

# 《心理学报》审稿意见与作者回应

题目：不确定性与预期误差对情绪强度的作用：来自行为、眼动和 fMRI 的证据

作者：袁梦琪、邹迪、代乔、万蓓、王福顺、刘庆科

## 第一轮

审稿人 1 意见：

**意见 1：**请在正文中注明，所有描述性数据是以  $M \pm SD$  还是  $M \pm SE$  的呈现形式。例如，2.1.1 中被试“平均年龄为  $21.06 \pm 4.5$  岁”；2.2.1 中反应时结果，“高不确定性条件下 ( $1534 \pm 102\text{ms}$ )”，以及“被试作预期选择的反应时显著大于低不确定性条件 ( $1226 \pm 67\text{ms}$ )”等。

**回应：**非常感谢审稿专家的建议。已补充数据呈现规范的说明，统一全文使用  $M \pm SD$  格式。具体是在本文第一处出现  $M \pm SD$  格式的报告处做了标明，并统一说明数据呈现格式：“本文所有描述性数据均以均值  $\pm$  标准差 ( $M \pm SD$ ) 形式报告。”位置在“2.1.1 被试”段落中。

**意见 2：**2.1.3 实验流程中提到，在第二张卡片的数值揭晓后，先呈现情绪状态的评估界面，再呈现被试输赢判断界面。请作者给出这样设置实验顺序的理由。按照我的理解，被试在知道第二张卡片的数字后就已经知道了自己的输赢情况，理应先做出输赢的判断然后再对输赢结果进行评分。

**回应：**非常感谢审稿专家的建议。根据您的意见，我们在“2.1.3 实验程序”中补充了设计理由：“在第二张卡片揭晓后先进行情绪评分，旨在捕捉被试对输赢结果的直接情绪反应，避免后置的输赢结果呈现带来的认知加工干扰，从而获得更真实的主观情绪体验。”即我们在正文中补充了情绪评估顺序的设置原因，是为捕捉第一时间的情绪反应，避免后置的输赢结果呈现引入额外认知干扰与污染，提升主观报告的信度。因此优先记录情绪评分。

**意见 3：**为方便理解，建议在 2.1.4 统计分析中补充该研究分析了哪些因变量指标。

**回应：**非常感谢审稿专家的建议。已清晰列出所有分析的因变量指标，在原“2.1.4 统计分析”开头处补充了：“本实验分析的因变量指标包括行为结果（主观唤醒评分、效价评分）和瞳孔结果（瞳孔扩张变化值）。”以增强读者对分析重点的理解。

**意见 4：**请调整写作顺序。建议将 2.1.4 统计分析的第一段与 2.1.2 实验材料统一合并为“2.1.2 实验仪器”。同理，将 3.1.4 统计分析的第一段与 3.1.2 实验材料统一合并为“3.1.2 实验仪器”。

**回应：**非常感谢审稿专家的建议。已按建议将原“实验材料”与仪器设备及采集设置整合，统一为“实验仪器”。即将“2.1.4 统计分析”的第一段与“2.1.2 实验材料”统一合并为了“2.1.2 实验仪器”；将“3.1.4 统计分析”的第一段与“3.1.2 实验材料”统一合并为了“3.1.2 实验仪器”。

**意见 5：**研究结果中相对缺乏行为结果和眼动指标的关联分析；如有可能，希望可以分别提供主观效价和唤醒度报告水平与瞳孔结果之间关系的结果。

**回应：**非常感谢专家提出的宝贵建议。我们十分认同在多模态数据中探讨主观情绪体验与生

理指标之间的对应关系对于揭示情绪加工机制具有重要意义。因此，我们根据建议，通过皮尔逊相关分析，结果显示：瞳孔扩张变化值与主观唤醒评分呈边缘显著正相关（ $r = 0.262$ ,  $p = 0.081$ ），但与效价评分无显著关联（ $r = 0.056$ ,  $p = 0.711$ ）。这表明瞳孔扩张主要反映情绪唤醒度的生理变化。

**意见 6:** 总讨论中提到“在实验一中，我们还发现了高不确定性条件对唤醒度和瞳孔扩张的一致性降低效应，而在实验二中，我们并未发现存在参与该过程加工的大脑区域，这与我们的实验预期不一致”。作者认为其原因是“不确定性对情绪的影响受到了预期机制的调节”。出现不一致是否还有可能是因为眼动数据只给出了瞳孔扩张方面的指标，如果提供其他指标可能会发现相关脑区的活动，例对感兴趣区的注视时间。

**回应:** 非常感谢审稿专家的建议。在总讨论中补充：“值得注意的是，在 fMRI 中未观察到与高不确定性条件下瞳孔及唤醒一致性降低相关的特定脑区，可能是因为实验一眼动数据仅包含瞳孔扩张，若结合注视时间等多眼动指标验证，或可更敏感地揭示相应神经基础。”已在总讨论中补充说明，未发现预期区域可能因眼动数据单一，未来可纳入如注视时间等更多指标以提升神经行为对应关系的解析度。

**意见 7:** 请参照实验一，在实验二中补充使用 G\*Power 计算得出的计划样本量。

**回应:** 非常感谢审稿专家的建议。已在实验二方法部分补充计划样本量的计算依据与结果，即在“3.1.1 被试”中添加如下内容：“本实验使用 G\*Power 3.1 计算样本量。基于  $\alpha = 0.05$ ，效应量  $f = 0.25$ ，统计功效  $(1 - \beta) = 0.80$ ，重复测量设计，所需最小样本量为 16 人。”并核查了实验实际样本量满足样本量的规范性要求。

**意见 8:** 为了更直观的展示数据，请在图 3，图 4，图 9 中以散点形式呈现每个被试的数据。”

**回应:** 非常感谢审稿专家关于图示改进的建议。我们非常认同在群体统计分析中增加个体散点展示，有助于提升数据透明度与结果解释的直观性。在当前版本中，我们已在图 3、图 4 及图 9 中使用“均值  $\pm$  标准误 (SE)”误差条的方式，体现各条件下的组间差异，并保证图像的清晰性与专业性。考虑到本研究中部分图表涉及条件较多、对比维度复杂，若同时叠加每个被试的散点点位，可能造成图像信息过于密集，影响整体可读性与呈现效果，因此我们在当前排版中未采用该方式。再次感谢专家提出的改进方向，促使我们在图示表达方面更加严谨。

.....

审稿人 2 意见:

**意见 1.** 研究问题的推演过程还不够清楚，也并未如作者所称“支持并拓展了情绪二维理论。这一发现深化了对情绪结构的理解，再次验证了情绪强度衡量标准的独立性”的那样，导言和讨论都没有说明实验设计、实验结果如何能“支持并拓展了情绪二维理论”。也没有引入既有情绪理论进行验证。

**回应:** 非常感谢审稿专家指出我们在理论推演与理论拓展方面表述不够清晰的问题。我们已在导言与总讨论部分进行系统性补充与梳理，并明确本研究对情绪二维理论的支持与拓展路径如下：我们前期研究在国际上首次提出了情绪强度是唤醒度和效价的向量和，该理论支持并拓展了情绪二维理论。本研究首次系统性地验证了情绪强度和唤醒度和效价之间的关系，并且发现了预期误差与不确定性对情绪唤醒度和效价两个独立维度上的不同影响。该研究还进一步验证了情绪强度是依赖于情绪唤醒度和效价的交互作用。另外，本研究还通过结合行

为、眼动及 fMRI 技术，观察了不确定性与预期误差在情绪加工中涉及额回、扣带回、角回、缘上回、尾状核及脑干等多个关键脑区的不同及协同作用。这为先验信念为情绪强度、唤醒度和效价的神经基础提供了新的理论视角，丰富了情绪二维理论框架。最后，本研究发现在高不确定性条件下预期误差对情绪唤醒度和瞳孔扩张反应的影响精度。这一机制性发现揭示了不确定性在情绪调控中的对情绪强度和唤醒度和效价的不同影响，拓展了对情绪强度影响因素的理论解释。

**意见 2.** 本质上讲，本研究仍是个 2 因素设计，二因素的交互作用影响可以看着二因素设计的必然结果之一，为何这就能成为本文创新的焦点了呢？而且变量操控表述不清楚，比如概率确定性为实际为 7 个水平，认为划分为 2 个的理由何在？是否可以 7 个水平做回归分析？

**回应：**感谢审稿专家的建议。我们完全认同本研究在实验设计上为经典的  $2 \times 2$  设计，正如文中所述，本研究的理论创新性主要体现在交互效应的“心理机制解释”及其神经机制扩展，而非仅仅在于是否发现交互作用。传统研究往往将情绪强度等同于唤醒度或效价的单一维度，且极少系统性考察这两者在表述情绪强度时的贡献。有研究首次提出：情绪强度是唤醒度和效价的向量和（基于我们前期已发表研究），并进一步验证该理论中两个维度如何被上游认知变量以交互形式影响，在 fMRI 层面，我们首次发现这种交互机制涉及额中回—扣带回—脑干的功能环路整合。因此，我们的研究在设计层面属于双因素范式，在理论建构、神经机制的揭示，以及整合行为—眼动—脑成像三种层面上具有系统性创新。并且为了验证预期机制在唤醒和效价上的不同表征，我们将原有任务稍作了修改，从 2 个 block 内分别单独报告效价、唤醒改为每一个 block 内同时报告效价和唤醒；结果发现，预期误差与不确定性之间的交互作用在两个因变量效价、唤醒之间存在不同表现。

此外，感谢专家的严谨建议。我们在正式实验前对不确定性分布进行了预实验 ( $N=25$ )，发现在卡牌任务中，数值 2~3、7~8 的预期赢/输比例呈明显偏斜（如 25% vs 75%，或 88% vs 12%），而数值 4、5、6 则更加接近 50%/50%。因此我们依据概率分布的不对称性，将 2~3、7~8 归为“低不确定性”、4、5、6 归为“高不确定性”。这一划分符合《不确定性在情绪加工中影响机制》(Qiao et al., 2018) 中的方法逻辑。关于是否采用回归模型，我们在数据分析中进行了探索性回归，但受限于每个数值水平的试次数量极不均衡（低至每人每条件 3~4 个试次），导致单个水平的误差较大，模型显著性不稳定，故最终使用了“二分类”以增强统计效能和可解释性。

**意见 3.** “预期”的操控不够明确和纯净，跟确定性共变，比如低确定性时，预期赢的概率就低，那这样如何平衡各个条件的试次？在研究一中，研究者陈述“符合预期条件占比较多”符合概率效应，仅能说明在该实验设计中，被试更可能由不符合预期的刺激引起惊讶。但是无法解决在统计角度，试次数差异过大导致的潜在问题。“符合预期”的条件总共占比 69%，期望值 38.64 试次，“不符合预期”的条件总共占比 31%，期望值 17.36 试次。研究二同样存在类似问题，如此试次量的差异是否导致对随机误差的平衡存在较大差异？这可能严重影响实验结果。

**回应：**非常感谢专家指出“预期是否符合”与“结果不确定性”在本实验中确实存在统计上的共变关系，以及由此带来的试次数分布的不均衡。我们非常赞同审稿专家的建议，进行了该部分实验操作的实验改进与理论优势分析：1. 模拟真实情绪加工的“预期—验证”动态结构：在现实生活中，情绪产生往往并非由均匀随机的事件引发，而是取决于预期形成的置信度（确定性）和结果是否违背预期（PE）的联合模式。我们认为，“确定性”与“是否符合预期”之间的共变关系，正是人类情绪系统自然的认知结构。因此我们保留了这一结构性特征，以增强任务的心理生态效度。2. 构建“非对称刺激概率”是为了激发情绪差异的关键条

件：我们设计符合预期的条件出现概率更高（约 69%），不符合预期更低（约 31%），这一设置与“概率效应”理论高度一致（Kamp & Donchin, 2015），即：罕见事件更易引发显著的主观惊讶和唤醒反应。这一非对称概率结构是研究中“预期误差”能否引发情绪效应的前提条件，而非噪声。这种结构性分布正是用于探索“效价 vs 唤醒”的相对贡献量：如果采用正交分布（如平衡预期/不预期的试次数），则每种情绪状态被人为等量，可能掩盖自然条件下唤醒和效价的真实比重。我们正是希望通过这种概率引导下的自然分布，来揭示唤醒与效价在表征情绪强度中的非对称贡献关系，而不是将两者人为对等处理。3. 统计处理层面已控制不平衡影响：我们采用被试内设计 + 重复测量 ANOVA，确保各个条件对比基于同一被试内部的差异 4. 补充稳健性分析计划：为回应专家建议，我们已在过去的暑假期间重新采集了近 50 名被试的数据，以验证主要结论的稳健性，并进一步增加了试次数以排除干扰可能。再次感谢专家的敏锐关注，让我们进一步明确并强化了实验设计中的理论逻辑，也有助于提升稿件的严谨性。

**意见 4.** 在研究一中，每个阶段的有效试次数为 56，因此每个被试的“低不确定性不符合预期”条件数期望值为 6.72，部分被试该条件下的试次数低于该数值。这样的试次数是否足够平衡随机误差？进一步地，研究者还在结果分析中将“奖赏”作为第三个自变量纳入分析，这将导致条件下的试次数进一步减少，研究者并未报告该情况下的最少试次数是多少，该情况下的试次数是否足够平衡随机误差？

**回应：**感谢专家提出关于“奖赏”作为第三因素引入后试次数过少的问题，我们认真反思了这一变量在本研究中的定位与作用并进行了改进。我们在原设计中确实记录了“输赢”结果，但并未明确将其作为独立变量进行实验操控或实验条件设置。因此，研究中我们始终是引导被试在看到结果之后，主观感受该结果是否符合预期，并基于这种认知偏差（PE）来报告自身的唤醒与效价体验。从而保证“奖赏”并非一个理论驱动的自变量，而是结果刺激本身在自然语义上的属性，并未经过系统操控。而且，我们在修订稿中去除将“奖赏”作为正式因素的统计分析和相关解释，聚焦于本研究的理论核心——不确定性与预期误差如何共同作用于情绪强度的形成。删除该因素后，我们的统计模型结构更为简洁明晰，同时也能减少分析维度和试次数被分割过细所带来的误差问题，有利于增强结论的稳定性和清晰度。衷心感谢专家指出该问题，促使我们进一步聚焦于研究问题本身。

**意见 5.** 研究一未将眼动数据与情绪效价与唤醒度的数据进行联合分析，仅依据方差分析差异模式的相似性便判断瞳孔扩张与情绪数据的关联，证据不足。建议补充与此相关的分析结果。

**回应：**非常感谢审稿专家的建议。已补充相关性分析：“2.2.3 行为与眼动结果相关分析，结果显示：瞳孔扩张变化值与主观唤醒评分呈边缘显著正相关（ $r = 0.262$ ,  $p = 0.081$ ），但与效价评分无显著关联（ $r = 0.056$ ,  $p = 0.711$ ）。这表明瞳孔扩张主要反映情绪唤醒度的生理变化。见图 5：瞳孔扩张与唤醒评分散点图（附回归线）”

**意见 6.** 作者将奖赏作为自变量之一纳入结果分析，并在摘要、研究一讨论与结论中提及，但是并未在引言中综述相关研究，也未在总讨论中提及，建议补充。

**回应：**非常感谢专家的细致建议。经过进一步梳理，我们意识到“奖赏”并非本研究理论假设所涉及的核心变量，其本质上是结果属性的一部分，而非独立被操控的实验因素。在本修订稿中，我们已删除了将奖赏作为自变量的分析，并相应删减了摘要、讨论等部分中对“奖赏效应”的引用与解释，以保证研究聚焦于“预期误差  $\times$  不确定性”的主效应与交互机制。

**意见 7.** 研究二仅有神经数据的分析结果,研究者是否收集了情绪效价与唤醒度的数据?在没有将情绪数据与神经数据进行联合分析的情况下,将神经层面发现的结果解释为情绪加工的认知神经机制,证据不足。并且研究者在总讨论中提及“在实验一中,我们还发现了高不确定性条件对唤醒度和瞳孔扩张的一致性降低效应,而在实验二中,我们并未发现存在参与该过程加工的大脑区域,这与我们的实验预期不一致”,进一步说明了以研究一结果直接为研究二神经数据结果做解释的不恰当性,建议补充与此相关的分析结果。

**回应:** 非常感谢审稿专家的建议。在实验二中,我们确实收集了被试对每个试次的唤醒度与效价评分,但由于篇幅所限未在初稿中报告。在修订稿中,我们已补充该部分分析内容(见第 3.2 节),并将神经激活结果与主观情绪评分进行了相关分析,结果表明:在多个关键脑区(如前扣带回、额中回、脑岛等)激活水平与主观唤醒评分间存在正相关,进一步支持这些脑区在情绪加工中扮演的角色。该补充增强了我们对“认知-情绪加工整合机制”的神经解释基础。

**意见 8.** 摘要自明性不高,比如文章称“不确定性与预期误差交互式地影响着情绪的唤醒度、效价和瞳孔扩张反应”,但这并未说明具体是如何影响的。

**回应:** 非常感谢审稿专家的建议。我们已对摘要进行修改,使其更清晰地说明“不确定性 × 预期误差”对情绪唤醒、效价及瞳孔反应的具体影响模式,包括方向性与交互关系。

**意见 9.** 本研究是实验室实验,但是却在自检报告中回回应了否,与事实不符,建议修改。

**回应:** 非常感谢审稿专家的建议。已在自检报告中将实验类型修改为“实验室实验”,与实际研究一致。

**意见 10.** 实验二没有写明实验设计,也没有报告预估被试量计算方法,如果实验设计与实验一相同,那被试总样本量不少于 24 人,16 人就不足了。

**回应:** 非常感谢审稿专家的建议。实验二为 fMRI 任务态设计,考虑到核磁数据质量与资源限制,我们进行了功效分析,基于中等效应量( $f = 0.3$ )与  $\alpha = 0.05$  条件下,预估最低样本量为 16 人。

**意见 11.** “认为潜在的威胁进行超过其可能严重程度的预测导致过度负性预期”这句话很不通顺,建议修改。

**回应:** 非常感谢审稿专家的建议。已在正文中进行修改。

**意见 12.** “也有研究发现不确定性可以影响正性情绪,结果发现不确定性加剧了参与者的兴奋感,说明了不确定性可以放大正性情绪”这句话很不通顺,建议修改。

**回应:** 非常感谢审稿专家的建议。已在正文中进行修改。

**意见 13.** 在研究一中,研究者使用反应时差异论证对高、低不确定性的操纵,证据不够直接。此外,被试引用参考文献说明“做决策选择时更快的反应时是个体自信状态的指标”,来论证反应时与不确定性关联并不合适,自信状态与不确定性并不完全等价。是否有其他结果能够说明对不确定性的操纵?

**回应:** 非常感谢审稿专家严谨的审阅意见。我们诚恳地承认,当前使用反应时差异作为“高/低不确定性”操控检验的指标,确实存在一定的间接性和理论上的不充分之处。在本研究中,“不确定性”的操作方式是通过卡牌任务中数值所对应的赢/输概率分布实现的(如:2

号卡牌为 88%赢/12%输, 6 号卡牌为 50%赢/50%输)。这种以客观概率差异引导不确定性体验的方式, 在已有研究中被广泛使用 (如: Hsu et al., 2005; Grinband et al., 2006), 虽然其对主观感知的映射并非完全一一对应, 但在实验心理学中仍属主流范式。我们使用反应时的差异主要作为一种任务感知复杂性与决策迟疑程度的行为指标, 而非直接测量主观不确定性感知。我们完全同意专家所指出的“自信”与“不确定性”之间并不等价的逻辑问题, 因此我们在修订稿中已删除原引用, 并对该解释进行了相应修改。鉴于本研究的设计重心在于构建认知预测偏差所引发的情绪机制, 我们今后研究将进一步考虑引入被试主观不确定性评分或试次后信心打分等直接测量方式, 以增强操控验证的理论严谨性。再次感谢专家的专业建议, 帮助我们更清晰地界定操作变量与其验证方法之间的边界与逻辑关系。

**意见 14.** 统计量应统一使用斜体, 作者文中有的统计量使用斜体, 有的没有。

**回应:** 非常感谢审稿专家的建议。已在正文中进行修改。

---

## 第二轮

### 审稿人 1 意见

作者对审稿人的意见进行了较好回复, 但是就文章的问题论证和贡献方面的回复还比较模糊, 没有直面问题, 比如何以证明情绪强度衡量标准的独立性? 其充要条件是什么? 文章真提供了足够证据了么? 等问题还缺乏系统论证。具体来看下述问题。

**回应:** 非常感谢审稿专家对我们上次修改的肯定, 并耐心地提出了文章的进一步修改的建议。我们完全同意您指出的“情绪强度衡量标准的独立性及其充要条件”问题, 确实触及了本研究理论基础的深层环节。在这次的修改中, 我们梳理了理论框架, 并在修改稿中补充一项新的验证性实验以更好地支持当前研究。

**意见 1、**“情绪强度独立于唤醒/效价”与作者“向量和”定义矛盾, 需澄清概念。

**回应:** 感谢审稿专家指出的表述问题, 这一点非常关键, 我们完全同意您的看法。我们在原文中“情绪强度独立于唤醒/效价”这一表述确实存在歧义, 在修改版本中, 我们进行了修改, 澄清了概念。

**意见 2、**核心变量“情绪强度”无直接测量, 用“向量和”直接替代合理性不足。

**回应:** 感谢审稿专家的深刻意见。针对您指出的核心问题, 我们在此次修改中已通过补充实验进行直接测量。我们新增了一项行为学实验, 要求被试在体验情绪刺激后, 使用直观的自评量表 (如“感受到的情绪有多强烈?”) 对“主观情绪强度”进行直接评分。相关数据、分析流程及与原有“向量和”计算结果的比较分析, 已作为“补充实验与数据验证”部分完整呈现于论文附录中。

**意见 3、**瞳孔-唤醒相关受单点离群拖累且仅边缘显著, 应报告剔除标准与敏感性。

**回应:** 感谢审稿专家的细致审阅。您指出的问题非常重要, 确保统计分析的严谨性是得出可靠结论的基础。我们已在修改稿中针对此点进行了补充和修正。详情请见图 5。

**意见 4、**未说明低不确定条件期望 6.72 能否抵消随机误差。

**回应:** 感谢审稿专家提出的深刻问题。我们已通过补充统计分析, 直接证明了“6.72”这一值能够抵消随机误差。我们将 (6.72 次) 与上述随机基线 (14.22 次) 进行了比较。单样本  $t$  检验表明, 6.72 显著低于 14.22 ( $t(35) = -27.606, p < 0.001$ )。这证明被试的预测行

为有效地利用了任务中的概率信息,从而将“不符合预期”的结果系统性地减少到远低于随机误差的水平。

**意见 5、**声称已补做稳健性分析,却无任何结果,请补数据或删除表述。

**回应:**感谢审稿专家指出此问题。我们在此前的修改中,在扩大样本的基础上尝试了相关的稳健性分析,但该分析的结果未能充分、严谨地支持“稳健性”这一结论。我们完全同意您的意见,这些初步结果不宜作为正式论据呈现。为此,我们已在文中删除了关于已补做稳健性分析的相关表述,并为之前表述的不严谨深表歉意。感谢您的严格审阅,这使我们的论文表述更加准确。

**意见 6、**研究一 vs 二“不确定→唤醒”显著性不一,结果不稳健,讨论应谨慎。

**回应:**感谢审稿专家的意见,针对研究一、二的结果不完全一致,我们已慎重修改讨论及对应部分。见讨论 4。

**意见 7、**研究二表格/图示碎片化、cluster 数对不上,请统一格式并标注阈值。

**回应:**感谢审稿专家对结果呈现规范性的细致审阅。根据您的建议,我们对研究二结果部分进行了系统整理与核对,具体修改如下:图示合并:为减少图示分散带来的阅读困难,我们将原图 6、7、8 合并为组图图 6 (abc),以更清晰地呈现整体激活模式。统一对应关系:我们逐一核查了表 4、正文描述与图 6c 柱状图的脑区对应。通过合并属于同一 Cluster 的子脑区(由 Voxel size 列标识),使表格展示的 10 个统计簇与文中描述的激活模式完全一致,解决了 Cluster 数对不上的问题。规范化术语与标注:统一校对了脑区解剖学译名(如将“Middle Temporal Gyrus”统一更正为“颞中回”),并在组图图 6 (abc)、表 3、表 4 的图注、表注中明确标注了统计阈值。相关修改已在修订稿 3.2 结果部分组图图 6(abc)、表 3、表 4 完成。

**意见 8、**神经-主观相关仅报唤醒、缺效价,且未给全矩阵,需补全并检验处理效应方向一致性。

**回应:**感谢审稿专家的重要指正,我们理解该建议旨在进一步检验神经活动与主观体验之间的关联性。然而,该评分任务并非实验二的核心操控变量,且在对实验二数据完整性与分析稳定性进行进一步复核时,我们发现该行为数据在试次数覆盖率与信度方面存在一定局限,难以支持系统的神经—主观联合分析。因此,在本次修订中,我们未将该部分结果作为正式分析内容纳入正文报告。实验二中所指“唤醒”与“效价”是基于文章对相关脑区(如显著性网络与价值相关区域)的功能界定,并非源自被试的主观报告。我们理解审稿人希望探究神经与主观体验关联的意图。为避免误解,修订稿中已对表述作出系统性调整,删除可能被解读为神经—主观直接关联的措辞。

**意见 9、“条件内相关”理由不足,需说明是否满足测量等值。**

**回应:**感谢专家的提醒。我们认同,在跨实验条件比较心理评分时,测量等值是一个需要明确说明的重要方法学前提。我们对此进行了深入论证并在正文中作了明确说明:操作等值:实验一中所有条件均采用完全一致的 9 点量表、指导语和任务流程,从操作层面确保了测量框架的跨条件一致性。逻辑适用性:本研究采用被试内相关(within-subject correlation),关注的是个体内部心理评分与生理指标波动的耦合关系,而非跨条件间的均值绝对比较,因此本分析对“跨条件测量等值”的依赖较低。局限性补充:已在修订稿 2.1.3 补充了操作一致性说明,并在 4.1 讨论了该方法学假设的限制。

**意见 10、功效分析工具不明；G\*Power 不适用于双因素被试内，请换 Monte-Carlo 并公开参数。**

**回应：**感谢审稿专家在统计方法学层面的专业指正。我们完全认同，对于双因素被试内设计，尤其是在关注交互效应的情况下，G\*Power 在功效估计上的确存在方法学局限，其结果高度依赖于对效应量和被试内相关结构的先验假设。您提出的 Monte-Carlo 模拟方法，无疑是处理此类复杂设计更为精确和稳健的策略。鉴于本研究数据已完成采集，且主要结论基于实际观测效应而非事前样本量规划，我们认为在修回阶段基于观测数据进一步开展 Monte-Carlo 模拟进行事后功效估计，可能会引入额外假设并造成对结果稳健性的误导性解读。权衡再三，我们采取了更稳健的处理方式：在修订稿中淡化了基于 G\*Power 的功效表述，转而通过在“总讨论”中详细阐述样本量局限、交互效应的观测效应量及其统计敏感性限制，来客观呈现结果的边界。我们非常珍视审稿专家的专业建议，并将在未来研究中采用基于 Monte-Carlo 模拟的事前功效规划，以更恰当地应对复杂被试内设计下的样本量估计问题。

审稿人 2 意见

作者对提出的疑问都做出了较好的回应和相应的修改。

---

### 第三轮

审稿人 1 意见：作者已经较好地回应了我的疑问，没有进一步的意见。

编委意见：同意审稿人的意见，请作者对照学报对图表的要求修改完善。

**回应：**感谢责任编辑的意见。我们已根据学报对图表的相关规范要求，对全文中的图表进行了系统性检查与修改完善，具体修改包括：

- （1）统一图表编号、标题及字体格式；
- （2）补充并规范图注及数据来源说明；
- （3）优化图表排版与布局，使其符合版面要求；
- （4）对个别表达不清或信息不完整的图表内容进行了完善。

相关修改已体现在正文中对应位置，主要涉及正文部分：第 6 页图 1、第 7 页图 2 及表 1、第 8 页图 3 及表 2、第 9 页图 4、第 10 页图 5、第 14 页表 3、第 15 页图 6、第 16、17 页表 4；附录部分：第 29 页图 1、第 31 页图 2 及图 3。

再次感谢您的细致指导。

主编意见：可以发表