

《心理学报》审稿意见与作者回应

题目：灰姑娘能变成白雪公主吗？感知到的信任对他人面孔表征的影响

作者：李庆功，方激，胡超，石德君，胡晓晴，傅根跃，王乾东

第一轮

责编委意见：

我很感谢两位审稿人给出了专业的意见。他们认为您的稿件有意义有创新，但还存在一些问题。

意见 1： 其中，一位审稿人指出实验三使用的是与实验一和实验二相同的面孔，如何在实验一和二中体现了个体心理表征，但在实验三中体现了群体心理表征？我认为这是一个十分重要的问题。而且三个实验使用了相同的面孔，得到相关的结果似乎并不奇怪。建议您使用其他更多的面孔重新做一下实验三，采用另外一位审稿人给出的新的指标，以得到更确定的结论。

回应：感谢编委的意见。本研究中获得的是个体还是群体的心理面孔表征是基于实验任务的操纵。实验 1(修改后的论文合并了实验 1 和实验 2)首先给被试观看个体的面孔照片，然后让被试回忆并从一系列两张叠加噪音的面孔中选出更像该个体的面孔，因此得到的是被试对于该个体面孔的心理表征。而实验 2(原论文中的实验 3)的被试并没有观看个体面孔照片，也无需再认个体面孔，他们直接在添加了噪音的面孔中选择可信面孔，因此是对于可信/不可信群体的面孔表征(可信/不可信群体的面孔原型)。我们已在“2 实验 1”和“3 实验 2”的第一段增加了相应的描述。

关于“使用相同面孔得到相关”这一问题，我们在数据分析中，首先提取出两个实验产生的分类图像，并对其面孔区域内的噪音模式的像素亮度值进行向量化，随后对其进行皮尔逊相关检验(Dotsch & Todorov, 2012)。在这一过程中，我们并没有直接对分类图像的像素值做相关(所有同性别的分类图像均由相同的面孔底片叠加各自的噪音模式产生)，这样就能降低相同面孔底片带来的冗余相关。在上一版的论文中，我们未对这一分析过程进行详述，现已在“3.1.4 数据处理”部分第一段加以强调。此外，在上一版论文中，诸如“被描述为可信(或不可信)的目标人物的面孔表征与可信(或不可信)群体的面孔表征有更多的相似性”的写法是有误导性的，会让读者误以为我们直接对分类图像做相关分析(实际上我们用的是分类图像的噪音模式做相关)，因此相关语句已改成如“被描述为可信(或不可信)的目标人物的面孔表征特征与可信(或不可信)群体的面孔表征特征有更多的相似性”。之所以选用“特征”这个词是因为噪音模式的叠加能改变面孔底片的样貌，噪音模式可以认为是不同面孔表征独有的特征(也已在“3.1.4 数据处理”部分第一段加以说明)。

但是，毫无疑问，面孔底片的选择会影响面孔表征的样子(比如，如果换一张面孔底片做出来的分类图像绝对与当前研究得到的会不一样)，使用反相关图像分类技术获得的面孔分类图像只能算是真实心理表征的近似值，这也是该技术的一个局限(Dotsch & Todorov, 2012)。如何将混杂的因素分离并得到真实的心理表征需要进一步的探索。基于此，我们非常认可编委的意见，选用相同的面孔底片会是本研究的一个局限，但是我们也不认为这会对结果的解读造成影响。采用反相关图像分类技术做的研究往往关注的是类别间的相对差异(比如这一类别的面孔表征是不是比另一类别面孔表征更有吸引力)，而不是某类面孔表征的

绝对样貌。在本研究中，我们主要关注的是可信/不可信操纵造成的个体面孔表征的差异或是可信/不可信个体与可信/不可信群体面孔表征相似性的差异。虽然同性别面孔的底片是一样的，但最终差异还是显著的，说明心理表征的变化主要来源于对于目标面孔可信度的描述不同。我们在讨论中加入了对此问题的说明，详见“讨论”部分第倒数第二段。此外，获取个体和群体面孔表征时使用相同的面孔底片也具有相对优势。试想如果使用两个不同的面孔底片，他们的关键区域位置（比如眼鼻嘴等，这些区域往往是面孔识别的诊断区域，见我们按照审稿专家意见新增的“2.2.4 面孔识别诊断区域”部分）没对齐或不重合，做噪音模式像素间的相关或是相似性分析也会失去意义。基于以上原因，我们没有使用其他更多的面孔重做实验 2。

最后，我们也非常感谢审稿专家提出的结构相似度指标，我们进行了相应的计算(方法描述见“3.1.4 数据处理”这一部分)，并且将结果放到了修改后的文稿中(见表 3)。这一指标的结果与我们的预期一致。

意见 2: 另外，您的研究方法描述得不是十分具体，两位审稿人给出了具体的疑问或修改建议。建议您相应进行修改，并思考假如有人要重复这些研究，现有方法的描述是否足够让他重复。

回应: 非常抱歉关于方法的描述不够详细，我们已在“2.1 研究方法”以及图 2 的流程图上增加了许多细节，相信修改后的研究方法可以被其他研究者很好的重复。

意见 3: 除了两位审稿人提出的问题和建议外，我也有一些其他疑问。

结论说实验一和实验二的结果类似，也就是男性女性的结果类似。但从表 2 看，男性在不可信群体的那两个相关系数的关系，与女性在不可信群体上的两个相关系数的关系，似乎是相反的。对于女性来说，可信个体与不可信群体的相关甚至大于不可信个体与不可信群体的相关。请作者核对是否是数据错误还是其他？

回应: 感谢编委的审阅和指导。经核实，表 2 的数据确实有问题。由于我们的疏忽，无论是男性还是女性目标面孔，不可信个体和可信群体之间的相关值与可信个体和不可信群体之间的相关值互相放错了位置。我们已修正这个错误，并重新计算相关的差异(见表 2)。新的结果发现针对男性目标面孔，结果与之前类似；但是，针对女性目标面孔，不可信个体与可信群体的相关却大于不可信个体与不可信群体的相关，与我们的预期相反。在编委和审稿专家的建议下，我们进一步使用结构相似性指标计算了各个面孔表征特征之间的结构相似性(见表 3)，结果与我们的预期一致。我们对女性目标面孔这一矛盾的结果做了相应讨论，提出了两种解释：第一，正如审稿专家提到的可能是因为对图像像素值向量化后求相关会丢失图像像素区块之间的空间联系，导致它不能很好地表征图像之间的相似程度。第二，人们会认为女性要比男性更可信(Buchan et al., 2008; Dong et al., 2018)，因此，可能也会把更多的可信特征添加到不可信女性目标面孔上，这也可能是导致女性结果并不稳定的原因。详细见“讨论”部分第四段。

此外，我们也仔细核对其他数据，并无类似的错误。

意见 4: 实验一和实验二的被试的写法让人疑惑，应写清楚究竟每组（男面孔可信、男面孔不可信、女面孔可信、女面孔不可信）被试多少人。

回应: 我们非常同意编委的建议并做了相应的修改。见“2.1.1 被试”这一部分的第一段。

意见 5: 应给出所采用的中等吸引力面孔的具体的评分，以及与 5 的单样本 t 检验结果。

回应: 感谢编委提出的宝贵意见。男女目标人物和男女干扰人物的平均吸引力评分分别为

5.03、5.18、5.13、4.90,与吸引力中间值 5 做比较的单样本 t 检验结果分别为 $t(39)=0.11, p=.910, t(39)=0.76, p=.455, t(39)=0.47, p=.644, t(39)=-0.41, p=.682$ 。我们在“2.1.2 实验材料”部分第一段补充了这些信息。

意见 6: 图 2 中最好在流程图中给出刺激呈现的示例,而不是文字描述。

回应: 感谢编委提出的宝贵意见。我们按照您的意见进行了修改,见图 2。

意见 7: 第一阶段中的目标人物的面孔是“目标面孔”还是“噪音面孔”,还是“面孔底片”?

回应: 抱歉我们的表述引起了疑惑,第一阶段的目标人物的面孔是“目标面孔”。为了避免混乱,我们把“2.1.2 实验材料”部分的目标面孔改成目标人物面孔,并明确注明“目标人物面孔是被试需要观察和再认的面孔(见实验流程)”。此外,我们还觉得使用“噪音面孔”这个词可能会与面孔底片中叠加的“噪音”产生混淆,因此把“噪音面孔”改成了“干扰面孔”。

意见 8: 最好在文中点明为何本文的研究内容跟灰姑娘能否变成白雪公主有关系。

回应: 感谢编委的建议。我们修改了正文最后一句话使之与文章题目相呼应:因此,灰姑娘能变成白雪公主:即使一个人的长相一般,只要拥有美好的心灵,人们心目中的这个人的相貌也会更具吸引力。

.....

审稿人 1 意见:

本文采用反相关技术探讨自上而下的因素如何塑造心理面孔表征。研究问题有意义,但由于一些细节和概念不明,建议进一步改善。

意见 1: 心理表征的概念是什么?基于物理特征评价的可信度和吸引力不是心理表征的内容吗?

回应: 感谢审稿专家提出这一直接关系到本文主题的问题。我们在正文中增加了相应的描述介绍心理表征的概念。心理表征是指外部事物在心理活动中的内部再现,它一方面反映客观事物,另一方面又是心理活动进一步加工的对象(见“1 引言”部分第 2 段)。根据这个定义,基于物理特征评价的可信度和吸引力不是心理表征的内容。

意见 2: 噪音面孔的选择决定了合成后的面孔特征,换个噪音面孔很可能改变了合成的面孔,那么如何确定是目标面孔的心理表征?噪音参数选择对最终合成的心理表征有何影响?因此,反相关技术构建的心理表征需要更多的澄清或补充。

回应: 我们非常同意审稿专家的观点:噪音的选择确实很有可能影响面孔表征的样子。使用反相关图像分类技术获得的面孔分类图像只能算是真实心理表征的近似值,因为面孔底片以及噪音的选择会影响面孔表征的样貌,这也是该技术的一个局限(Dotsch & Todorov, 2012)。因此,如何将混杂的因素分离并得到真实的心理表征需要未来研究进一步的探索。但是相比于面孔底片,噪音的选择对面孔表征样貌的影响要小很多。Dotsch 和 Todorov(2012)的研究使用了两组随机产生的噪音(每组都是 300 对正负噪音),发现结果很一致。这可能是因为反相关分类图像技术往往需要非常多的试次进行平均(文献中一般都是多于 300 试次,本研究用了 640 个试次),这降低了噪音选择的影响。此外,采用反相关图像分类技术做的研究往往关注的是类别间的相对差异(比如这一类别的面孔表征是不是比另一类别面孔表征更有吸引力),而不是某类面孔表征的绝对样貌。本研究主要关注的是可信/不可信操纵造成的个体

面孔表征的差异。在实验过程中，我们选取的同性别面孔底片是相同的，同时，在底片基础上叠加的噪音也是相同的。最终得到的结果差异显著，这说明心理表征的变化主要来源于对于目标面孔可信度的描述不同。我们在讨论中加入了对这一问题的说明，见“讨论”部分倒数第二段。

我们也同意审稿专家提出对反相关图像分类技术做进一步补充的建议。我们在“引言”部分第四段中补充了反相关技术的介绍：为了将面孔的心理表征可视化，本研究使用了反相关图像分类(reverse correlation image classification technology, RCIC)技术(Dotsch & Todorov, 2012; Gosselin & Schyns, 2003)。该技术是数据驱动的，需要被试在多轮试次(往往至少 300 试次)中从两张模糊的面孔(同一张面孔底片叠加不同的随机视觉噪音)中选择最能代表目标个体或特定社会群体的面孔。通过在许多试次中做出这些选择，被试本质上提供了与他们的心理表征相关的图像特征的信息。对被试选择的模糊面孔的噪音叠加平均后与面孔底片合成即获得了被试的分类图像。由于该图像显示了驱动感兴趣的社会判断的刺激特征，因此被视为内在的心理表征。此外，在研究方法部分我们也补充了具体操作的详细信息。

意见 3: +-噪音的选择结果有何差别？在试次间如何分配？请给出相关数据。

回应: 感谢审稿专家提出这一问题。每个试次将并排呈现两张模糊的面孔(底片叠加正负相反的随机噪音)，让被试进行选择。我们已经在图 2 的流程图中给出刺激呈现的示例。并且我们也在“实验材料”部分补充采用相反随机噪音模式的优势：相反的随机噪音模式能将所呈现的两张刺激图片间的差异最大化，提升视觉对比效果，从而减少实验试次数(Dotsch & Todorov, 2012)。

意见 4: 三个实验都是相同的面孔，如何既是个体心理表征又是群体心理表征？两者是什么关系？像素向量化做相关如何避免系统刺激带来的相关？

回应: 感谢审稿专家的意见，能帮助我们描述更加清晰。获得的是个体还是群体的心理面孔表征是基于实验任务的操纵。实验 1(修改后的论文合并了实验 1 和实验 2)首先给被试观看个体的面孔照片，然后让被试回忆并从一系列两张叠加噪音的面孔中选出更像该个体的面孔，因此得到的是被试对于该个体面孔的心理表征。而实验 2(原论文中的实验 3)的被试并没有观看个体面孔照片，也无需再认个体面孔，他们直接在添加了噪音的面孔中选择可信面孔，因此是对于可信/不可信群体的面孔表征(可信/不可信群体的面孔原型)。我们已在“2 实验 1”和“3 实验 2”的第一段增加了相应的描述。

审稿专家提到系统刺激可能带来相关，我们也非常同意。为了降低系统刺激的影响，我们在数据分析中，首先提取出两个实验产生的分类图像，并对其面孔区域内的噪音模式的像素亮度值进行向量化，随后对其进行皮尔逊相关检验(Dotsch & Todorov, 2012)。在这一过程中，我们并没有直接对分类图像的像素值做相关(所有同性别的分类图像均由相同的面孔底片叠加各自的噪音模式产生)，这样就能避免相同面孔底片带来的冗余相关。在上一版的论文中，我们确实也是这么做相关的，但我们并没有强调出来，现已在“3.1.4 数据处理”部分第一段加以强调。此外，在上一版的论文中，诸如“被描述为可信(或不可信)的目标人物的面孔表征与可信(或不可信)群体的面孔表征有更多的相似性”的写法是有误导性的，会让读者误以为我们直接对分类图像做相关分析(实际上我们用的是分类图像的噪音模式做相关)，因此相关语句已改成如“被描述为可信(或不可信)的目标人物的面孔表征**特征**与可信(或不可信)群体的面孔表征**特征**有更多的相似性”。之所以选用“特征”这个词是因为噪音模式的叠加能改变面孔底片的样貌，噪音模式可以认为是不同面孔表征独有的特征(也已在“3.1.4 数据处理”部分第一段加以说明)。最后，我们也想说明本研究主要关注的是可信/

不可信个体与可信/不可信群体面孔表征相似性的差异，而不是某个单独的相关值。虽然相同性别面孔的底片是一样的，但最终差异还是显著的，说明心理表征的变化主要来源于对于目标面孔可信度的描述不同。

意见 5: 实验 1 和实验 2 的结果统计分析，有的地方作为一个实验分析，有的地方作为两个实验分析，为什么不统一？

回应: 感谢审稿专家提出这一点。为了统一，我们把实验 1 和实验 2 合并成一个实验，合并后的实验增加了目标面孔的性别这一自变量。

意见 6: 任务 2 是一对面孔评价完所有维度在评价另一个对面孔吗？如何避免多维度同时评价的影响？

回应: 我们确实没写清楚任务 2 的过程，感谢审稿专家提出这一问题。在任务 2 的每个试次中，一组面孔对以及一个评价维度的问题(如您认为下面两张面孔哪张看起来更可信)同时呈现，被试按键选择。这个任务总共有 18 个试次(2 类性别×9 组评价)，试次呈现顺序随机，每个试次中面孔呈现的左右位置也随机。我们在正文中增加了这部分描述(见“第二阶段 评估分类图像”部分)。同时，图 2 也展示了刺激呈现的示例。

.....

审稿人 2 意见:

灰姑娘能变成白雪公主吗？感知到的信任对他人面孔表征的影响 本研究采用反相关分析方法，探究“自上而下”的信任感对他人面孔表征的影响及其潜在机制。本研究，通过实验 1 和 2 的描述个人特质（可信/不可信）操控被试对目标面孔的表征，并得到相对应的可信/不可信的分类图像；以及发现被试对不同的面孔图像会产生不同的评价：可信的面孔往往具有更高的吸引力，被赋予积极的评价；而不可信的面孔吸引力减弱，给人以消极的感受。而实验 3 中，作者对信任度的个体目标面孔表征（实验 1 和 2 所得）与群体面孔表征（实验 3 所得）作了相关性分析，进一步论证了面孔加工过程中，个体对面孔的信任度是通过与信任/不信任的群体面孔表征整合而最终被感知的。文章内容表述较为清楚，科学问题提出合理。研究内容具有一定的创新性，希望未来能在这方面继续深入探索。与此同时，个人认为文章还存在一些问题，建议完善。

意见 1: 首先方法部分能否提供更多的实验细节？不是每个读者都熟悉这个领域，也不一定会查阅(Dotsch et al., 2008)的文章。还有主试者在实验时使用的指导语原文最好能给出。

回应: 感谢审稿专家提出的建设性意见，我们在方法部分补充了更多实验细节，在实验流程图(图 2)上也补充了许多细节（如附上了指导语和刺激呈现示例等）。

意见 2: 实验 1 和 2 中只使用了 1 位个体的目标面孔，那么实验结果是特异与该个体还是对于大部分个体都成立，需要解释和讨论。

回应: 感谢审稿专家的意见。本研究结果能否概化到大部分个体需要未来研究的进一步验证，我们在“讨论”部分倒数第二段进行了讨论：本研究主要探究个体面孔表征受到可信/不可信描述的影响，但又不可能把所有个体研究一遍，因此本研究的结果是特异于实验 1(修改后的论文合并了实验 1 和实验 2)选择的两个个体还是对于大部分个体都成立需要未来研究进一步拓展和验证。特别是本研究选取了具有中等吸引力面孔的个体，因此能不能推广到本身吸引力较高和较低的面孔上值得一探究竟。

意见 3: 图 3 结果是个体结果还是所有受试者平均结果, 需要注明。图 4 同样需要说明。如果能够采用一些统计手段来比较可信与不可信条件下面孔表征的差异, 就更好了。

回应: 图 3 和图 4 (现为图 5) 结果都是所有受试者平均的结果, 我们已在正文和图注中都进行了说明。关于面孔表征的差异, 我们对受试者的主观评价进行了统计检验。但是, 使用客观的基于图像的算法去比较可信与不可信条件下面孔表征的差异, 我们没能想到好的方法, 因此没做这方面的统计。

意见 4: 第 5 页倒数第 4 行“后续分析发现实验 1 和实验 2 的结果一致, 提示本研究探索的现象与面孔的性别无关。”作者在这里应该附上统计学结果。

回应: 基于审稿专家 1 的意见, 我们合并了实验 1 和实验 2, 这句话也由此删除。此外, 在数据分析中, 我们将面孔性别纳入分析, 发现面孔性别并不影响结果。详见实验 1 结果部分的“2.2.3 多维特质评价”。

意见 5: 第 6 页 2.1.1 关于被试的描述不清晰。首先, 在目标面孔的选择阶段中 40 名受试者是否参加了其他实验; 在第二阶段中, 任务一的 40 名受试者和 50 名受试者是否从第一阶段的 155 名受试者中进行筛选产生还是重新招募的, 是否存在可信组和非可信组偏倚? 请说明。此外, “任务一/二”的书写形式与下文及图 2 中 (“任务 1/2”) 应该保持一致。

回应: 感谢审稿专家提出这一问题, 根据审稿专家的意见我们修改了表述使得意思更加清楚。我们已在“2.1.1 被试”这一部分明确写明第二阶段招募的所有被试均未参加第一阶段的实验。并且第二阶段的任务 2 的被试与任务 1 并没有重合, 是重新招募了 50 名被试。此外, 我们也在流程图(图 2)中清晰地表明这些被试是不重合的。我们也修改了“任务一/二”的书写形式改为“任务 1/2”并在全文中统一。

意见 6: 第 7 页第 2 行, “方向”应为“方位/朝向(orientation)”。

回应: 感谢审稿专家的指导, 我们已对该说法进行了相应修改 (见 2.1.2 节第三段)。

意见 7: 第 8 页 2.2.1 可信度操纵的有效性: 这一步作者除了对可信度进行评分, 还增加了“活跃度”这一指标。作者在这一步应该说明为何要使用“活跃度”而非其他。640 个试次做 reverse correlation 是否足够获得可信结果, 需要说明。

回应: 在“第一阶段 获取分类图像”这一部分, 我们已补上选择“活跃度”作为对照的原因: “此外, 为了确保信任操纵不影响与信任无关的特质, 我们也让被试对目标人物的活跃度(1=非常不活跃—9=非常活跃)进行评分。选择活跃度作为对照指标也可以检验另一个自变量(目标面孔性别)所引起的特有心理效应: 我们预期大众会觉得男性比女性更活跃。”

本研究共使用 640 个试次, 数量远多于以往相似的实验范式(如 Dotsch & Todorov, 2012)。同时, 我们采用相反的随机噪音模式也能将所呈现的两张刺激图片间的差异最大化, 提升视觉对比效果, 从而在一定程度上减少不必要的实验试次数(Dotsch & Todorov, 2012)。我们已在“2.1.2 实验材料”部分做了说明。基于以上原因, 我们相信 640 个试次做 reverse correlation 足够获得可信结果。

意见 8: 第 9 页 2.2.2 吸引力评价: “无论是实验 1 还是实验 2 产生的分类图像, 所有被试均选择可信条件下的分类图像更有吸引力。”对于该结果, 作者应提供评分结果和统计学分析的结果加以佐证。

回应: 我们已补上相应的统计分析结果, 请见“2.2.2 吸引力评价”部分。

意见 9: 第 10 页“3.1.3 实验流程和数据处理”: 作者使用“噪音的像素亮度值向量化后求皮尔逊相关”, 应当加上该方法的出处, 似乎是 (Dotsch & Todorov, 2012)? 并简单说明为何使用“像素亮度值”。此外, 对于图像之间的相似程度, 皮尔逊相关也许不是最好的表征方法。因为对像素值向量化会丢失二维图像像素区块之间的空间联系, 或许可以考虑 The structural similarity index measure (SSIM) 之类的图像相似度指标。

回应: 非常感谢审稿专家的建议。我们使用皮尔逊相关的方法确实是因为参考了 Dotsch & Todorov (2012) 的文献, 我们已在正文中引用了该文献。此外, 我们也非常同意审稿专家提到的结构相似度指标可以更好地反应图像之间的相似程度, 我们进行了相应的计算(方法描述见“3.1.4 数据处理”这一部分), 并且将结果放到了修改后的文稿中(表 3)。

另外, 我们需要说明下, 由于我们的疏忽, 无论是男性还是女性目标面孔, 不可信个体和可信群体之间的皮尔逊相关值与可信个体和不可信群体之间的皮尔逊相关值互相放错了位置。我们已修正这个错误, 并重新计算相关的差异(见表 2)。新的皮尔逊相关结果发现针对男性目标面孔, 结果与之前类似; 但是, 针对女性目标面孔, 不可信个体与可信群体的相关却大于不可信个体与不可信群体的相关, 与我们的预期相反。审稿专家建议的基于结构相似性指标的结果与我们的预期一致。我们对女性目标面孔这一矛盾的结果做了些讨论, 详细见“讨论”部分第四段。简单说来, 第一, 正如审稿专家提到的可能是因为对图像像素值向量化后求相关会丢失图像像素区块之间的空间联系, 导致它不能很好地表征图像之间的相似程度。第二, 人们会认为女性要比男性更可信(Buchan et al., 2008; Dong et al., 2018), 因此, 可能也会把更多的可信特征添加到不可信女性目标面孔上, 这也可能是导致女性结果并不稳定的原因。

意见 10: (Dotsch & Todorov, 2012) 的研究对分类图像的面孔区域像素与面孔特征进行了分析 (pixel cluster test, 参考文献“Diagnostic Facial Regions”部分)。为使研究更加完善, 作者为何不在目前研究内容的基础上, 进行同样的分析?

回应: 我们根据审稿专家的意见进行了该指标的计算与分析, 并做了些讨论。该分析方法见“第一阶段 获取分类图像”这一部分最后一段, 结果见“2.2.4 面孔识别诊断区域”这一部分, 有关讨论见“讨论”部分第三段。感谢专家的意见!

意见 11: 第 11 页“4 讨论”第 2 句: “实验 1 和实验 2 发现当同一个目标人物被描述成可信时(相比于不可信), 人们对其面孔的表征会更具吸引力。”在这两个实验中, 第一阶段通过描述将图像分类为可信/不可信; 第二阶段的被试再对这些图像进行特征评价。但是值得注意的是, 在第一阶段中, 被试被随机分入了可信组和不可信组, 但是在第二阶段的面孔评价, 一部分被试(任务 1)则需要对可信/不可信的面孔均进行吸引力评价, 或新的被试(任务 2)进行特征评价。也就是说, 第二阶段的大多数被试进行特征评价时, 不依赖第一阶段的“描述”过程。因此, 这里更符合逻辑的表述是否应当是“实验 1 和实验 2 发现同一张目标面孔被赋予不同的可信度描述时, 被认为可信(相对于不可信)的面孔表征, 也被认为更具吸引力。”? (实验方法上两个阶段是分离的, 而原文表述时, 给人被试是对描述内容进行吸引力评价的感觉)

回应: 我们非常同意审稿专家的意见, 在“4 讨论”第 2 句按照审稿专家的意见做了修改。

意见 12: 英文摘要与中文摘要内容不完全相符, 建议修改。

回应: 感谢审稿专家指出这一点, 我们修改了英文摘要使其与中文摘要相符合。

第二轮

审稿人 1 意见:

作者较好回应审稿人的问题，但我仍然有如下 3 个问题:

意见 1: 反相关图像是一种可视化的心理表征，但并不是心理表征本身。建议作者调整绝对化的表述;

回应: 我们非常赞同审稿专家的意见。我们在“1 引言”第四段强调了反相关图像分类技术能够将心理表征的内容可视化，并直接给出了本文中心理面孔表征的操作性定义为使用反相关图像分类技术获得的分类图像作为可视化的心理面孔表征。

此外，我们在正文中把“获得心理表征”改成了“对心理表征可视化”，以突出反相关图像是一种可视化的心理表征。

意见 2: 关于相同底片面孔带来的系统偏差，建议采取偏相关来排除系统偏差对相关的贡献;

回应: 非常感谢审稿专家提出这一方法。按照审稿专家的意见，我们使用偏向关的方法，把面孔底片的像素值向量化后纳入控制变量，再次求分类图像噪音模式之间的相关。由于偏相关系数的差异检验无法使用 Zou (2007)等人的方法，我们使用了置换检验的方法。偏相关的结果与使用相关的结果一致。为了保证统计检验方法的一致性，我们对相关的结果也补充了置换检验(见表 2, Zou (2007)和置换检验方法的结果是非常一致的)。偏相关的方法和结果见“3.1.4 数据处理”和“3.2 结果”部分。

意见 3: 关于+-噪音的选择结果有何差别的问题：原意是希望作者解释+噪音产生的表征图像与-噪音参数的表征图像，对于结果有什么影响。

回应: 感谢审稿专家再次提出这一问题。图 1 展示了叠加正负噪音后产生的刺激图。结果会发现如果一张刺激图片的某些像素点比较亮，另一张对应位置的像素点就会比较暗，反之亦然。因此，在一个试次中同时呈现正噪音以及其对应的负噪音能产生差别比较大的两张刺激图片。但是，如果在一个试次中呈现的两张刺激叠加的噪音是完全随机的，它们的差异可能就不会这么明显，这会导致需要更多的试次才能揭示不同的心理表征。因此，在一个试次中同时呈现叠加正负噪音的刺激是目前反相关图像分类技术的标准做法。

审稿人 2 意见: 修改稿中作者已经回复了之前提出的问题，建议接收。

回应: 再次感谢审稿专家建设性的意见!

第三轮

审稿人 1 意见:

作者已经回应了提出的问题，支持发表。

编委意见: 建议发表。

主编意见: 同意外审和编委意见，建议录用。