

《心理学报》审稿意见与作者回应

题目：前额 P3 幅度调控对情绪调节的促进作用

作者：李宜伟，汤煜尧，王庭栋，张丹丹

第一轮

审稿人 1 意见：

该研究采用神经反馈技术系统考察了前额 P3 幅度对情绪调节的潜在因果作用，前者长期以来被视为情绪调节等执行控制任务早期阶段前额资源调用程度的神经相关物。结果发现，与训练前相比，实验组通过 NF 训练可成功上调前额 P3 幅度，NF 训练后的情绪调节过程能更有效调用前额 P3 指征的执行控制功能，从而达到更佳的情绪调节效果。研究具有较好的创新性和科学价值，实验设计严谨、数据分析正确，结论可靠，讨论深入，有助于深化对情绪调节脑机制的理解。研究将认知神经科学的研究与神经调控的应用需求紧密结合，应用价值凸显性强。以下问题建议作者思考和修改：

意见 1：摘要中，“对照组则呈现相反方向的前额 P3 及情绪调节效果改变。”需要直接说发生了什么改变，以便于阅读。对“同时结果表明，P3-NF 训练主要通过增加情绪调节过程的前额 P3 幅度以提高情绪调节效果。”是否使用如下表达：P3-NF 训练主要通过增加认知重评过程前额 P3 幅度的改变以提高认知重评调节效果。

回复：我们已将该表述修改为更具体的描述：“对照组则相反：NF 训练后情绪调节期间前额 P3 幅度下降，负性情绪自评和 LPP 均显著提高。”（见第 1 页 10-11 行）。

我们已经修改此处表述，见第 1 页 11-12 行。

意见 2：引言部分提到“根据过程模型理论，情绪调节是包含多个认知阶段的过程：情境选择、情境修改、注意分配、认知改变和反应调节(Gross, 2015)”，但后文并未进一步展开或引用该理论框架。建议删除此句或在后续内容中与该理论建立更明确的联系。

回复：我们已对引言这一部分进行了修改，删除了该理论框架，直接引出认知重评策略一本文核心关注的情绪调节策略。

意见 3：引言中，“相比于 EEG 的自发或引发节律”建议修改为“相比于 EEG 的自发或诱发节律”，这样表述更为严谨规范。

回复：我们认为 EEG “引发节律”的表述更准确。

意见 4：“第一，考虑到现有研究尚未得到一致性的 ERP-NF 训练效果，我们认为有必要同时考察相反调控方向的 NF 训练过程和训练效果，从而更全面地评估 ERP-NF 的科研和应用价值。因此，本研究让两组被试分别上调(实验组)和下调(对照组)P3 幅度。”需要更清楚说明相反调控为何能够解决以往的不一致。

回复：我们对该部分进行了扩充说明，见第 3 页 14-15 行。

意见 5：引言末尾提到“探索性纳入多个潜在的 NF 训练效率影响因素”，但未具体说明考察了哪些因素，也未提供选择这些因素的理论依据。建议在引言中至少列举拟考察的主要个体差异变量（如基线认知控制能力、情绪调节习惯性策略使用频率等），并简要说明其理论相关性。

回复：感谢您的建议，据此我们已增补了论述内容，见第 3 页 24-27 行。

意见 6：方法部分，请明确说明情绪调节任务中的负性情绪评分、认知重评难度、前额 P3 幅度、顶区 LPP 幅度这四个指标是否均采用基于 single trial 的线性混合效应模型(LMER)进行统计检验。此外，“被试随机效应结构包含随机截距和前后测主效应的随机斜率(见公式 2)”这一表述建议修改为“被试的随机效应结构包含随机截距和前后测主效应的随机斜率(见公式 2)”，以保持语法的一致性。

回复：本研究中所有因变量，包括 NF 训练阶段的前额 P3 幅度、情绪调节任务中的负性情绪评分、认知重评难度、前额 P3 幅度和顶区 LPP 幅度，均采用基于 single trial 的线性混合效应模型进行统计检验。已修改相应表述，见第 8 页 7-8 行。

我们已将表述修改为“被试的随机效应结构包含随机截距和前后测主效应的随机斜率(见公式 2)。”（见第 8 页 9-10 行）。

意见 7：结果部分，在中介模型分析中，当以顶区 LPP 幅度作为因变量时，其结果是否与以负性情绪评分为因变量时保持一致？请补充说明。

回复：我们已在结果部分补充了“以顶区 LPP 幅度作为因变量的中介分析”的结果，见第 12 页 19-25 行。

意见 8：讨论部分描述“两组均表现出了典型的非线性学习轨迹：前三个训练 session 中 P3 幅度变化显著(上调组快速增加，下调组快速减小)，随后逐渐趋于稳定”，但这一表述似乎与结果不完全吻合。结果显示两组在最后一个 session 存在差异，且实验组（上调组）和对照组（下调组）在后期均表现出前额 P3 的上升趋势。建议作者进行解释。

回复：图 2A 的对照组曲线，在最后一个 session 确实存在一个 P3 幅度上升的趋势，但是并无统计显著性，目前很难推测这一结果的原因。不过大致上看，确实两组都比较符合经典的学习轨迹。为表述准确，我们修订了讨论的相应内容，见第 13 页 11-12 行。

意见 9：讨论部分对 NF 训练效率个体差异的分析很有价值，这也是 TMS 等其他脑调控技术面临的共性问题。建议补充报告有多少比例的被试能够成功完成 NF 训练（即达到预设的训练标准），这对评估该技术的临床应用潜力至关重要。

回复：已补充，见第 14 页 24-26 行。

意见 10：讨论部分关于临床应用前景的论述具有较强的说服力和现实意义。建议作者进一步阐述以下问题：其他 NF 研究（特别是情绪调节相关研究）的训练持续时间和追踪周期大

致是多少？P3-NF 训练效果能否迁移到其他情绪调节策略（如注意分散、表达抑制等）？这些信息将有助于更全面地评估 P3-NF 的临床应用价值。

回复：感谢您的建议。我们在讨论部分补充讨论了其他情绪调节 NF 研究的训练持续时间和追踪周期，见第 15 页 20-25 行。

由于前额 P3 指征的是执行控制功能，其增强可能对多种涉及认知控制的情绪调节策略都有或多或少的促进作用。我们在讨论增加了论述，见第 16 页 25-27 行。

.....

审稿人 2 意见：

研究采用靶向前额 P3 的 ERP 神经反馈训练，并在情绪调节任务中检验其训练过程和效果，选题具有潜在的理论意义和应用价值。但论文在核心概念界定、统计分析参数设定及因果解释逻辑等方面存在如下问题，建议修改：

意见 1.问题提出的逻辑性问题：摘要中作者提到“实验组通过 NF 训练可成功上调前额 P3 幅度，NF 训练后的情绪调节过程能更有效调用前额 P3 指征的执行控制功能，从而达到更佳的情绪调节效果”，存在循环论证嫌疑。作者用神经反馈训练改变 P3 幅度，但同时又把“P3 上调/下调”当作执行控制能力增强或减弱的证据，并用它来解释情绪调节变好或变差的原因。易让读者困惑：P3 是训练带来的变化（结果），还是情绪改善的原因（机制）？

回复：我们在此澄清本研究的逻辑：前额 P3 是执行控制功能的神经标志物（Bernat et al., 2011; Moser et al., 2014; Yuan et al., 2015, 2023），通过 NF 训练被试学会了上调 P3 幅度（这是训练的直接结果），而 P3 幅度的增加反映了其所指征的执行控制功能的增强（这是认知层面的变化）。因此，在后续情绪调节任务中，由于执行控制功能得到增强，被试获得了更好的情绪调节效果。换言之，P3 幅度变化既是 NF 训练的直接结果，也是情绪调节效果变好的原因。

为了更清楚的表达上述观点，我们修改了摘要的相应表述，见第 1 页 8-9 行。

意见 2.引言第二段提到“前额 P3/LPP 是情绪调节等执行控制功能早期阶段认知控制资源调用程度的公认神经标志物。”作者为什么选择上调 P3 幅度，而不选择 LPP 作为靶点呢？

回复：原文中同时提及前额 P3 与 LPP，可能引起了误解。其实学界更认可前额 P3 是执行控制功能的电生理指标。只是由于个别研究会将 P3 的时间窗延迟到比较晚的时刻，例如超过经典 P3 的 500 ms，我们在介绍时才加入了 LPP。为避免歧义，我们已在修改稿中将表述改成“前额 P3”，见第 2 页 11 行。

意见 3.核心概念界定不准确：引言第二段作者提到“ERP 的前额 P3 成分也因其时间跨度超过经典的 300~500 ms (通常持续 2~3 s)被称为前额晚正成分 (late positive potential, LPP”。前额 P3 也被称为前额 LPP 不合适，两者在功能和头皮分布并不相同，LPP 不是因为“持续时间长”才被称为 LPP，建议作者在文中明确 P3 与 LPP 之间的关系。

回复：非常感谢您的指正。我们已在修改稿中删除了所有将 P3 与 LPP 混淆的表述，现在明确区分了这两个成分的认知功能意义和头皮分布，见第 2 页 11 行、第 7 页 19-22 行。

意见 4.因果推断提前: 引言第二段的最后,作者提到“这可以因果性地证明前额 P3 指征的认知控制系统是情绪调节的关键性启动和实施准备脑网络”,研究还没有提出解决问题的方法、排除替代解释的策略,便在引言的问题提出中说“可以因果性地证明”“关键脑网络”。容易让读者产生因果推断提前,逻辑证据链尚未建立就进行判断的疑问。

回复: 感谢您指出的问题。已修改此处,见第 2 页 21-22 行。

意见 5.EEG 与 ERP 之间关系需明确: 引言第三段作者提到“相比于 EEG 的自发或引发节律,事件相关电位(event-related potential, ERP)被认为具有更高的时间分辨率,基于 ERP 的 NF 可以为我们了解认知过程提供更精确的时间信息”。表达不准确。ERP 的时间分辨率并不优于 EEG,因为 ERP 本质上是基于 EEG 信号的特定事件相关分析结果。

回复: 感谢您的指正。我们修改了此处表述,见第 3 页 3-5 行。

意见 6.假设不合适: 引言最后一段作者提出假设“ERP-NF 的训练效率或可基于某些因素进行一定程度的预测”,但当前表述并未明确具体预测变量,也缺少对应理论依据,导致该假设难以检验。

回复: 感谢您的宝贵意见。NF 训练效率的个体差异一直是 NF 应用领域最为关注的问题(Haugg et al., 2020; Hellrung et al., 2022; Sitaram et al., 2024)。根据近期关于 NF 训练个体差异的元分析研究(Alkoby et al., 2018; Haugg et al., 2020)和考察了个体差异影响因素的相关研究文献(Li et al., 2025a),我们提出了本研究探索的两个预测变量:基线时期的前额 P3 幅度(Alkoby et al., 2018; Haugg et al., 2020)、日常使用认知重评策略的频率(Li et al., 2025a)。此处表述已进行了修改,见第 3 页 24-27 行。

意见 7.统计结果有疑问: 表 1 显示,在特质焦虑中,实验组平均得分 40.618,对照组平均得分 45.121,统计结果为 $t=0.206$ 。根据其他变量计算模式推测,均值差为实验组-对照组, t 值应为负数。但在特质焦虑这一变量中,均值差与 t 值方向不匹配,是否存在计算或抄写错误,请全文仔细检查。

回复: 感谢您指出此处错误。特质焦虑(STAI-T)实验组平均得分为 45.618,我们之前确实抄错了,已修改(第 4 页)。

意见 8.统计结果有疑问: 2.3 部分第一段,“控制 2 个子集图片的效价和唤醒度无差异(效价: $t(58) = 0.000, p = 1.000$)”,一般在真实数据中, t 值几乎不可能等于 0.000(除非所有被试两条件下每个数据点完全相同),这在随机抽样中出现的概率极低,这个材料评价结果有点近乎“完美”,请介绍材料评价的具体结果。

回复: 这个结果确实看起来过于“完美”,但图片分组是人为匹配进行的,因此统计结果确实可以如此完美。为了更透明地呈现材料信息,我们在下方附表中提供了所有使用图片的 IAPS 编号及其对应的效价和唤醒度评分。

附表 前后测刺激材料

IASP 图片编号	效价	唤醒度	子集	IASP 图片编号	效价	唤醒度	子集
6836	3	5.44	A	9181	2.98	5.02	B
6840	2.98	4.4	A	9594	2.95	5.14	B
2120	2.95	5.23	A	3181	2.92	4.75	B
2700	2.95	4.28	A	1310	2.91	6.5	B
6230	2.91	5.82	A	1945	2.91	5.71	B
6260	2.9	6.33	A	2722	2.89	4.48	B
2691	2.87	4.9	A	2205	2.86	4.76	B
9182	2.86	5.52	A	2141	2.79	4.28	B
9250	2.86	5.5	A	9102	2.77	4.74	B
2053	2.77	5.5	A	9600	2.73	5.58	B
8231	2.77	4.84	A	9500	2.7	5.83	B
6510	2.75	5.85	A	9490	2.63	5.46	B
1201	2.67	5.43	A	3180	2.63	4.73	B
9330	2.61	4.36	A	9180	2.61	5.45	B
9253	2.57	5.61	A	3230	2.57	4.63	B
9415	2.57	4.85	A	9584	2.54	5.02	B
6313	2.56	5.87	A	9480	2.5	5.16	B
9373	2.37	4.1	A	6300	2.46	6.48	B
3530	2.36	6.19	A	9430	2.43	5.22	B
3301	2.28	4.88	A	9280	2.42	3.69	B
6570	2.24	6.2	A	6370	2.36	5.17	B
6010	2.2	5.17	A	9571	2.32	5.15	B
9300	2.12	4.71	A	6550	2.26	6.6	B
9400	2.12	5.26	A	8230	2.14	5.65	B
1280	2.11	5.52	A	3550	2.02	5.86	B
9140	2.11	5.28	A	3062	2	5.61	B
9410	2.02	4.83	A	6831	2	6.26	B
3030	1.86	5.83	A	9320	1.8	5.38	B
9042	1.69	6.04	A	3350	1.77	5.83	B
2800	1.45	5.54	A	9570	1.61	5.98	B

意见 9.NF 训练中策略的自由使用是否会影响结果？2.4 部分第一段，作者提到“参与者需在观看情绪图片时探索并使用适用于自己的策略(包括但不限于：调节图片带来的负性情绪，回忆，转移注意，心算，抑制等)，以提高温度计中红色液体的高度。”NF 训练目标是让被试学习调控 P3 幅度，但被试可以自由使用各种策略（如分心、心算、回忆、情绪抑制等）。这种“自由使用各种策略”设定是否会影响对 P3 作用结果的解释？

回复：“自由策略”是目前 NF 训练的常用训练设置（Li et al., 2025a, 2025b; Lubianiker et al., 2026; Peng et al., 2020）。例如，Lubianiker 等人（2026）的研究中，被试可以自由选择 NF 训练策略（包括想开心的事情、算术、关注身体某个部位等），只要该策略可以提高中脑边缘

系统的 BOLD 信号; Li 等人 (2025a, 2025b) 的研究也采用了类似的设计。

NF 训练的目的是按要求改变靶点神经活动, 被试需要通过探索找到对自己而言的优化策略 (每个被试所用的策略存在个体差异)。我们认为“自由策略”不会影响对 P3 作用的解释, 因为通过事后采访以及我们团队的实验研究经验, 被试采用的策略确实都与认知控制有关。

参考文献:

Lubianiker, N., Koren, T., Djerasi, M., Sirotkin, M., Singer, N., Jalon, I., Lerner, A., Sar-El, R., Sharon, H., Shahar, M., Azulay-Debby, H., Rolls, A., & Hendler, T. (2026). Upregulation of reward mesolimbic activity and immune response to vaccination: a randomized controlled trial. *Nature Medicine*, 10.1038/s41591-025-04140-5. Advance online publication.

Peng, W., Zhan, Y., Jiang, Y., Nan, W., Kadosh, R. C., & Wan, F. (2020). Individual variation in alpha neurofeedback training efficacy predicts pain modulation. *Neuroimage. Clinical*, 28, 102454.

意见 10.数据分析中的更新规则缺少依据: 2.6 部分作者提到“实验组的更新规则”, 文中温度计刻度的换算方式是本研究自定义的规则吗? 是否有研究用过类似换算规则? 或者该规则是否相关理论或实证依据?

回复:文中温度计刻度的换算方式并非完全自定义, 而是基于已有 ERP-NF 研究的方法并结合本研究的预实验结果确定的。NF 温度计存在一个基线值和目标值, 分别对应温度计的 0 刻度和最高刻度。1) 基线值的设定: 我们参考了之前两项 ERP-NF 研究 (Chang et al., 2014; Pei et al., 2020)。2) 目标值的设定: 我们根据预实验结果确定目标值为基线值+1 μV , 因为预实验结果显示, 大部分被试在单轮次的前额 P3 变化幅度相较于基线值在 $\pm 1 \mu\text{V}$ 范围内, 鲜有被试能超过+1 μV 。

我们在修改稿中添加了更新规则的依据, 见第 7 页 3-4 行。

意见 11.训练阶段编码中参数设定缺少依据: 2.8 部分作者提到“训练阶段采用二项式编码: 将 session 4 (+1/4)、5 (+1/4) 与 session 1 (-1/6)、session 2 (-1/6)、session 3 (-1/6) 进行比较”, 为什么 session 4/5 赋值 +1/4, 而 session 1-3 赋值 -1/6? 以及“前后测采用二项式编码, 前测任务赋值为-1/2, 后测任务赋值为+1/2。”这些赋值方式是否合理? 赋值方式是否有相关理论和实证依据? 这种编码方式是否会影响模型解释?

回复:我们使用的编码方式是正交对比编码, 这是混合线性模型因素编码的常用方法 (Wu et al., 2022)。具体而言, 将 session 4-5 赋值为+1/4、session 1-3 赋值为-1/6, 是为了比较训练后期 (session 4-5) 与训练早期 (session 1-3) 的差异, 同时保持编码的正交性, 即权重之和为 0: $2 \times (+1/4) + 3 \times (-1/6) = 0$ 。这种编码方式可以使用编码权重反映各水平的样本量比例, 并且保证了对比的统计独立性。类似地, 前后测的-1/2 和+1/2 编码也是标准的二分类正交编码。我们已在本轮修改中补充了编码方式的依据, 见第 7 页 29 行。

意见 12.语言表达前后不一致: 2.8 部分最后一段作者提到“为确定 NF 训练效率是否中介 NF 训练导致的后测前额 P3 变化与情绪调节行为改变之间的关系, 采用 bruceR 工具包(版本: 0.8.9)中的 PROCESS 函数进行中介分析(Bao, 2022)。在该分析中, NF 训练效率作为自变量, 情绪调节任务中前额 P3 变化(后测减前测)作为中介变量, 负性情绪评分或 LPP 幅度变化(后测减前测)作为因变量”。前面一句说“为确定 NF 训练效率是否中介 NF 训练导致的后测前额

P3 变化与情绪调节行为改变之间的关系”，即 NF 训练效率是中介变量。但后面又说“NF 训练效率作为自变量”，前后矛盾，请重新梳理逻辑。

回复：非常感谢您指出的表述错误。我们已更正，正确的表述为：“为确定 NF 训练效率是否通过调节情绪调节期间前额 P3 幅度变化来改变情绪调节效果。”（见第 8 页 15-16 行）。

意见 13.表达逻辑问题：在讨论部分第三段，作者引用了一项 TMS 研究(Cheng et al., 2024)是通过干扰前额叶功能，发现一旦前额叶在情绪加工早期受到干扰，负性情绪体验就会增强，这说明前额叶在情绪调节早期阶段是不可缺少的。这类结果回答的是“如果这一功能被破坏会发生什么”。而本文的神经反馈研究并不是破坏前额叶功能，而是在功能存在的基础上，通过调控前额 P3 这一具体神经反应，考察情绪调节是否能够得到改善，回答的是“在功能正常的情况下，把相关神经反应调强或调弱是否会带来更好的调节效果”。因此，将 TMS 研究和神经反馈研究直接用“与此相似”连接，容易让读者误以为它们属于同一种因果证据或反映相同的机制，这在逻辑上并不严谨。

回复：感谢您的仔细阅读。我们已将其修改为“在此基础上”，见第 14 页 14 行。

意见 14.语言表达不严谨：全文多处（结果、讨论等部分）仅依据相关分析的结果，便使用“预测”等因果的表述。比如：3.2 部分第一段作者提到“相关分析显示，两组被试的 NF 训练效率与情绪调节任务下的前额 P3 幅度的变化(后测减去前测)显著相关($r = 0.281, p = 0.024$; 图 3B)，提示 NF 训练效率可以预测情绪调节的前额 P3 幅度”，用相关结果直接得出“预测”的结论不严谨。相关分析只能说明 NF 训练效率与 P3 变化在统计上存在正向联系，不建议使用“可以预测”等因果性语言。

回复：我们已删除了结果部分的不当表述，修改了讨论部分的相关内容，见第 13 页 19、25 行、第 14 页 16 行。

意见 15.文中存在多处语言表达、错别字、统计符号书写等不规范之处，比如：2.4 部分第一段，作者提到“NF 训练采用双盲设计”，双盲设计的表述不准确，双盲指的是主试和被试均不了解实验目的、实验分组，而非仅仅不了解分组，并且，在实验过程中被试能够看到温度计反馈，不算严格意义的双盲。又如：多处样本量等统计符号没有斜体，还有一些错别字、中英文标点混用等问题。建议全文仔细检查修改。

回复：感谢您的建议。本轮修改，我们全文仔细检查，并修改了所有不规范之处，例如 6 页 5-6 行（之前写到“双盲”）。此外，我们对全文斜体、错别字等进行了检查和修正。

第二轮

审稿人 1 意见：

作者的回复已经解决了我对论文的问题。没有其他意见。

审稿人 2 意见：作者对论文进行了修改，但还有如下问题与作者探讨：

意见 1: 引言中对采用相反调控解决 ERP-NF 训练效果不一致问题的修改不充分。作者在正文第 3 页 14-15 行修改为：“相反调控方向可有效排除期望效应、安慰剂效应”。1) 如果没有设置无反馈控制条件，仅凭“上调组”和“下调组”很难直接排除安慰剂或练习效应。同时，请作者明确说明：相反方向设计究竟用于检验“方向特异性”的哪一条证据链，以及该设计如何针对以往 ERP-NF 不一致结果提供解释性信息。

2) 此外，建议作者补充：被试是否被告知训练方向（上调/下调）及其预期效果？两组是否在反馈形式、任务负荷、奖励规则、实验接触时间上完全等同？

回复: 感谢您指出的问题。

1) 以往 ERP-NF 研究多采用单方向调控加假反馈对照组的设计。一方面，单向调控的效应量肯定没有两个反向组之差的效应量大，因此实验组和对照组的组间差异较小，有的研究没有发现 NF 的训练效果。另一方面，真反馈和假反馈两组被试在 NF 过程中的心态不同，例如假反馈组发现自己的努力不能造成脑内信号往希望的方向改变，会有挫败感，而这又会导致安慰剂效应（如果有的话）等混淆变量对两组实验结果的影响程度不同。这些因素都对结果产生了干扰。这可能是已有研究结果不一致的原因之一。

基于此，本研究采用上调与下调的相反方向设计，我们的逻辑在于：既然 P3 表征执行控制，两组反向 NF 的结果会导致情绪调节效果也呈现出相反方向的变化。本轮修改，我们在结果中补充了两组前后测的比较（见第 10 页 13-16 行、第 12 页 5-8 行、23-26 行）。结果表明，两组的情绪调节效果的确改变方向相反。同时，上调 vs. 下调 NF 组别的设计还可以排除您提到的安慰剂效应等混淆，因为两组被试的实验流程和实验设计完全一致，且所有被试接收到的都是真反馈，他们发现通过自己努力可以使脑信号往规定的方向改变，此时安慰剂和练习效应在两组中的方向相同（对指标造成均提高或均降低的影响）、程度相当，可以通过减法原则消除干扰。因此，真反馈的两个组别正好可以排除您提及的混淆因素。

我们已在引言中补充了相关表述（见第 3 页 13-18 行）。

2) 已在方法部分补充：“实验前和实验过程中主试与被试都未被告知分组信息及其预期效果。两组在反馈形式、任务负荷、奖励规则、实验接触时间上完全等同。”（见第 6 页 11-13 行）

意见 2: 论文的逻辑结构仍然需要进一步明确。当前逻辑是：情绪调节任务时 NF 训练 → P3 幅度改变 → 说明执行控制增强 → 执行控制增强导致情绪调节改善。作者在文中补充：“前额 P3 是执行控制功能的神经标志物（Bernat et al., 2011; Moser et al., 2014; Yuan et al., 2015, 2023）”，但 NF 训练时的任务是情绪调节（“参与者需在观看情绪图片时探索并使用适用于自己的策略，包括但不限于：调节图片带来的负性情绪，回忆，转移注意，心算，抑制等”），不是执行功能任务，为什么执行情绪调节任务时改变的 P3 波幅，说明执行功能改变呢？P3 波幅在不同的任务范式中代表的含义是不同的。NF 训练任务为什么不选择执行功能任务呢？

回复: 我们建议被试在 NF 训练过程中采用各种高度依赖执行功能的任务，包括但不限于调节图片带来的负性情绪，回忆，转移注意，心算，抑制。这里为被试提供多种执行功能任务的选项是考虑到被试的个体差异，他们可以选择自己最能驾驭且 NF 训练效果最好的任务。同时，不同被试间采用不同的执行功能任务可以减弱固定任务对非靶点脑区神经信号的固定改变，后者会干扰因果推论（Kvamme et al., 2022）。此外，不使用固定的 NF 训

练任务也是 NF 研究常用的方法 (Lubianiker et al., 2026), 这也写入了 NF 研究共识 (Ros et al., 2020), 我们在修改稿中补充了这点 (见第 6 页 17-19 行)。

参考文献:

- Kvamme, T. L., Ros, T., & Overgaard, M. (2022). Can neurofeedback provide evidence of direct brain-behavior causality? *Neuroimage*, 258, 119400.
- Lubianiker, N., Koren, T., Djerasi, M., Sirotkin, M., Singer, N., Jalon, I., Lerner, A., Sar-El, R., Sharon, H., Shahar, M., Azulay-Debby, H., Rolls, A., & Hendler, T. (2026). Upregulation of reward mesolimbic activity and immune response to vaccination: a randomized controlled trial. *Nature Medicine*, 32(2), 572–581.
- Ros, T., Enriquez-Geppert, S., Zotev, V., Young, K. D., Wood, G., Whitfield-Gabrieli, S., Wan, F., Vuilleumier, P., Vialatte, F., Van De Ville, D., Todder, D., Surmeli, T., Sulzer, J. S., Strehl, U., Sterman, M. B., Steiner, N. J., Sorger, B., Soekadar, S. R., Sitaram, R., Sherlin, L. H., ... Thibault, R. T. (2020). Consensus on the reporting and experimental design of clinical and cognitive-behavioural neurofeedback studies (CRED-nf checklist). *Brain*, 143(6), 1674–1685.

意见 3: 论文在修改稿中补充: “本研究探索的两个预测变量: 基线时期的前额 P3 幅度 (Alkoby et al., 2018; Haugg et al., 2020)、日常使用认知重评策略的频率”, 将日常使用认知重评策略的频率作为预测变量不太合适, 因为这不是本文探讨的核心问题。

回复: 将日常使用认知重评策略的频率作为预测变量是因为已有文献发现日常使用认知重评策略的频率会影响靶向 VLPFC 神经反馈的训练效率 (Li et al., 2025a)。本研究想验证这一结果, 并没有将之作为研究的核心变量和核心分析目标。

意见 4: 关于 $t=0$ 的问题, 作者解释称图片分组为人为匹配, 因此统计结果可以“如此完美”。但即便人为匹配, 在真实数据中出现 $t=0.000$ 、 $p=1.000$ 仍比较罕见, 容易引发读者对匹配方式的疑问。建议在正文中详细说明匹配的具体方法。例如, 图片是否基于 IAPS 标准评分进行匹配? 是逐一配对匹配, 还是整体均值匹配?

回复: 感谢审稿人严谨的关注。图片分组的效价唤醒度编号等信息见第一轮审稿审稿人 2 意见 8 附表, 图片评分的统计结果见下图。确实出现了 $t=0.000$, $p=1.000$ 。

图片匹配流程如下: 1) 纳入 IAPS 图片 100 张, 效价控制在 2.5 分左右, 唤醒度控制在 5 分左右。2) 对于类型相似的图片 (例如都是蜘蛛或者都是枪), 每组各分配一张。3) 对各不相同的图片逐一匹配好效价, 分配至两组。4) 计算两组图片的唤醒度 p 值 (独立样本 t 检验), 如果有显著差异, 每次从 IAPS 图片库重新选择、替换单组 1~2 张图片, 直到两组图片的唤醒度无显著差异 (见第 5 页 18-22 行)。

	分组	个案数	平均值	标准差	标准误差平均值
效价	A	30	2.4827	.41925	.07654
	B	30	2.4827	.38587	.07045
唤醒度	A	30	5.3093	.59512	.10865
	B	30	5.3373	.68213	.12454

		莱文方差等同性检验		平均值等同性 t 检验							
		F	显著性	t	自由度	显著性 (双尾)	平均值差值	标准误差差值	差值 95% 置信区间	下限	上限
效价	假定等方差	.628	.431	.000	58	1.000	.00000	.10403		-.20824	.20824
	不假定等方差			.000	57.605	1.000	.00000	.10403		-.20827	.20827
唤醒度	假定等方差	.375	.543	-.169	58	.866	-.02800	.16528		-.35883	.30283
	不假定等方差			-.169	56.953	.866	-.02800	.16528		-.35896	.30296

上图：子集 A 与 B 效价和唤醒度的独立样本 t 检验结果。

意见 5: 2.1 部分，作者基于 $\eta^2 = 0.071$ 使用 G*Power 进行事前样本量估计，并报告统计功效达 99%。该部分仍需澄清以下问题：

(1) G*Power 需输入 Cohen's f，作者是否对偏 η^2 进行了转换？请报告具体转换过程与最终参数。

(2) 功效达 99% 偏高，建议补充 α 水平、测量次数及相关系数等关键设定参数。

(3) 功效估计基于重复测量 ANOVA，而本文实际分析采用 LMER，这两种方法在假设条件、模型结构和数据处理方式等方面存在一定差异。由于方法不一致，基于重复测量 ANOVA 得到的样本量估计可能仅为近似参考。建议作者说明这种近似参考的合理性和局限性，并考虑是否采用更合适的方法进行样本量估计。

回复：

1) 样本量估算流程如下：在 G*Power 3.1.9 软件中 (F tests, ANOVA: repeated measures, within-between interaction)，基于已有 ERP-NF 研究报告的 $\eta_p^2 = 0.071$ 计算 Cohen's f，并将其导入样本量计算的参数设置中。具体参数设定为： $\alpha = 0.05$ ，power = 0.99，组别数量 = 2，重复测量次数 = 2，重复测量间相关系数 = 0.5，非球形校正系数 (nonsphericity) = 1。输出结果显示：所需总样本量为 64 人，真实 power = 0.9917。

2) 已在本轮修改稿的 2.1 被试部分进行补 α 水平、测量次数及相关系数等关键设定参数，并额外计算了 Power = 95% 时所需的最小样本量 (见第 4 页 10-13 行)。

3) 我们使用基于重复测量 ANOVA 得到的样本量估计基于以下两点原因：第一，与本研究最相关的 NF 研究 (Eschmann et al., 2020; Eschmann & Mecklinger, 2022; Pei et al., 2020) 均采用方差分析进行结果统计，因此能用于本研究样本量估计的效应量值，均来自 ANOVA，我们采用同一套统计方法 (即 ANOVA) 去进行样本量估计，得到的样本量最准确 (因为无需换算)。第二，目前尚未有成熟的、针对线性混合模型的专用样本量估计软件，基于重复测量 ANOVA 进行样本量估算，是目前采用线性混合模型研究的普遍做法 (例如 Li, J. et al., 2026; Li et al., 2025b)。此外，样本量估计采用 ANOVA 方法，但数据分析使用贝叶斯线性混合模型的做法也见于近期高水平文献 (如 Lubianiker et al., 2026)。

本研究采用 LMER 数据分析，实际提高了统计效力，这也增强了我们对估算的样本量得出实验效应的信心和能力。

参考文献:

- Li, J., Chen, S., Zhang, L., Weng, L., Lin, X., Tu, Y., & Peng, W. (2026). Dynamic expectation strength and precision shape human pain perception through shared and dissociable α -oscillatory mechanisms. *PLoS Biology*, 24(3), e3003675.
- Lubianiker, N., Koren, T., Djerasi, M., Sirotkin, M., Singer, N., Jalon, I., Lerner, A., Sar-El, R., Sharon, H., Shahar, M., Azulay-Debby, H., Rolls, A., & Hendler, T. (2026). Upregulation of reward mesolimbic activity and immune response to vaccination: a randomized controlled trial. *Nature Medicine*, 32(2), 572–581.

意见 6: 3.4 部分，作者进行的中介分析以 NF 训练效率作为自变量，探讨其通过情绪调节任务中的前额 P3 变化影响情绪改善。然而，NF 训练效率并非实验随机操纵变量，而是被试层级的连续指标，其本身可能部分反映组别差异。因此，目前尚不清楚该中介路径是否在统计上替代了实验操纵效应。1) 建议作者说明 NF 训练效率与组别之间的关系强度，2) 并明确在中介模型中是否控制了组别变量。3) 同时，主分析基于组 \times 前后测的实验设计，而中介分析基于被试层级聚合变量，作者需进一步阐明不同分析层级之间的逻辑衔接，以避免因果解释过度外推。

回复: 感谢审稿人宝贵的意见。1) NF 训练效率是训练过程中 P3 幅度的改变程度和方向。上调组的 NF 训练效率为正，下调组的效率为负。因此，组别确实显著影响了训练效率的值。

2) 在本轮修改中，我们在中介模型中控制了组别变量，将其作为协变量纳入中介模型（见第 9 页 8 行、第 13 页 4-9 行）。

3) 采用基于被试层级的中介分析，原因有以下两点：一、NF 学习成功与否，目前普遍采用训练轮次的回归斜率来衡量（见元分析 Haugg et al., 2020），大概 50% 的被试不能从 NF 中达到明显训练效果。在本研究中，NF 训练过程中 P3 幅度的调控程度直接影响了情绪调节效果的变化——即使在上调组内，未能学会上调 P3 的被试（NF 效率低）其情绪调节效果也没有显著提升。因此，我们认为使用更直接的 NF 效率作为中介分析的自变量更为合理。第二，本研究分为上调和下调组两个相反方向的调控。因此，合并两组的 NF 训练效率数据，不仅在统计上涵盖了实验操纵效应，还因样本量的增加提升了中介分析的统计效力和估计的可靠性。我们修改和增补了统计部分的表述，以阐明不同分析层级之间的逻辑衔接（见第 8 页 26-28、第 9 页 1-6 行）。

意见 7: 结果部分，作者统计报告同时使用了 t 值和 Z 值（例如主模型报告 t，事后比较报告 Z），但文中未说明不同统计量的来源及其计算方法。鉴于本研究主要采用线性混合效应模型（LMM，连续因变量），通常报告 t 值及相应自由度即可。目前统计指标使用不统一，可能引起读者对检验方法一致性的疑问。

回复: 感谢审稿人细致的意见。我们使用 lme4 进行线性混合效应模型输出因子间主效应和交互效应，如果主效应或者交互作用显著，我们使用 emmeans 包进行事后比较。这一分析流程是目前最通用的线性混合模型输出和结果报告方法（Duque et al., 2023; Li, Y. et al., 2025b）。之前我们事后比较 emmeans 使用 asymptotic 的方法（输出为 Z），现在我们改为了 satterthwaite（输出为 t），并修改了结果部分。此外，为了方便读者的理解，我们在方法部分补充说明了事后比较的方法（见第 8 页 12-14 行）。

线性混合效应模型分析代码样例:

```
power = lmer(data ~ group*session + (1|subj), data=a)
```

事后比较代码样例:

```
emm_options(lmer.df = "asymptotic")已修改为 emm_options(lmer.df = "satterthwaite")
```

```
pairwise_comparison<-emmeans(power, pairwise~group|session, adjust = "fdr")
```

参考文献:

Fernandez-Duque, M., Hayakawa, S., & Marian, V. (2023). Speakers of different languages remember visual scenes differently. *Science Advances*, 9(33), eadh0064.

意见 8: 结果部分, 作者报告了多项相关分析 (包括 NF 效率与 P3 变化、情绪变化、LPP 变化、重评难度等), 但文中未见对多重比较进行校正的说明。建议作者明确说明是否对相关分析进行多重比较校正 (如 FDR 或 Bonferroni); 若未进行校正, 请说明理由。

回复: 感谢您的建议, 我们本轮使用 FDR 对所有相关分析进行了多重比较校正, 并修改了方法部分 (见第 8 页 23-24 行) 和结果部分 (见第 10 页 17 行和第 12 页 9-10、16-17、27-28 行)。

第三轮

审稿人 2 意见: 作者的回复已经解决了我对论文提出的问题。

编委意见: 同意外审意见, 建议录用。

主编意见: 同意发表。