

《心理学报》审稿意见与作者回应

题目：道德概念垂直空间隐喻理解中的映射：双向性及其不平衡性

作者：鲁忠义 贾利宁 翟冬雪

第一轮

审稿人 1 意见：

意见 1：在前言中：(1)在介绍道德概念的相关研究时，转折比较突兀，建议进行修改润色。

(2)本研究使用四个实验，采用了不同的范式，结果实验 1 没有发现效应，而通过改动范式，结果发现了效应。可见范式在本研究也是一个重要的因素。因此，建议研究者在前言中适当增加研究范式的介绍以及范式可能存在的影响。

回应：(1) 谢谢专家的建议！在介绍了道德概念隐喻之后，我们用“在上述始源域中的具体概念里，对抽象概念的建构起重要作用的是人们对身体和空间的认知”这句话引导到下一段关于始源域问题的研究。(2) 感谢审稿专家的建议。我们在前言中确实没有对影响实验结果的研究范式进行介绍，但在综合讨论 6.1 中进行了阐述。不过我们还是接受专家的意见，在引言中从实现研究目的而避免由于实验范式不同带来的无关变量的影响的角度，对此作了介绍，见正文蓝色字体。为了再审方便，将增加的这部分内容复制如下：

另外，要弄清道德概念垂直空间隐喻的双向性映射，还必须在研究从始源域向目标域映射和从目标域向始源域映射时，使用统一的研究范式。在以往的研究中，Meier 等人(2007)在实验 1 中使用的是 IAT 范式，使用这种范式分析实验结果时，把“道德与上和不道德与下”这一组的联结强度（相容性的平均反应时）同“道德与下和不道德与上”的联结强度（不相容的平均反应时）进行比较，充其量只是证明了“道德与上和不道德与下”有内隐的隐喻联结，而不能说明道德概念和空间位置是怎样映射的。而使用空间 Stroop 范式时，实验程序和方式也不尽相同。一种是把道德词或不道德词随机呈现在电脑屏幕的上方或下方，让被试按不同的键做词汇分类判断（如 Meie et al., 2007 实验 2；Hill et al., 2009）。另一种空间 Stroop 范式是引导式的，即将注视点从屏幕中央逐渐引导到屏幕的上方或下方，然后再在相应的位置呈现道德词或不道德词（王程和鲁忠义，2013）。这两种研究范式实质上与 IAT 范式是一样的，把道德词和不道德词与空间的上或下放在了一起，这可以证明道德概念垂直空间隐喻的心理现实性，但不易说明道德概念垂直空间隐喻的映射情况。因此为了避免实验范式不同带

来的隐喻双向性的差异，我们采用了统一的分离式空间 Stroop 范式，即在屏幕中央呈现启动项，而在屏幕上方或下方呈现反映项（详见实验 1 和实验 3 的实验程序），这样才能更好地观察一个项目对另一个项目的启动，从而去了解道德概念垂直空间隐喻中的映射。

意见 2: 在文章结构上：本研究的研究思路大概是从两个角度，通过在实验 1 和实验 3 中使用相同的范式来验证道德概念垂直空间隐喻存在双向的关系。实验 1 验证始源域到目标域的映射，实验 3 验证目标域到始源域的映射。结果实验 1 没有发现，实验 3 有发现。实验 2 中采用另一种迫选范式，证明了始源域到目标域也存在映射，只是这种映射关系比较弱。实验 4 中研究者又使用面孔图片，探讨空间位置对人物道德性的影响。实验 2 和实验 4 为何选择这种研究范式，特别是对于实验 4 进一步使用面孔图片进行探究的原因和价值的阐述较少，使得文章的结构不够清晰，降低文章的系统性和整体感。

回应: 关于四个实验之间的关系，我们在前言、四个实验之后的讨论中都有交代，这里综合起来说明一下几个实验之间的逻辑关系。实验 1 探讨无意识词义判断条件下，道德概念垂直空间隐喻中始源域向目标域的映射，但没有发现这一映射。这一结果的出现意味着什么呢？是说人们不是通过上下这一始源域向道德这一目标域的映射来理解道德概念的，还是说这种方向的映射是存在的，只不过弱一些而没有显现出来罢了。为解决这一问题我们从两个方面进行了思考：一是改变研究范式，增强被试对始源域与目标域的意识性，做实验 2。实验 2 的结果证明了这种映射在明确的意识状态下是存在的。第二个方面的思考是：采取同实验 1 一样性质的研究范式，与另一个方向的映射（目标域向始源域的映射）作对比，所以做了实验 3。实验 3 的结果表明，道德概念垂直空间隐喻的目标域向始源域的映射是存在的。实验 3 采用了与实验 1 相同的实验范式和实验材料，结果显示存在着道德概念垂直空间隐喻的目标域向始源域的映射。这说明在道德概念垂直空间隐喻中，始源域向目标域的映射是弱的，在无意识状态下没有显现出来（见实验 1 结果），只有在意识状态时才显现出来（实验 2 结果），而目标域向始源域的映射在无意识状态下就自动地显现出来了（见实验 3 结果）。这说明道德概念垂直空间隐喻的映射是双向的，但二者的力量是不平衡的。

按说实验到此是可以结束的，但为了进一步证实实验 3 的结果的可靠性我们做了实验 4。这里的实验逻辑是：如果实验 3 得到的可以从抽象道德概念向空间位置方向进行自动化映射的结果是可靠的，那么就应该存在具身效应，就是说空间位置会影响到个体道德认知的判断。据此，我们进行了实验 4。

意见 3: 在研究方法上: (1)实验 1 中有两个任务, 包括道德概念判断和空间位置判断任务, 那么在数据处理中, 先删除正确率低于 80%的 2 名被试数据是依据那个任务删除的? 此外, 对于错误反应的数据, 研究者又是如何处理的? (2)实验程序中材料的呈现时间没有描述清楚。是呈现具体的某个时间段, 还是被试按键后消失? 对于实验 1, 研究者认为是一种无意识的加工, 但是, 一般的无意识加工是指材料呈现较短时间, 本研究中材料的时间是没有明确说明, 似乎材料呈现是不限时间, 是按键后消失的, 那如何证明被试没有意识到空间位置的加工呢? 建议补充说明。

回应: (1) 在本实验中, 删除的 2 名被试数据是依据被试在词义判断任务中正确率低于 80%的试次进行删除的, 在字母分类判断任务中被试的正确率均在 80%以上。

对于错误反应数据处理是这样的: 本实验的两个任务中, 只要一个任务判断错误, 就不能清晰说明是哪些因素造成的对实验结果的影响。因此, 在本实验 28 名被试的数据中, 删除 2 名被试 (正确率低于 80%) 的所有实验数据, 其余 26 名的被试的数据在统计分析时, 词义判断和字母分类判断两个任务的错误反应数据均予以删除, 即一个实验试次中两个判断任务只要出现一次错误该试次数据就予以删除, 这样正确率就达到了 100%。

(2) 本实验中, 词义判断环节, 实验材料呈现在屏幕中央, 被试自定步速阅读, 按键反应后消失。同时设定最长反应时限为 3000ms, 若在 3000ms 内被试不做反应, 实验材料同样消失, 在进行统计计算时, 该试次按错误反应计算。

关于无意识加工。无意识加工的实验范式很多, 如人工语法范式、任务分离范式、加工分离范式、IAT 范式、Stroop 范式以及掩蔽范式。作者所说的短时呈现确实是一种研究无意识加工的方法, 一般在掩蔽实验中就是这么做的, 靶刺激的呈现时间几十毫秒, 之后进行掩蔽。我们的实验用的是空间 Stroop 范式, 这在空间隐喻的研究中, 不失一种无意识自动加工的合适方法。

意见 4: 在讨论上: 文章中实验 1 和实验 2 使用不同的范式, 结果在实验 2 中发现了始源域到目标域也存在映射, 而在实验 1 中并没有发现, 研究者解释为实验范式的原因。在讨论中主要列出了不同实验范式的研究结果, 但是对于不同范式的之间的差异, 以及对道德垂直空间隐喻的映射的内在加工机制, 还缺少一定的阐述。

回应: 这问题与第一个问题类似, 见对第一个问题的回答。

意见 5: 在文章写作规范细节上:

(1)在 2.2.3 的实验材料描述中，多出了一个括号；

(2)文章的结果分析中，F1, F2 是否对应被试检验和项目检验，应该在文中清楚报告；

(3)在 2.4 的分析讨论中存在错别字，比如“第二个方面的思考时：”；其他部分也应仔细检查。

回应：谢谢审稿专家细心的审阅！除了根据这些问题修改外，我们再一次通读了全文，进行了修改（见正文红色字体）。

审稿人 2 意见：

意见 1：实验 1 与实验 3 均为 2×2 的被试内设计，但正式实验一共才 80 试次，每种实验条件下才 20 试次，再去除反应错误的数据与极端数据之后，每种实验条件下可能不足 20 试次，这对于反应时实验来说试次数太少，可能会导致结果不可靠。

回应：试次多少确实没有什么标准，但是从实验结果看，从显著性达到 0.05 水平的实验结果看，效应量 η^2 在 0.207 到 0.597 之间，实验效果还是不错的。当然增加试次，显著水平和效应量应该会相应提高。

意见 2：实验 2 的任务首先出声读出屏幕中央的汉语词，然后判断屏幕上方和下方的两个希腊语非词中哪一个与汉语词的词义相同，这一任务的反应太随意，受被试偏好影响比较大。

回应：本实验设计参考了 Lakens, Semin 和 Foroni (2011) 的研究。实验采用两因素被试内设计，平衡了每个希腊语非词的空间呈现位置，避免了实验任务受被试反应偏好的影响。本实验的实验任务能够反映出被试在词义加工后对空间位置加工的影响。

意见 3：实验 2 与实验 4 数据统计部分将所有被试在 40 个试次在的选择的总和放在一起，进行卡方分析不妥当。建议计算每个被试在每个实验条件下选择道德或不道德的比例，然后进行 T 检验即可。

回应：因为数据类型属于非连续的离散变量，因此还是应该用卡方检验。对这个问题我们咨询了国内的一些统计方面的专家，他们都认为应该用卡方检验。

审稿人 3 意见：

意见 1: 实验一，作者对实验材料的选取做了详细的说明，但在 2.1 的研究目的、2.2.3 的实验设计以及 2.4 的讨论中，对于实验设计的意义说明不是很清楚，实验一的具体假设是什么？通过怎样的逻辑推理得到 2.4 的讨论和分析中的结论，还请作者能够在文中详细说明。

回应: 感谢审稿专家的建议，我们根据专家建议做了修改。见 2.1 蓝色字体。为再审方便，先复制如下：

采用空间 Stroop 研究范式，探讨在无意识层面道德概念垂直空间隐喻的理解中，是不是存在着道德概念垂直空间隐喻中始源域向目标域的映射。如果这种映射存在，那么呈现启动词（道德或不道德词）就可以激活目标词（空间方位的上或下），说明始源域可以向目标域映射。如果不能发现这种映射，则不一定说这种映射不存在，而可能是一种弱映射。这需要在后面的实验中采取另一种研究范式继续对此进行探讨。

这种实验假设是探索性的。这是基于我们利用实验 1 的空间 Stroop 研究范式做了许多次研究而无结果才猜测这个方向的映射可能是一种弱映射。后面的研究的确证明了我们研究的设想，因此得出了双向映射的不平衡性。这就是实验 1 研究的意义所在。

意见 2: 实验一，2.3 的实验结果中，为什么要删除正确率低于 80% 的被试数据，有何依据和标准？

回应: 在本实验中，主要探讨被试先前的词义加工是否会对后续的空间位置的加工产生影响，若被试在先前词义判断任务中出现错误，就不能说明被试在该试次是否准确加工了词义。同样，对于后续的字母判断就不能说明是哪类词的加工对其产生了影响。所以，对正确率过低的被试予以删除。数据的删除标准参考了王程,鲁忠义(2013)《心理学报》上的《道德概念的垂直空间隐喻及其对认知的影响》等相关研究。

意见 3: 实验二，作者在实验中使用的希腊语非词，是否在实验前进行过前测，保证所用的 80 个希腊语非词给个体的效价感受是完全中性的，比如这些希腊语非词的拼写或者长度等？或者之前的相关研究已经证明了利用希腊语非词不会给被试带来不同的正性或者负性的感受。另外，实验二的结果中为何只提及了被试选择上、下方希腊语非词个数的分析，而没有不同条件下，反应时的差异比较？

回应：实验参考了 Lakens, Semin 和 Foroni (2011) 的研究，本实验对每个非词的长度做了控制，由 6-8 个希腊字母组成。没有评定它的正负性，是基于这样的考虑：（1）被试没有学习过希腊语；（2）而且它是由字母构成的非词；（3）这些非词是随机呈现在屏幕的上方或下方的。因此，我们认为，这些希腊语非词不会给被试带来什么影响。

在本实验中，着重探讨的是被试在有意识迫选任务条件下，道德词是否对后续的空间位置选择产生影响。因此，本实验数据分析中，对个数的分析比反应时更加直观和有意义，所以就没有分析反应时数据。

意见 4：实验三、作者在这个实验中使用了圆环随机出现在屏幕的上下方，而在实验一中，作者使用的是 p/q 随机出现在屏幕上下方，请问为何在实验三中选取了圆环而不是 p/q 随机出现在屏幕的上下方，这一改变会不会影响研究整体对作者想要考察的内容的统一性。具体而言，实验一使用 p/q，被试需要对 p/q 的不同字母、以及这两个字母的位置进行识别，而在实验三中被试只需要对圆环（而非不同字母或者图形）的位置进行识别，这一不同是否可能会造成实验一和实验三，被试在反应时的变量上，不同的显著性差异，进而影响对始源域和目标域之间映射关系，尤其是信息映射不对等的理解。

回应：虽然实验 1 使用 p/q 而实验 3 使用圆环，但不会造成反应时的差异，因为在这两个 Stroop 范式中启动项和反应项是不同的，所以这二者不好进行比较。在实验 1 中启动项是词汇，目标项是上下空间位置，被试对呈现在屏幕上下的 p/q 判断反应时作为因变量；而在实验 3 中是通过圆环的上下位置来代表启动项，所以实验程序中要让被试首先对圆环在屏幕中的上下位置做出判断，然后观察上下位置对道德词语义判断的影响，因变量是词汇判断的反应时。这里并不关心圆环本身，而是关心圆环呈现的位置是上还是下。

5、意见 5：实验三中，作者的其中一个重要发现是抽象的道德概念向空间位置进行自动化的映射。结合作者在文中引用的一篇 2013 年发表在《心理科学》的题为“为什么快乐在‘上’，悲伤在‘下’”的文献，是否可以理解上/下的概念与效价普遍相关，而非道德/不道德的概念。是否可以理解为正性的效价（如道德、快乐）与上有关，而负性的效价（如不道德、悲伤）与下有关？作者是否可以利用现有的数据说明上/下与道德/不道德的概念

之间的特异性关系？另外为何实验三的结果中，空间位置的主效应显著，对上的反应时普遍快于对下的反应时？

回应： 谢谢专家对本研究的深入思考和提出的两个问题。

(1) 是否可以利用现有的数据说明上/下与道德/不道德的概念之间的特异性关系。我们的回答是不能的。我们知道，道德概念的构成成分至少有道德认识和道德情感（当然有知情意行四成分说），因此本实验中使用的道德词和不道德词，一定含有道德情感效价。如果将道德的情感效价分离出来，这种道德概念可能就不称其为道德概念了。因此本研究中道德概念是含有情感成分的道德概念。但是我们在另一项“欺负儿童”的道德概念垂直空间隐喻的研究中，发现他们的道德概念垂直空间隐喻与正常儿童是反转的，调查其原因发现主要是道德认知造成的。目前我们也在尝试做道德概念垂直空间隐喻中道德认知和道德情感孰轻孰重、孰先孰后的研究。但将道德认知和道德情感完全分离出来，探讨不含有道德情感的“道德”与上下的特异性关系尚没有好的办法。

(2) 在极性研究中，正效价（如上、道德、高兴）等反应时确实快于负效价，但我们并不关注主效应，而要研究的是“道德是上，不道德是下”。极性方面有一项研究，Lakens, D. (2012). Polarity correspondence in metaphor congruency effects: Structural overlap predicts categorization times for bipolar concepts presented in vertical space. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 38(3), 726–736.

意见 6： 在所有的实验中，作者并未明确指出被试是否均为右利手，或者没有限定被试的左/右利手？而由于被试在实验中需要按键作反应，而往往不用利手的被试会对不同位置的按键有不同的反应时，作者是否考虑了这个因素对实验结果的影响？

回应： 本研究的 4 个实验中，除实验 2 为单一手上下按键外（因为实验 2 记录个数而不是反应时），其余三个实验均需被试左右手按键，每一实验在进行实验设计时，所有实验条件在左右按键上均在被试间进行了平衡。此外，在正式实验开始前，被试均需进行一定次数的练习实验，确保被试对实验的熟悉性，以此来避免被试利手因素对实验结果造成的影响

意见 7： 请作者对文中的一些拼写再行检查，比如 E-Prime 的软件，不是 E-preime；英文摘要中的大小写和语法问题，如第一段第一句应为过去时，Four studies investigated；最后一

段, All together, The results 应为 the results; 第二行的“bidirectional, It was”应为 bidirectional. It was 等等。

回应: 谢谢! 我们已做了相应修改, 并对全文进行了再次检查和修改 (见正文红色字体)。

第二轮

审稿人 1 意见:

意见 1: 作者对审稿意见进行了很好的思考和回应, 对文章也做了一定的修改, 但还有一些回应并没有体现在文章的修改中, 比如删除被试的具体标准、错误反应的处理方法、实验程序上的具体细节等, 建议作者再进一步的思考先前提出的建议, 在文章中也进行更好的修改, 最好每一个回应中都明确在文章中多少页进行了相应的补充或修改。另外, 有些意见, 希望作者不是简单的回应“我们在文章中都有交代, 在这里再说明一下”, 而是仔细思考一下文章中怎么表述可以让读者更清楚要说明的问题。

回应: 我们认真阅读了审稿专家再次提出的修改意见, 在回应一审意见时有些问题做了说明而确实没有在文中进行修改。根据二审意见和上一次只做了说明而没有修改的地方都做了修改, 并将修改处用蓝色字体做了标记, 请审阅。

1.关于删除被试的依据

在本实验 1 中, 被试有两个任务——词义判断和字母分类判断, 只要其中一个任务判断错误, 就不能清晰说明是哪些因素造成的对实验结果的影响。因此, 删除的 2 名被试是依据被试在这两个任务中任何一个任务的正确率低于 80%来进行删除的。这样, 就整个被试来说, 在词义判断和字母分类任务中正确率均达到 80%以上。具体的修改见正文第 6 页, 页码不含回应和自检报告等内容, 下同。

2.关于被试的删除标准

这方面似乎没有统一的规定, 根据已有文献, 删除被试的范围一般在其反应的正确率低于 65%-80%, 就是说至少要保证每个被试反应的正确率在 65%以上, 现列几篇国内外有关的几篇文献:

Kousta, S.-T., Vigliocco, C., Vinson, D. P., Andrews, M., & Del Campo, E. (2011). The representation of abstract words: Why emotion matters. *Journal of Experimental Psychology: General*, 140(1), 14-34.

Xie, J. S., Huang, Y. L., Wang, R. M., & Liu, W. J. (2015). Affective valence facilitates spatial detection on vertical axis: Shorter time strengthens effect. *Frontiers in Psychology*, 6,

1-11.

Xie, J. S., Huang, Y. L., Wang, R. M., & Chang S. (2014). The Mechanism of valence-space metaphors: ERP evidence for affective word processing. *PLoS ONE*, 9, 1-16.

张积家, 王娟, 刘鸣. (2011). 英文词、汉字词、早期文字和图画的认知加工比较. *心理学报*, 43(4), 347-363.

3.关于错误反应数据的处理

删除的依据如同删除整个被试的数据。因此，在本实验 28 名被试的数据中，删除 2 名被试（正确率低于 80%）的所有实验数据后，其余 26 名的被试的数据在统计分析时，词义判断和字母分类判断两个任务的错误反应数据均予以删除，就是说一个实验试次中两个判断任务只要出现一次错误该试次数据就予以删除。（修改处见正文第 6 页）

4.关于实验程序

本实验中，在词义判断环节，实验材料呈现在屏幕中央，被试自定步速阅读，按键反应后消失。同时设定最长反应时限为 3000ms，若在 3000ms 内被试不做反应，实验材料同样消失，在进行统计计算时，该试次按错误反应计算。（具体在文中的修改见第 6 页）

5.在介绍道德概念的相关研究时，转折比较突兀，建议进行修改润色

修改处见正文第 1-2 页。

6.关于增加不同研究范式对实验结果的影响的建议。

对这一部分我们原来做了修改，增加的部分见第 3-4 页。

7.关于四个实验之间的实验逻辑

接受审稿专家的意见。在第一次回应时，我们只是对四个实验之间的逻辑关系做了综合说明，而没有在正文中作出补充和修改。这一次在“引言”的最后一部分补充了这方面的内容，见第 4 页。

8.关于写作规范细节

对写作上的一些细节问题，原来在正文中做了修改，是用红色字体标出的。

9.关于实验 1 是研究目的

原来在正文中做了修改，见第 5 页

审稿人 2 意见：

意见 1： 对于之前的第五个问题：“实验三中，作者的其中一个重要发现是抽象的道德

概念向空间位置进行自动化的映射。结合作者在文中引用的一篇 2013 年发表在《心理科学》的题为“为什么快乐在‘上’，悲伤在‘下’”的文献，是否可以理解上/下的概念与效价普遍相关，而非道德/不道德的概念。是否可以理解为正性的效价（如道德、快乐）与上有关，而负性的效价（如不道德、悲伤）与下有关？作者是否可以利用现有的数据说明上/下与道德/不道德的概念之间的特异性关系？另外为何实验三的结果中，空间位置的主效应显著，对上的反应时普遍快于对下的反应时？”

对这个问题，作者的回复是道德概念与道德情感很难分离。的确要把道德概念和道德情感成功地剥离开非常困难，那么基于这样的困难，作者是否应该考虑对本文章的中讨论和结论做出相应的说明。结合 2013 年的《快乐在上，悲伤在下》的研究结果，我们不难推出“上”很可能与积极的，正向的内容有关，而“下”则与消极的，负向的内容有关，这并不仅仅局限在道德情感上，而作者在回复中的“主效应”是否仅仅是道德，还是效价，亦或是道德和效价的结合，作者应该结合已有的文献和研究，详加讨论和分析。

回应：接受审稿专家的意见。我们分两个方面来加以说明。其一是道德概念垂直空间隐喻中的道德性和情绪性问题。与情绪效价有关的一些空间隐喻的研究(Meier & Robinson,2004; 吕军梅, 鲁忠义,2013)，都发现情绪效价和垂直空间维度之间存在隐喻联结，表现为“积极-上、消极-下”。我们的实验材料是道德概念，从道德概念的成分看，它必然含有情绪效价。因此本研究所考察的道德概念不单纯是道德性这一维度，也包含了道德概念的组成成分——情绪。因此本研究的结果很有可能是道德性和情绪性共同作用而致。当然，把道德概念垂直空间隐喻中的道德性和情绪性分离开来，研究其在道德概念垂直空间隐喻理解中孰轻孰重、孰先孰后将是今后研究的一个重要课题，我们也在尝试这方面的研究。（文中的具体修改见第 17-18 页）

其二是，对研究结果的解释还要考虑到极性的影响。Lakens(2012)指出，维度组成的极性对立(如，道德-不道德或上-下)存在一种固有的极性优势效应：由于正极在语言中往往被更频繁地使用，正极(如道德、上)的心理加工比负极(不道德、下)具有更大优势。因此本研究的实验三中被试对上的反应时普遍快于对下的反应时可能就是这种极性的表现。而当概念和感知维度重合时(正极词呈现在上方和负极词呈现在下方)会比其不重合时(正极词呈现在下方和负极词呈现在上方)的加工处理更有优势。但是，极性又不能完全解释隐喻一致效应，例如 Hill 和 Lapsley(2009)研究就发现不道德词出现在空间下方时，反应时快于出现在上方的情况。（文中的具体修改见第 18 页）

据此，我们在结论中也增加了一条“道德概念和垂直空间的隐喻联结可能是道德和效价共同作用的结果，同时也可能受到极性结构重叠的影响。”见第 18 页

第三轮

审稿人 1 意见：作者根据审稿意见进行了认真修改，同意发表。

审稿人 2 意见：感谢作者的修改，没有其他意见。

编委意见：本文经过多轮审稿，审稿人非常仔细认真地提出相关的问题，作者根据这些问题对文稿进行了补充和修正，在理论背景、研究逻辑、讨论方面都较为完善。目前存在的主要问题是经过多轮修改后，文章显得过于冗长，文章主体部分字数达到 16000 余字，作者需要进行删减，在保证重要问题得到论述的同时精简文字，尤其是前言和讨论部分，数据呈现方面，一些过于简单的数据表可以合并到文字中，直接在条件比较时呈现平均数和标准差。作者进行适当的删减后，论文可以发表。

回应：根据编委的意见，我们对文章进行了删减。一是对不很重要的段落直接做了删除（引言部分）；二是对全文的文字进行了精炼，总的删除了 1500 多字。但数据表还是保留了，主要是想如果发表的话，对读者全面了解数据有帮助。删减后的稿子同本文作者去年在《心理学报》发表的一篇文章的字数相当。

第四轮

主编意见

意见 1：请作者报告每个检验结果的效应估计值。

回应：实验 1 的各项实验结果的统计检验均不显著，因此不需要效应量 η^2 ，实验 3 原来有 η^2 已实验 3 和实验 4 的 η^2 补上了。

意见 2：作者在正式实验前进行了 16 个试次的练习，是不是练习次数太少？请作者最好能回应一下设置如此少的次数的理论依据，还有，作者在每次实验数据的删除中，删除的数据占总数据超过 6%，是不是删除的数据有点太多，会不会影响结果的准确性。

回应：关于练习 16 个试次的问题。练习多少次是根据任务的难度估计出来的，是一种经验

数据。本实验中，应该是在正式实验开始前每位被试至少做 16 个试次的练习，若被试不能充分理解实验要求和实验操作需要再次练习，直到对实验操作有充分理解之后才开始正式实验。

删除数据 6%，删除的实验数据为被试的错误反应数据，由于每个实验试次都为双任务判断，导致部分被试错误反应次数较多，但进入实验统计的数据均为被试正确反应的数据，不会影响实验的可靠性。