 但是诚如专家的意见, 考虑到实验控制的严谨性, 词性确实应该是进行严格匹配和控制的。为此, 笔者在研究局限中也对这一问题进行了相应说明。具体如下:

“再次, 本研究中的靶子词虽然是根据测量刻板印象的经典方法, 经过严格前测甄选出来的, 但是前测中主要考虑和控制刻板词的典型性, 而并未对其词性进行严格匹配和控制。尽管后续的补充数据统计分析中并未发现词性对实验结果的显著影响, 但是未来研究可以对靶子词的词性进行严格的控制和匹配, 对该研究结果进行进一步验证和推进。”

最后, 再次诚挚地感谢两位审稿专家花费宝贵的时间和精力阅读和评审本文, 为本文的修改和完善提出了严谨而重要的意见和建议, 也为我们未来的研究提供了难得的启示!

第五轮

编委意见:

意见 1: 该论文经过几轮的修改, 基本已经达到发表的要求, 但是目前论文中仍存在一些书写格式问题, 例如“平均数 ± 2 SD”这种中英文混用的情况, 请作者仔细修改。

回应: 感谢编委的细致建议! 我们在对全文进行仔细审读之后, 修改了中英文混用的情况, 以期论文表述更符合中文阅读习惯。修改部分已用蓝色字体标出。

如 2.1.3 实验设计部分“采用 4(社会群体: HW-HC, HW-LC, LW-HC, LW-LC) \times 2(一致性: 一致, 冲突)的被试内实验设计。”，“如表 1(左半部分实验 1)所示, 每种实验条件下有 30 个试次”。

2.1.4 实验程序部分“采用 E-prime 2.0 编程。在正式实验前, 先进行 24 个试次的练习, 短暂休息后进入正式实验。在正式实验中, 被试共完成 240 个随机呈现的试次, 每 80 个试次休息一次。所有试次呈现刺激的顺序如图 1 所示: 首先呈现注视点“+”(500 ms), 紧接着呈现启动刺激(700 ms), 其后间隔 500 ms 空屏, 然后呈现靶刺激(350 ms); 靶刺激之后呈现空屏, 待被试做出反应后随机间隔 600~800 ms, 开始下一个试次”。

2.2 结果部分“以反应时和正确率为因变量, 进行 4(社会群体: HW-HC, HW-LC, LW-HC, LW-LC) \times 2(一致性: 一致, 冲突)的重复测量方差分析, 所有因素均为组内变量, 反应时在分析前剔除错误反应的以及超出平均数正负两个标准差的数据”。

3.1.2 实验设计部分“与实验 1 相同, 采用 4(社会群体: HW-HC, HW-LC, LW-HC, LW-LC)×2(一致性: 一致, 冲突)的被试内实验设计。因变量为正确率与反应时。”。

3.1.4 实验程序部分“实验 2 中的实验设备, 刺激的呈现方式(如字体、视角等), 实验的基本流程以及试次呈现方式与实验 1 完全相同。不同之处有两点: 一是正式实验的试次增加了一倍(即除了实验 1 中的 240 个“真词”试次, 还包括 240 个“假词”试次, 如表 1 所示), 因此实验 2 共需 30 分钟左右”。

3.2 结果部分“与实验 1 相同, 以反应时和正确率为因变量, 进行 4(社会群体: HW-HC, HW-LC, LW-HC, LW-LC)×2(一致性: 一致, 冲突)的重复测量方差分析, 所有因素均为组内变量, 反应时在分析前剔除错误反应的以及超出平均数正负两个标准差的数据”。

尾注 2“将按键左右手这一试间变量作为一个自变量, 对反应时和正确率为进行 2(按键左右手: 左手一致, 右手一致)×4(社会群体: HW-HC, HW-LC, LW-HC, LW-LC)×2(一致性: 一致, 冲突)的重复测量方差分析, 其中按键手为被试间变量, 其它所有因素均为组内变量”。

尾注 3“同实验 1 一样, 将按键左右手这一试间变量作为一个自变量, 对反应时和正确率为进行 2(按键左右手: 左手一致, 右手一致)×4(社会群体: HW-HC, HW-LC, LW-HC, LW-LC)×2(一致性: 一致, 冲突)的重复测量方差分析, 其中按键手为被试间变量, 其它所有因素均为组内变量”。

表一

社会群体	实验2(靶子包括4类群体的刻板印象词语及其对应的假词, 假词不分析)					
	实验1(靶子均为词)			由真词的形近字构成的假词		
	一致性	刺激举例	试次个数	一致性	刺激举例	试次个数
HW-HC	一致	军人-正义	30	一致	军人-止丈	30
	冲突	军人-瘦弱	30	冲突	军人-叟溺	30
HW-LC	一致	老年人-慈祥	30	一致	老年人-滋样	30
	冲突	老年人-挥霍	30	冲突	老年人-浑霞	30
LW-HC	一致	商人-有钱	30	一致	商人-肴线	30
	冲突	商人-贫穷	30	冲突	商人-贫罕	30
LW-LC	一致	乞丐-懒惰	30	一致	乞丐-濼隋	30
	冲突	乞丐-勤劳	30	冲突	乞丐-谨荧	30