

《心理学报》审稿意见与作者回应

题目：道德概念的黑白隐喻表征及其对道德认知的影响

作者：殷融 叶浩生

第一轮

审稿人 1 意见：该文探讨了道德概念与黑白颜色概念的隐喻联结存在的心理现实性，具有重要的意义，但文章存在着一些需要解释或修改的问题。

意见 1：三个实验之间的逻辑关系问题。作者认为从实验 1 到实验 2 是从词汇水平到信息加工水平的变化，而信息加工是较为复杂的任务。好像这是任务复杂程度的变化，如果仅仅是复杂程度的话就似乎没有什么多大的意义，并且前两个实验跟实验 3 是什么关系也不清楚，如果按作者的看法是想从黑白经验角度来探讨，那前两个实验设计中的黑白线索不也是经验吗？总之没有看到三个实验之间的内在逻辑。

回应：将三个实验的逻辑关系进行了更为详细的整理与解释。实验 1 主要是探讨在概念加工任务中，人们关于黑白颜色与道德词和不道德词的匹配或表征倾向；实验 2 则是道德判断过程中，颜色刺激是否会对道德判断结果产生与隐喻映射方向一致的偏向性效应。实验 3 则将平面的黑白颜色扩展到环境亮度范畴，探讨亮度知觉对道德认知判断的影响。之所以进行实验 3，还有一个原因是本研究认为白好黑坏的隐喻表征关系是由人们关于“光明”与“黑暗”的经验塑造而成的，因此，道德概念颜色隐喻表征中也包含亮度范畴，亮度知觉与道德认知间的隐喻干扰效应同颜色知觉与道德认知间的隐喻干扰效应应具备共通性。关于实验 3 的意义也进行了更为详细的解释

意见 2：作者把实验 1 看做是词汇水平，把实验 2 的故事理解看做是信息加工水平是没有道理的，因为这两者都可以看做信息加工。

回应：修改后，实验 1 主要是词汇加工水平，实验 2 主要是道德判断水平。

意见 3：实验 1a 设有中性词，实验 1b 为什么就不设中性词了？

回应：实验 1a 与实验 1b 主要测查的变量不同。实验 1a 本质上是要求被试将黑色或色与词汇相匹配，因变量是被试选择匹配的比率，在这种情况下，实验材料中可以包括中性词，以中性词作为填充项一方面可以干扰被试，防止对实验目的进行猜测，另外也可以对实验结果进行更好地解释。但实验 1b 的任务是要求被试对词性做出快速判断反应，因变量是被试的反应时，如果同时设置三种类型的词汇，在按键反应上不能进行很好的平衡，容易发生其他的阻碍效应，因此只有两种选项。即反应时任务中常见的 Yes（道德词）/No（不道德词）选择模式。

意见 4：实验 2 为什么把不道德故事和两难故事这两个放在一起研究？如果按照此逻辑为什么不设道德故事？

回应：从判断难度角度考虑，不道德故事和两难故事是两种不同水平的材料，且在关于道德判断的研究中经常使用这两类材料，因此将道德故事和不道德故事作为了实验 2 的实验材料。而美德故事从判断难度角度来看和不道德故事属于同一类，因此没有使用。在文中增加了关于此内容的说明。

意见 5: 实验 2 的统计分析部分说,“分析显示,被试对于不道德故事评分的组间差异不显著, $t(50)=0.86$, $p=0.395$ ”;“黑色线索组被试对不道德故事的的不道德评分显著高于白色线索组被试的评分。”这是矛盾的

回应: 此处书写有误,已进行了更正。

意见 6: 把什么结果都看做是干扰效应是没有道理的,因为有匹配的一致性效应存在。

回应: 对具体用词进行了更正,如将干扰效应改为“认知偏移效应”或“隐喻一致性效应”。

意见 7: 实验 2 的一个结果是,依然出现了颜色知觉对道德认知判断的干扰效应,作者从两个实验的组内设计和组间设计来说明似乎没有道理。如果按照这种解释为什么不道德故事没有受到颜色的影响呢?

回应: 删去了关于组内、组间设计的讨论。对“为什么不道德故事没有受到颜色的影响”进行了更为详细的解释。

意见 8: 该文是研究道德的颜色隐喻的,但讨论部分缺少从这个角度来解释实验结果。在综合讨论中一段谈到了概念隐喻理论,但用的人际情感的例子。

回应: 在综合讨论部分添加了相关内容,从道德概念颜色隐喻表征的角度对实验效果进行了解释;更详细的探讨了道德概念黑白隐喻可能的形成机制。

意见 9: 国外也有这方面的研究,讨论中应进行比较

回应: 在文中各个部分都增加了一些实证研究范例,以进行对比或总结。同时,新增加了研究展望一部分,其中结合了很多关于概念隐喻理论的研究进行讨论。

审稿人 2 意见:

意见 1: 在实验 3 中,研究者认为“如果环境的亮度会对个体的道德认知判断产生影响,那么在个体思考道德相关问题时(合作者是否会公平地对待自己),黑暗视知觉体验会引发更多的负性道德特质联想”,这是一种可能的理论视角。另外一种可能性是,黑暗而匿名的环境可能引发人类的竞争性和自私性,因而在暗箱操作的情况下,做出自私自利的行为代价较小而收益较高。其实这一点,作者在最后的讨论中也有所提及。

回应: 修改后,在实验 3 的讨论部分(5.4)对这种可能性进行了讨论。同时,也结合了“暗增加自私自利性”探讨了塑造白好黑坏隐喻形成的因素。

意见 2: 作者在实验 3 的讨论中说“不同实验环境组被试在关于情绪的检查问题上评分无显著差异,这证明同明亮环境中的被试相比,在黑暗环境中的被试在实验过程中并没有更强的负面情绪感受。因此排除了环境差异引发被试不同的情绪感受、从而造成被试的认知判断差异这一可能性”。情绪所以没有影响,可能是因为研究者的实验设置并没有引起足够强烈的情绪,或者明暗操作对于情绪的影响本身就有限,远不如情绪相关线索比如看恐怖片或看恶心图片来得强烈。当然,作者的主要研究目的不是考察情绪的影响,而是排除明暗操作可能导致情绪差异,而这种差异导致了道德判断的差异这样一种替代假设。没有出现情绪影响的结果,除了情绪强度较弱之外,还有可能是有的相关情绪没有进行测查。建议作者在讨论时考虑这些可能。

回应: 在分析部分增加了对情绪因素的讨论,在总结与展望部分也进一步对此问题进行了明确。

第二轮

编委意见：作者根据审稿专家的意见做了适当的修改，或进行了解释与说明。修改后的稿子质量明显提升。但是，尚存在一些问题，需要修改后再审。

意见 1：研究的实验逻辑关系的问题。审稿专家 1 已经明确提出，“前两个实验跟实验 3 是什么关系也不清楚”，“没有看到三个实验之间的内在逻辑”。虽然作者做了一些修改或说明，但是目前的修改后的稿子这一问题依然没有很好解决。存在这一问题的根本原因是作者将两条逻辑线索混淆在一起使用。一条逻辑线索是“从黑白到明暗”，“从刺激材料的属性到刺激背景的属性”，以探讨道德隐喻的本质；另外一条逻辑线索是“从词义的道德判断到两难情境的道德判断再到社会关系中不公正对待可能性判断”，以探讨道德隐喻的功能，即对道德认知的影响。这两条线索的关系处理不好，自然研究的实验逻辑关系的问题就解决不好。期望作者好好理清这两条线索的关系，一方面在 3 个实验的研究目的的陈述中讲清楚，最重要的是在总的讨论中将两条线索分开讨论，各列一个标题分别讨论。

回应：根据编委的意见厘清了两条线索，在三个实验的目的及陈述中有所体现，同时，更主要的是对总的结论讨论进行了修改，将两条线索各列了一个标题分开讨论，一条线索是“在隐喻表征层面上，本研究的系列实验首先验证了道德概念黑白隐喻表征的心理现实性，进而探讨了塑造这一机制的主要因素”，另一条线索是“从社会心理认知研究的层面看，本研究的系列实验则主要探讨了不同的刺激因素方式（平面的黑白颜色、环境中的明暗亮度）对不同类型道德认知判断任务的影响。”

意见 2：关于研究结果的推论问题。该问题与上述问题 1 密切相关。该研究的 3 个实验的行为指标的结果是客观的，但是从行为结果来进行心理层面的推理的时候却是主观的，所以要特别注意推论要有理有据。具体地说，黑白（或明暗）的隐喻是客观存在的，但是这种隐喻是道德认知的一个成分（必要条件），还是与道德认知密切联接的相关成分（充分条件），这一点上应该仔细考虑。作者的实验 2 和实验 3 明显是针对“黑白（或明暗）的隐喻对道德认知的影响”（观点 2），作者在摘要中也说“采用不同的研究方法在不同水平上考察道德概念的颜色隐喻表征，探讨了颜色知觉对道德认知的影响”，但是作者在文章中却多次表达出“这种隐喻是道德认知的一个成分”的观点，例如，作者在实验 2 的讨论中说“黑白颜色线索会激活关于道德范畴的黑白颜色隐喻表征，并对被试的道德认知判断产生同化的影响作用。”请作者仔细考虑：实验 2 和 3，究竟是反应冲突（隐喻关系是道德认知的必须成分）还是启动效应（不是必须成分但是有自动联接——内隐联想关系）。

回应：对每个实验的结论和讨论的表述进行了修改，强调实验效应是由于被试无意识的联结倾向，另外特别增加了总讨论的 6.1 部分的最后一段，强调黑白颜色范畴并不是表征概念范畴的必要成分。

意见 3：关于标题与关键术语的问题。文章的标题是“道德概念的颜色隐喻表征及其对道德认知的影响”，但是实际上 3 个实验并未涉及色彩因素，似乎标题应该改为“道德概念的黑白隐喻表征及其对道德认知的影响”，以免引起阅读时的误解。（就像一般人认为黑白电视不是彩色电视一样。）文章中的相关的关键术语似乎也应该做改动，例如，“颜色隐喻效应”似乎应该改为“黑白隐喻效应”，“颜色知觉”应该改为“黑白（或明暗）知觉”。

回应：将标题和文中部分术语进行了修改，突出了“黑白”。

意见 4：关于统计方法的问题。实验 1a“以翻译词类型（道德词，不道德词，中性词）和备选词颜色（白色，黑色）为自变量，以被试选择的百分比为因变量进行重复测量方差分析”，

似乎是不恰当的（这样会导致“翻译词类型”主效应为 0，作者在文章中报告的“备选词颜色的主效应不显著”或许也与这有关系）。如果以“不道德选择的频次（或百分比）”作为因变量指标就可以用这种方法分析。如果一定要以被试选择的百分比为因变量进行检验，就应该进行多元卡方检验。

回应：对实验 1a 的统计方式进行了修改，只计算被试选择白色词汇的比率，使用了单因素方差分析。

意见 5：讨论的深度和恰当性的问题。和问题 1、2 和问题 3 密切相关的一个问题是，要重新考虑各个实验的讨论和最后的总讨论，要思考和回答“两条逻辑线索”上讲 3 个实验的结果分别说明了什么规律，还要结合多次出现的交互作用来分析这种规律性（例如，实验 2 出现“不同组别被试对不道德故事的评分不存在组间差异”的结果，要结合“两条逻辑线索”的实验逻辑来思考和讨论）。此外，讨论的恰当性问题也要考虑，作者在实验 1b 中，报告“不道德词的判断反应时在词汇颜色水平上差异边缘显著， $t(23)=1.64$ ， $p=0.12$ ”的时候，不顾下错误结论的风险（ $p=0.12$ ）认为“实验结果同实验 1a 相符合，进一步证明，道德概念与黑白颜色范畴的隐喻关联具有心理现实性。”并在随后的 3.4 讨论中进一步确认。实验 3 也存在这样的风险（ $p=0.059$ ），作者说“而当假设由计分的合作者决定酬劳分配时，不同实验条件组被试对酬劳预期的组间差异达到边缘显著， $t(53)=1.93$ ， $p=0.059$ ，暗环境组被试对自己所获得酬劳的预期会少于明环境组被试的预期。暗环境组被试会比明环境组被试更倾向于认为自己在酬劳分配中会受到不公正的对待。”并在“5.4 讨论”中进一步确认。

回应：结合两条线索对实验结果进行了讨论。另外，特别对实验 1b 和实验 3 进行了补充，作者认为，实验中的一些效应不显著或边缘显著是由于被试样本容量不足引起的，因此，对实验 1b 又增加了 7 位被试，对实验 3 增加了 10 位被试，实验程序及实验环境、场所与之前实验一致。补充实验后一些差异效应达到了边缘显著或显著性。对相应的讨论、结论都进行了修改。

意见 6：无关变量的严格控制问题。一些至关重要的无关变量的控制应该更清楚、更全面地交代。例如，实验 3 中，暗适应的问题是否进行了实验控制（或，是否需要控制）？实验两个阶段（任务测试阶段，问题回答阶段）都是在不同明暗条件下进行的，不同明暗条件对任务测试阶段的影响有多大（是否可以把最后的实验效应都归结为不同明暗条件对“问题回答阶段”的影响）？被试任务测试阶段的实际成绩应该报告，被试主观评定的“客观程度”或“主观程度”有多大，是否受不同明暗条件的影响？再例如，每套材料都包括两组 8 张刺激材料，实验中两套图片的呈现顺序是如何安排的，是否会出现顺序效应的干扰？

回应：对变量的控制进行更清楚、详细的交代陈述。暗适应与明适应的控制文中进行了说明，对材料的呈现顺序也进行了更为详细的说明，如实验 2 中 16 张图片按照完全随机的顺序进行呈现。另外，对实验 3 中被试在推理测试中的成绩进行了报告和统计分析，同时计算了被试的实际成绩与主观成绩标准化分数后的差值，对差值进行了独立样本 t 检验。并根据结果增加了相应的讨论。