

《心理学报》审稿意见与作者回应

题目：第三方惩罚的神经机制：来自经颅直流电刺激的证据

作者：殷西乐 李建标 陈思宇 刘晓丽 郝洁

第一轮

审稿人 1 意见：

本研究通过经颅直流电刺激技术对第三方惩罚行为的心理机制进行了较深入的探讨，试图区分其中情绪与自利两种主要动机的不同作用，实验设计比较合理，所得结论对原有理论有一定的发展。现就文章可能存在的若干问题与作者商榷。

意见 1：文章标题为：“社会规范遵从的神经机制：来自经颅直流电刺激的证据”，但通观全文，实际上只是探讨了第三方惩罚行为本身的神经机制。第三方惩罚虽然是社会规范遵从（或社会规范维护）的一种典型行为，但并不能等同于社会规范遵从，如果文章要探讨的是社会规范遵从的神经机制，我认为至少需要再分析一两种其它的社会规范遵从行为。

回应：首先，非常感谢审稿专家的评论和意见。

正如审稿专家所言，相比第三方惩罚，社会规范遵从是一个更加宽泛的概念。诸如最后通牒博弈中作为优势方的提议者的提议行为(Ruff et al., 2013)，或者作为劣势方的回应者（第二方）的拒绝或惩罚行为(Knoch et al., 2008; 王益文等, 2014)，抑或是作为利益无关的第三方乃至第四方的惩罚行为等(Fehr & Fischbacher, 2004a, 2004b; 陈思静等, 2015; Li et al., 2018)，都能反映个体对社会规范的遵从。在这些行为反应中，第三方惩罚被认为是社会规范得以大规模维系的基石和关键(Fehr & Fischbacher, 2004a; Fehr & Gächter, 2002)。Fehr & Fischbacher (2004b, p. 186)认为“第三方惩罚可以十分‘干净’地刻画和研究社会规范，因为规范违反者的行为不会以任何方式影响第三方的经济利益”，因此研究者经常将第三方惩罚和社会规范遵从交替使用。基于这一考虑，我们在原稿题目中采用了“社会规范遵从的神经机制”。

然而，正如专家所言，第三方惩罚毕竟只是社会规范遵从的一种典型行为，其并不能等同于社会规范遵从。例如，Ruff 等(2013)在 *Science* 发表的“Changing social norm compliance with noninvasive brain stimulation”，事实上考察的是提议者的提议行为。因此，为了更加精确反映本文研究问题，我们在修改稿中将题目改为“第三方惩罚的神经机制：来自经颅直流电刺激的证据”。基于此，我们也修改了正文的相关内容，在修改稿中，引言的第一段即引

出第三方惩罚及其与社会规范的关联,同时将正文有关社会规范遵从的表述改为第三方惩罚或者第三方的社会规范遵从。我们依据专家意见修改的内容已在正文用红色标注。

最后,审稿专家的评论实际上指出了本文的拓展方向:本文对第三方惩罚神经机制的探讨是否同样适用于其他的社会规范遵从行为,这是值得进一步探讨的话题,也是我们后期探索的重点。再次感谢专家意见。

意见 2: 自检报告中对计划样本量的说明是简单地对比其它研究的样本量,但不同实验设计和效应量的期望下样本数量可能有较大差别,不应做简单的数量对比。建议作者在自检报告中写明实验设计时参照的相关研究效应量及依据这些数据所计算出的预期样本数量以及所使用的程序或软件。

回应: 感谢审稿专家的建议。根据专家意见,我们使用 G*Power 软件计算了实验所需样本量。Zhang 等人(2016)和 Yin 等人(2017)发现,关于规范遵从的 tDCS 研究的效应量通常在中等和较大之间。基于此,我们将 F 检验的效应量首先设置为中等,根据 Cohen(1988)和 Lakens (2013)的建议,其对应的 η^2 数值为 0.06 (G*Power 转换后的 effect size f 为 0.2526),犯 I 类错误的概率 (α err prob) 为 0.05,组数和重复测量水平数均为 3,水平间相关 (Corr among rep measures) 默认为 0.5, Nonsphericity correction ϵ 为 1, 检验效能 Power(1- β err prob) 为 0.95 时,计算出的样本量 (total sample size) 为 51。保持其他数值不变,将效应量调整为大 ($\eta^2 = 0.14$) 时,样本量为 24。本文有效样本量为 89 人,满足上述设定要求。本修改稿 12 页自检报告部分已按上述表述重新修改。

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology*, 4, 863.

Yin, Y., Yu, H., Su, Z., Zhang, Y., & Zhou, X. (2017). Lateral prefrontal/orbitofrontal cortex has different roles in norm compliance in gain and loss domains: a transcranial direct current stimulation study. *European Journal of Neuroscience*, 46(5), 2088-2095.

Zhang, Y., Yu, H., Yin, Y., & Zhou, X. (2016). Intention modulates the effect of punishment threat in norm enforcement via the lateral orbitofrontal cortex. *Journal of Neuroscience*, 36(35), 9217-9226.

意见 3: 实验采用了大学生被试,根据以往研究,某些专业(如经济学相关专业)大学生被试可能在经济博弈类实验中跟其它专业被试的行为有显著差异,建议作者用一句话报告一下被试专业的情况。

回应: 依据专家建议,本文实验 89 名有效被试中有 69 名被试专业为经管类(包括经济学、

金融学、会计学和企业管理等), 占比为 77.53%。另外, 被试专业类型在不同刺激组中的分布比较接近: 阴极、阳极和虚拟组分别包括 6、6 和 8 名非经管被试。我们在正文第 17 页的实验设计部分加入这一内容。

意见 4: 实验设计中“任务 2 中被试无需汇报情绪反应”, 这样设计是否会造成无关变量的影响——即任务 1 与任务 2 的行为差别可能是由于汇报/不汇报情绪反应而造成的?

回应: 感谢专家意见。正如专家所言, 我们在实验设计时需要做出权衡: 一方面, 要求被试在任务 1 和任务 2 均汇报情绪反应可以确保两个任务的行为差别不受汇报与否的影响; 另一方面, 被试在有成本设置的任务 2 中汇报的情绪反应容易遭受“污染”, 即其可能受到自利因素的干扰。况且, 被试面对相同的分配方案重复汇报情绪反应时可能产生学习效应或需求效应等(本实验中此类效应对情绪反应的影响可能比对惩罚决策的影响大得多, 因为我们通常无法对自我汇报的情绪反应进行“价值诱导”的实验激励)。就本文研究目的而言, 我们希望相对“干净”地比较不同 tDCS 组被试的情绪反应是否存在差异进而考察第三方惩罚的情绪机制, 因此更倾向于规避后者的影响。进一步, 由于不同刺激组的被试面临的实验场景和流程完全相同, 即便被试行为在任务 1 中受到情绪汇报的影响, 这一偏差也会在对不同刺激组的结果比较中很大程度上得以消除。基于以上考虑, 我们在实验设计时还是更倾向于让被试只在零成本设置的任务 1 汇报情绪反应, 再次感谢专家意见。依据上述分析, 我们在正文实验设计部分第 21 页的脚注 5 中进行了相应修改。

意见 5: “情绪反应是一个从非常愤怒到非常赞赏的 5 分量表”, 愤怒与赞赏并不能算是一对完全相反的情绪, 也就是说, 当我选择非常赞赏时, 可能是因为我想表达自己“一点也不愤怒”, 但并不是“非常赞赏”。为什么不用两个问题分别测量愤怒(从非常愤怒到一点也不愤怒)和赞赏?

回应: 非常感谢审稿专家的这一意见。首先, 作者核对了实验程序和说明, 发现此处存在笔误。原文应为“情绪反应是一个从非常愤怒到非常高兴的 5 分量表”。正如专家所言, 这一测度方式下被试可能存在通过选择正(负)的量表分值来表达没有其他负(正)性情绪的可能。对此我们进行了如下分析。本文结果显示, 当独裁者给出公平的分配方案时, 情绪值小于零的比例为 0, 当独裁者给出不公平的分配方案时, 情绪值大于零的比例仅为 5.62%。因此, 当被试存在负性(正性)情绪时, 他一般会选择相应的负(正)的分值, 其情绪汇报结果比较真实地反应了其正/负情绪的内在偏好。由于本文主要分析情绪特别是负性情绪(可

能不局限于愤怒)在第三方惩罚中的作用,我们更关注被试汇报的情绪分值是否存在系统性偏差(如将负性情绪汇报成正值),而不太关注该情绪的具体类别,而且通过自我汇报的方式去判别不同的情绪类别其说服力可能也不够强。我们在量表中使用愤怒作为负性情绪的测度指标,主要是基于已有的研究观点(Fehr & Fischbacher, 2004a; Ruff et al., 2013; Jordan, McAuliffe & Rand, 2016),如果要严谨地厘清哪一种情绪在第三方惩罚中起到的作用更大,那么审稿专家的建议方式就显得尤为必要(再辅以 FaceReader 等面部表情分析软件来判别具体情绪类型效果可能更好)。综上,本文的量表测度不会显著影响我们的实验结果,同时我们认为审稿专家的建议更加严谨,为后续研究提供了更多思路,再次感谢专家的细致意见。

意见 6:“第二,三种分配方案 0、30、50 中有两种结果重复呈现给被试,这样被试一共观察 5 次分配结果。这一分配结果重复呈现的设计有助于减少实验结果的人为操纵痕迹。”为什么这种设计可以减少实验结果的人为操纵痕迹?重复的两种结果是随机的吗?

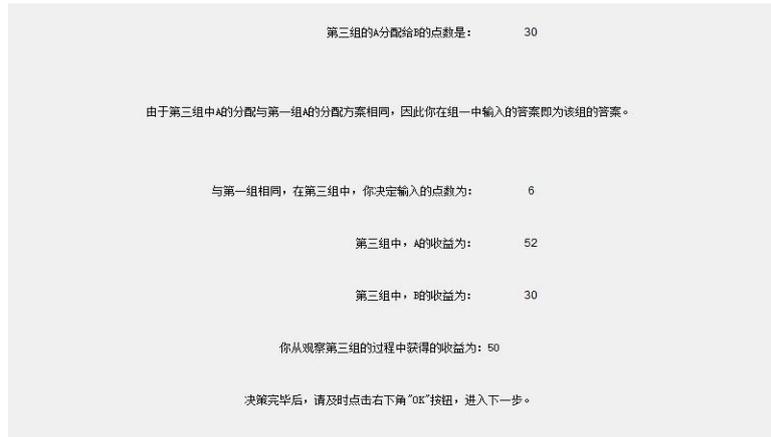
回应:感谢专家意见。如果本文没有重复设计,被试将观察到三组独裁者的分配方案恰好不同(分别为 0、30 和 50),而这一可能性相对较小。例如,Fehr & Fischbacher(2004a)的行为实验结果发现,第三方惩罚中独裁者给出的分配方案比较分散,其给出不同分配方案的比例大致在 10%-25%之间。依据这一结果假定独裁者的分配方案服从均匀分布(即其选择 0、30 和 50 的概率均为三分之一),那么三组的独裁者恰好选择了不同方案而没有任何重复的概率仅为九分之二(22.22%)。为了消除被试可能的怀疑,我们将其中两种分配方案重复一次。这样,五组当中有两组是重复呈现的,我们认为这有助于减少实验设计的人为操纵痕迹。

与已有研究类似(Ruff et al., 2013; 王益文等, 2014),本文采用伪随机方式控制随机设计对被试行为的影响。我们对五组分配方案的呈现次序进行了伪随机排列,规则是确保(0, 30, 50)三种分配方案至少出现一次且至多出现两次,且两个任务的呈现顺序在被试间 ABBA 平衡。具体方式为,第一场(session)实验中被试先进行零成本任务 1,进行零成本任务 1 时计算机按照上述规则生成 5 组分配方案的呈现顺序 A,随后在有成本任务 2 时按照上述规则随机生成 5 组分配方案的呈现顺序 B。在接下下的第二场实验中被试先进行有成本任务 2,且其 5 组分配方案的呈现顺序为 A,随后进行零成本任务 1,且 5 组分配方案的呈现顺序为 B。也就是说,计算机只在奇数场次实验(1, 3, 5……)的任务 1 和任务 2 生成 5 组分配方案的伪随机顺序,而接下下的偶数场实验(2, 4, 6……)将任务呈现顺序颠倒,但两个任务的 5 组分配方案的呈现顺序按照前一场实验的顺序(AB)进行。这一设计既控制了任务呈现的顺序效应,又确保了零成本设置和有成本设置的 5 组分配方案的呈现顺序在总体样

本上是一致的。

为了避免上述随机设计干扰 tDCS 效应，我们每一场实验（session）招募三名被试，且三名被试分别接受阴极、阳极和虚拟刺激。这样，上述设计既减少了人为操作痕迹，又确保了分配方案及其呈现顺序在不同 tDCS 组中完全相同。

另外，对于重复呈现的结果，被试无需再次进行决策，其决策和收益将由计算机自动按照其在前序实验的决策处理。如下图所示。



重复呈现分配方案的计算机界面

依据上述分析，我们在正文实验设计部分第 20 页的脚注 4 中进行了相应修改。

意见 7: 文中称：每观察一个小组，被试需要回答三个问题。①在看到独裁者的分配结果之前，被试需要回答他们对社会规范的信念，即“你认为 A（独裁者）‘应该’向 B（接受者）分配多少 G\$（即 A 向 B 分配多少 G\$是公平的）”。社会规范信念是稳定的还是短时间内会波动的？为什么不只在整个实验前测量一次而需要每一轮实验都测一次？另外，如果用独裁者博弈中的分配高于/低于第三方社会规范信念来定义“公平/不公平”是否比单纯的使用统一数值更合适？

回应: 谢审稿专家指出这一问题。正如审稿专家所指出，我们对社会规范信念的测度在实验伊始就进行测量且整个实验过程中只测量一次。我们在原文汇报相应结果时提到“单因素方差分析结果表明第三方对社会规范信念在不同 tDCS 刺激组间的差异不显著 ($F_{2,86}=0.65$, $p=0.524$)。”从自由度的汇报结果可以佐证我们对社会规范信念只测量了一次。我们非常感谢审稿专家发现这一错误，并在修改稿中 22 页进行了相应修改。

我们对公平的定义参考 Civai 等(2015)的分类，0 代表不公平的分配结果（unfair），30 代表中等方案（mid-value），50 则代表公平的分配方案（fair），其中不公平的分配结果明显地违反了社会规范。同时，审稿专家也指出了用社会规范信念与分配方案的比较作为分类标

准。我们认为两种分类方式都比较合理，因此两种分类方式所得结果应该一致。

本文结果发现，社会规范信念的均值为 32.584（方差为 16.60），非常接近我们对中等的定义值 30。同时按照专家建议的分类方式，我们重新定义了分配方案的公平与否（分配额小于社会规范信念定义为不公平，否则为公平），结果与本文已有结果相一致：当独裁者的分配方案不公平时，零成本设置的单因素方差分析显示不同刺激组的情绪（ $F_{2, 121}=3.6$ ， $p=0.03$ ， $\eta^2=0.096$ ）和惩罚值（ $F_{2, 121}=6.41$ ， $p=0.002$ ， $\eta^2=0.056$ ）均存在显著差异，且零成本和有成本设置的惩罚差异的 tDCS 效应也显著（ $F_{2, 121}=3.65$ ， $p=0.029$ ， $\eta^2=0.057$ ）。当独裁者的分配方案公平时，上述结果均不显著。

上述结果表明，对公平两种分类方式的效果是一致的。我们在修改稿中（26 页标红部分）将上述结果作为稳健性分析结果汇报。

意见 8：“当独裁者的分配方案不公平时，单因素方差分析显示不同刺激组的情绪存在显著差异，且 tDCS 主效应可以解释 21.9%的变异量（ $F_{2, 86}=12.07$ ， $p<0.001$ ， $\eta^2=0.219$ ）。两样本 T 检验结果显示，阴极组中第三方的情绪均值为-1.38，显著小于虚拟组的-0.97（ $t_{57}=2.12$ ， $p=0.019$ ， $d=0.553$ ），阳极组中第三方的情绪均值为-0.33，显著大于虚拟组（ $t_{58}=2.90$ ， $p=0.003$ ， $d=0.748$ ）。”这里及下文中的两两比较为什么不直接在单因素方差分析中进行（如用 SNK 法等），要再做两样本的 t 检验？

回应：感谢专家意见。按照专家建议，我们将对 tDCS 组间效应的两两比较更改为事后比较检验。结果显示直接采用单因素方差分析中的事后比较检验（SNK 和 LSD），和两样本 t 检验的结果是一致的。我们在修改稿中将两样本 t 检验结果更新为 SNK 的事后比较结果。

意见 9：讨论部分关于本文思路和逻辑的解释与分析应该放到引言中。

回应：感谢专家意见。按照专家建议，我们将原文关于我们如何在已有研究的基础上进行了改进等相关思路和逻辑的解释放到了引言中。具体见 14-16 页正文前言的标红部分。再次向审稿专家的宝贵意见表示感谢。

参考文献：

- 陈思静, 何铨, 马剑虹. (2015). 第三方惩罚对合作行为的影响: 基于社会规范激活的解释. *心理学报*, 47(3), 389-405.
- 王益文, 张振, 张蔚, 黄亮, 郭丰波, & 原胜. (2014). 群体身份调节最后通牒博弈的公平关注. *心理学报*, 46(12), 1850-1859.
- Civai, C., Miniussi, C., & Rumiati, R. I. (2015). Medial prefrontal cortex reacts to unfairness if this damages the self: a tDCS study. *Social cognitive and affective neuroscience*, 10(8), 1054-1060.

- Fehr, E., & Fischbacher, U. (2004a). Third-party punishment and social norms. *Evolution and human behavior*, 25(2), 63–87.
- Fehr, E., & Fischbacher, U. (2004b). Social norms and human cooperation. *Trends in cognitive sciences*, 8(4), 185-190.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137–140.
- Jordan, J., McAuliffe, K., & Rand, D. (2016). The effects of endowment size and strategy method on third party punishment. *Experimental Economics*, 19, 1–23.
- Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A., & Fehr, E. (2008). Studying the neurobiology of social interaction with transcranial direct current stimulation—the example of punishing unfairness. *Cerebral Cortex*, 18(9), 1987–1990.
- Li, J., Li, S., Wang, C., Zhu, C., Niu, X., Liu, X., & Wang, G. (2018). An event-related social evaluation of pro-social behaviors. *Frontiers in Psychology*, 9, 932.
- Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science*, 342(6157), 482–484.
-

审稿人 2 意见：

文章基于第三方惩罚的独裁者博弈框架，借助经颅直流电刺激（tDCS）技术对右侧背外侧前额叶区域（DLPFC）进行干预，同时依据第三方是否需要为其惩罚付出成本设计了零成本和有成本两个实验任务。旨在通过改变 RDL PFC 区域的神经活跃水平来影响第三方的负性情绪和自利倾向，从而观察第三方惩罚行为的变化，以此分析第三方惩罚行为的神经机制，即第三方遵从社会规范的行为是否受到情绪和自利机制的影响。

文章的实验设计具有较大的创新性，并提出了一个富有创见且得到一定程度验证的关于第三方惩罚行为的神经机制（负向情绪和理性自利动机的冲突）。文章所使用的科学方法（神经刺激技术）适当，实验设计规范，实验结果分析与讨论都比较充分，是一篇有较高学术价值和科学规范的论文。当然，文章在实验设计、研究结论方面还有一些细节之处需要澄清和解释。

意见 1：作者使用的是单侧的 tDCS 刺激方法，选取了 CZ 为回路脑区，RDL PFC 为目标脑区，这两个脑区之间的距离并没有相隔太远，作者如何避免或解释在刺激启动时，这两个脑区不存在电流的相互干扰情况。

回应：首先，非常感谢审稿专家的评论和意见。

我们选取 Cz 作为参考电极主要基于以下考虑。对于将 RDL PFC 作为目标脑区的 tDCS 研究，选取 Cz 作为参考电极位置是比较常见的选择（Ruff et al., 2013; Harty et al., 2014; Li et al., 2017）。头皮位置 Cz 对应于右侧和左侧中央沟的汇合，相比其他位置（如眼眶），该区

域及其附近的脑区对认知和情绪活动的影响通常较小。特别是，Spitzer 等（2007）的 fMRI 研究发现 Cz 附近的脑区没有在独裁者博弈中激活，Ruff 等（2013）也指出该区域作为 tDCS 参考电极不会干扰社会规范相关的神经活动。另外，Sellaro 等（2016）建议可以将参考电极设置在头部以外（如肩部三角肌位置或胳膊），这也是比较好的选择。但本实验进行时正值冬天，将浸润生理盐水的电极片放在被试肩部或胳膊不太具备可行性，或者可能会因这一设计给被试带来的不适感反而对实验结果造成干扰。基于上述考虑，我们选取了 Cz 作为回路脑区。最后，本文使用 Soterix 医药公司 (Soterix Medical Inc., New York) 提供的 HD-Explore 软件模拟了大脑区域接受刺激后的电流模式（图 1，箭头表示电流方向）。从中可以看出，无论是阳极刺激还是阴极刺激，均在右侧 DLPFC 区域产生了较高的场强 (Field Intensity)，这表明刺激有效改变了该区域的活跃水平。依据上述分析，我们在正文实验设计部分第 18 页的脚注 2 中进行了相应修改，且已用蓝色标注。

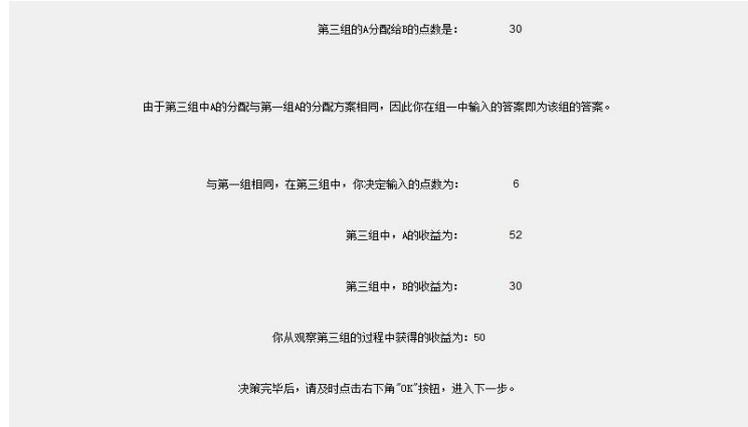
意见 2: 实验设计中，三种分配方案 0、30、50 中有两种结果重复呈现给被试，这样被试一共观察 5 次分配结果。这一分配结果重复呈现的设计是为了减少实验结果的人为操纵痕迹。那么，具体哪两种分配方案是重复的？文中并没有提及，以及任务 1 和任务 2 的分配方案呈现内容和顺序是否是一样的？这一设计对于实验结果有一定影响，需要说明清楚。

回应: 感谢专家意见。与已有研究类似 (Ruff et al., 2013; 王益文等, 2014)，本文采用伪随机方式控制随机设计对被试行为的影响。我们对任务 1 和任务 2 中五组分配方案的出现次序进行伪随机排列，规则是确保 (0, 30, 50) 三种分配方案至少出现一次且至多出现两次，且两个任务的呈现顺序在被试间 ABBA 平衡。具体方式为，第一场 (session) 实验中被试先进行零成本任务 1，进行零成本任务 1 时计算机按照上述规则生成 5 组分配方案的呈现顺序 A，随后在有成本任务 2 时按照上述规则随机生成 5 组分配方案的呈现顺序 B。在接下来的第二场实验中被试先进行有成本任务 2，且其 5 组分配方案的呈现顺序为 A，随后进行零成本任务 1，且 5 组分配方案的呈现顺序为 B。也就是说，计算机只在奇数场次实验 (1, 3, 5……) 的任务 1 和任务 2 生成 5 组分配方案的伪随机顺序，而接下来的偶数场实验 (2, 4, 6……) 将任务呈现顺序颠倒，但两个任务的 5 组分配方案的呈现顺序按照前一场实验的顺序 (AB) 进行。这一设计既控制了任务呈现的顺序效应，又确保了零成本设置和有成本设置的 5 组分配方案的呈现顺序在总体样本上是一致的。

为了避免分配结果重复及其顺序呈现的随机设计干扰 tDCS 效应，我们每一场实验 (session) 招募三名被试，且三名被试分别接受阴极、阳极和虚拟刺激。这样，上述设计既

减少了人为痕迹，又确保了分配方案及其呈现顺序在不同 tDCS 组中完全相同。

另外，对于重复呈现的结果，被试无需再次进行决策，其决策和收益将由计算机自动按照其在前序实验的决策处理。如下图所示。



重复呈现分配方案的计算机界面

我们在正文实验设计部分第 20 页的脚注 4 中加入了上述分析。

意见 3: 文章的结论提到，单因素方差分析结果表明第三方对社会规范的信念在不同 tDCS 刺激组间的差异不显著。这一结果排除了 tDCS 刺激通过改变第三方对社会规范信念和判断而改变其惩罚行为的可能性。但文中并没有明确说明是否在两个任务中，信念都不存在显著差异。

回应: 感谢审稿专家指出这一问题。我们在原文中提到“每观察一个小组，被试需要回答三个问题。①在看到独裁者的分配结果之前，被试需要回答他们对社会规范信念……”。此处存在表述错误。事实上，我们对社会规范信念的测度在实验伊始就进行测量且整个实验过程中只测量一次，而非每观察一个小组都需要测量一次。我们在原文汇报相应结果时提到“单因素方差分析结果表明第三方对社会规范信念在不同 tDCS 刺激组间的差异不显著 ($F_{2,86}=0.65, p=0.524$)。”从自由度的汇报结果可以佐证我们对社会规范信念只测量了一次。我们在正文实验设计部分的第 22 页进行了相应修改。

意见 4: 文章的主要结果为，第三方在零成本任务的情绪反应和惩罚显著受到 tDCS 设置的影响，且阴极刺激显著提升了第三方的惩罚值，这表明情绪机制对社会规范遵从有着重要影响。另外，第三方在零成本和有成本任务中的惩罚差异在不同 tDCS 设置之间也存在显著差异，这与社会规范还受到自利机制影响的观点相符。然而，文章结果并没有明确告知在零成

本和有成本两个任务中，情绪反应是否也在不同 tDCS 设置之间存在显著差异。

回应：感谢专家意见。我们在原文第 10 页（本文 21 页）提及：“任务 2 过程与任务 1 类似，不同之处在于任务 2 为有成本设置，被试输入的惩罚点数将影响他的收益，且任务 2 中被试无需汇报情绪反应。”也就是说，我们只在零成本设置中测度了被试的情绪反应，而没有在有成本设置中测度。原因在于，被试在有成本设置的任务 2 中汇报的情绪反应容易遭受“污染”，即其可能受到自利因素的干扰。况且，被试面对相同的分配方案重复汇报情绪反应时可能产生学习效应或需求效应等（此类效应对情绪汇报的影响可能比对惩罚决策的影响大得多，因为我们通常无法对自我汇报的情绪反应进行“价值诱导”的实验激励）。就本文研究目的而言，我们希望相对“干净”地比较不同 tDCS 组被试的情绪反应是否存在差异进而考察第三方惩罚的情绪机制，因此在实验设计时让被试只在零成本设置的任务 1 汇报情绪反应。我们在正文实验设计部分第 21 页的脚注 5 中加入了上述说明。再次感谢专家意见。

意见 5：文章的结论认为，当看到违反社会规范的行为时，第三方大脑中的第一系统（感性系统）会释放负性情绪，使其产生惩罚规范违反行为的冲动。进一步，如果对社会规范的遵从需要付出成本，第三方大脑中的第二系统（理性系统）会发挥作用，此时他对自利的理性思考会削弱其情绪的冲动作用。然而，虽然研究证实在有成本的条件下惩罚行为会得到减弱，但没有足够的证据说明就是理性系统在其中发挥了作用。

回应：感谢审稿专家意见。我们运用“双系统理论” (dual system)解释本文研究结论主要受启发于罗艺等（2013）和 Feng 等（2015），他们对最后通牒博弈的回应者行为进行了元分析和综述，认为回应者是否选择拒绝分配方案会先后调动双系统，是一个情绪和理性认知交互作用的复杂过程，这一观点与本文发现相契合，因此我们在原文的讨论和结论部分均阐述了这一观点。然而，鉴于学界对社会规范遵从的神经机制尚未达成共识，严谨起见这一观点更适合放在讨论部分作为理论解释，而不适合放在本文的研究结论部分。基于此，我们修改了结论部分关于“双系统”的表述。修改后的结论内容如下（对应 29 页）。

本文实验结果表明，DLPFC 是与第三方惩罚密切相关的重要脑区，该区域的活动会显著改变第三方的负性情绪反应和自利加工过程，进而影响第三方惩罚：当看到违反社会规范的行为时，第三方会受负性情绪驱使产生惩罚规范违反行为的冲动；进一步，如果对社会规范的遵从需要付出成本，第三方自利的理性思考会削弱其情绪的冲动作用，最终使得第三方惩罚即其对社会规范的遵从取决于其负性情绪和自利机制的权衡。

再次感谢审稿专家的宝贵意见。

参考文献:

- Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and human behavior*, 25(2), 63–87.
- Feng, C., Luo, Y. J., & Krueger, F. (2015). Neural signatures of fairness-related normative decision making in the ultimatum game: A coordinate-based meta-analysis. *Human brain mapping*, 36(2), 591–602.
- Harty, S., Robertson, I. H., Miniussi, C., Sheehy, O. C., Devine, C. A., McCreery, S., & O'Connell, R. G. (2014). Transcranial direct current stimulation over right dorsolateral prefrontal cortex enhances error awareness in older age. *The Journal of Neuroscience*, 34(10), 3646–3652.
- Li, J., Yin, X., Li, D., Liu, X., Wang, G., & Qu, L. (2017). Controlling the anchoring effect through transcranial direct current stimulation (tDCS) to the right dorsolateral prefrontal cortex. *Frontiers in Psychology*, 8, 1079.
- Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science*, 342(6157), 482–484.
- Sellaro, R., Nitsche, M. A., & Colzato, L. S. (2016). The stimulated social brain: effects of transcranial direct current stimulation on social cognition. *Annals of the New York Academy of Sciences*, 1369(1), 218–239.
- Spitzer, M., Fischbacher, U., Herrnberger, B., Grön, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, 56(1), 185–196.
- 罗艺, 封春亮, 古若雷, 吴婷婷, 罗跃嘉. (2013). 社会决策中的公平准则及其神经机制. *心理科学进展*, 21(2), 300–308.
- 王益文, 张振, 张蔚, 黄亮, 郭丰波, 原胜. (2014). 群体身份调节最后通牒博弈的公平关注. *心理学报*, 46(12), 1850–1859.

第二轮

审稿人 1 意见: 作者已按意见修改了论文的学术表达, 技术方法和数据分析与结论阐释部分, 我已没有进一步的修改意见。

审稿人 2 意见: 修改后的文章表述更为得当, 方法的介绍也更细致全面, 本人没有其它问题了。

编委复审意见: 经过两轮审稿, 修改稿达到发表的要求, 编委建议发表。

回应: 非常感谢编委老师对我们研究工作的肯定。

主编终审意见: 研究采用第三方惩罚的独裁者博弈, 设计零成本和有成本两个实验任务, 通过经颅直流电刺激 (tDCS) 改变右侧背外侧前额叶区域 (DLPFC) 的活跃水平, 探究右侧 DLPFC 对第三方惩罚的影响, 研究发现该区域的活动会显著改变第三方的负性情绪反应和自利加工过程, 进而影响第三方惩罚, 研究具有高水平的理论意义和学术价值。经过多轮修稿, 作者在语言表述、行文结构及逻辑框架方面进行了进一步修改和完善, 符合学报发表文章的要求, 建议发表。

回应: 非常感谢主编老师对我们研究质量的认可。