

《心理学报》审稿意见与作者回应

题目：社会互动视角下人际公平形成的脑机制
作者：张如倩; 刘洁琼; 李先春

第一轮

审稿人 1 意见：

论文采用 fNIRS 技术研究了人际互动视角下社会规范形成的脑机制。实验设计上最大的问题在于惩罚和不惩罚这两个条件是被试内设计，且分配者和接收者是规定的一对玩多轮。而这两个条件是相互影响的，难以分离的。惩罚条件下的分配者的分配金额会受到接受者的惩罚制约，但是在无惩罚条件下，分配者的分配方案如果不公平，他也有可能在下一轮或者接下来的惩罚条件下受到接受者的惩罚，因此他也是受惩罚制约。文中将分配金额差异显著的回合分为两个阶段，并称为是社会规范形成前和后，但是实际上这只是对分配达成了平衡。这两个阶段的划分在两个条件上的差异可能是条件的平衡点形成的时间点是不一样的，即不惩罚条件是否存在这样一个平衡点，而这个平衡点是在 10 个之前还是 10 个之后。也没有表述清楚惩罚条件下的 10 或 11 是否是第一次出现。

意见 1：建议给出 24 种阶段划分方法的在惩罚和不惩罚阶段的分配金额和惩罚额度或者惩罚频率的差异，可以更直观全面的看出这个平衡点的动态变化。

回应：感谢审稿专家的审阅。在修改稿中，我们对数据重新进行了分析，不再对惩罚条件下的试次进行阶段的划分以区分社会规范形成前与形成后，主要原因有以下 3 点：

(1) 原稿中社会规范是指公平规范，公平规范是业已存在的一种社会规范，因此原稿中我们将惩罚条件下的 25 个试次划分成 2 阶段，命名为“社会规范形成前”和“社会规范形成后”有不妥之处，以下我们将其改之为“违反阶段”和“遵守阶段”。原稿中我们对转折点的寻找是依据前后两个阶段的平均分配金额和惩罚频率是否存在显著的差异，但正如审稿人指出的那样，不同条件出现转折点的时间点可能不同，且这样的转折点可能存在多个。按照审稿人的建议，我们给出了 24 种阶段划分的具体情况，如图 1 所示，第 6~10 个点都满足转折点的要求。但是仅以“分配更多”和“惩罚更少”来定义“违反”和“遵守”并不充足，并且如审稿专家的解读，这些转折点可能仅是前后两阶段达成平衡的时间点而已。因此，通过上述方式定义公平规范的形成并不合适。

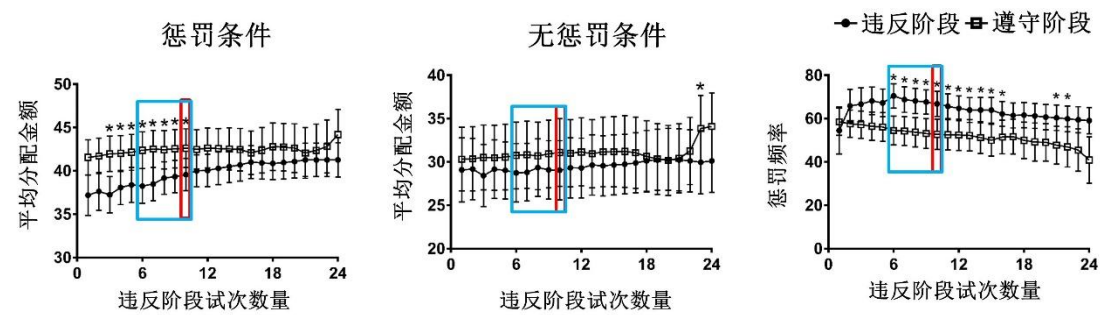
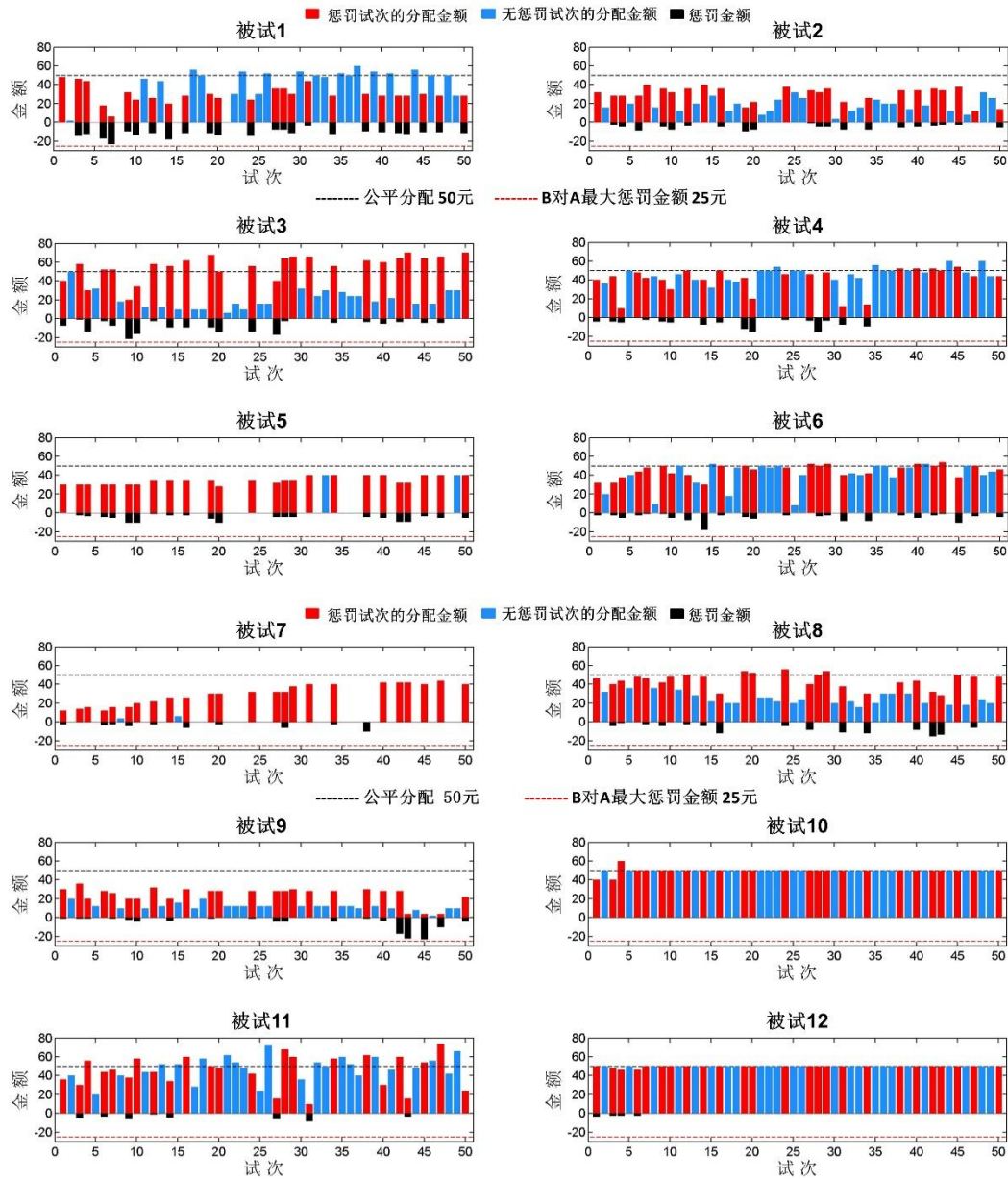


图 1 24 种划分情况

注：蓝框-满足转折点的第 6~10 个试次；红框-满足转折点的第 10 个试次。

(2) 为了更直观地了解每一对被试 50 个试次的分配和惩罚情况，我们将 22 对被试任务过程中的互动情况一一作图。从图 2 中可以看出，对于大部分的被试来说，惩罚条件下提议者对公平规范的遵从与违反存在反复交替出现的现象，违反与遵从两阶段并不明确，所以在修改稿中，我们不再对试次进行阶段的划分。



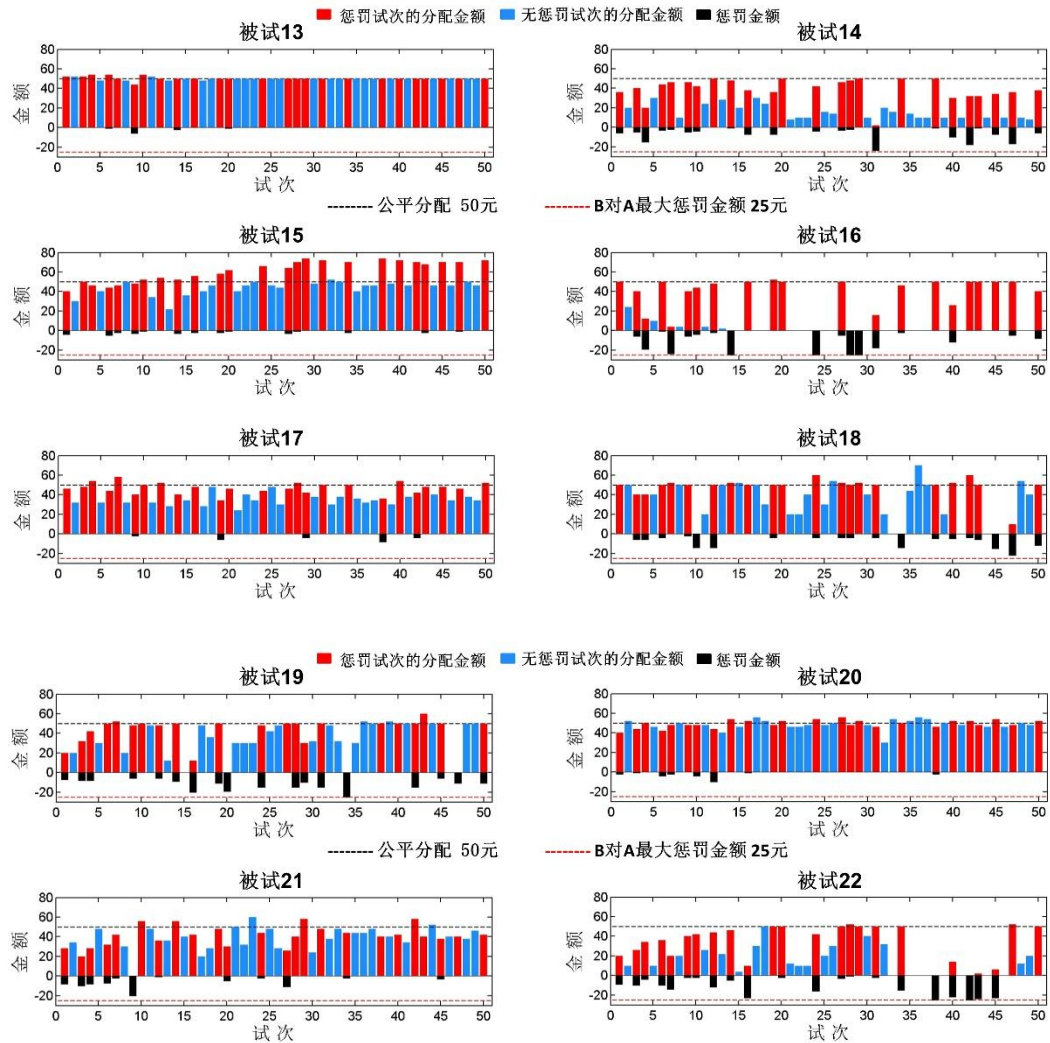


图2 22对被试50个试次的分配金额与惩罚金额

(3) 我们直接比较了惩罚条件和无惩罚条件下的分配金额,发现惩罚条件下提议者的分配金额 (41.39 ± 9.28) 高于无惩罚条件 (30.27 ± 17), $t(21)=3.91$, $p<0.001$, Cohen's $d=0.83$,更接近平均分配。已有研究显示,在最后通牒博弈任务中,提议者往往会遵守公平规范,其分配金额一般占总金额的 40%~50%(Ruff, Ugazio, & Fehr, 2013; Spitzer, Fischbacher, Herrnberger, Gron, & Fehr, 2007),而接受者也会接受高于 40%的分配金额(Wu et al., 2014)。另外,鉴于以往研究中,研究者对人际公平的操作定义为分配金额占总金额的 40%~50%(Baumgartner, Knoch, Hotz, Eisenegger, & Fehr, 2011; Halko, Hlushchuk, Hari, & Schürmann, 2009; Kirk, Downar, & Montague, 2011; Vieira et al., 2014; Wei, Zhao, & Zheng, 2013; Wu, Zang, Yuan, & Tian, 2015)。在本研究中,我们同样采取了上述操作定义,因此在修改稿中,我们直接比较惩罚与无惩罚条件下脑间活动同步性的差异,以揭示社会互动下人际公平形成的脑-脑机制。

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience*, 14(11), 1468-1474.

Halko, M.-L., Hlushchuk, Y., Hari, R., & Schürmann, M. (2009). Competing with peers: Mentalizing-related brain activity reflects what is at stake. *Neuroimage*, 46(2), 542-548.

Kirk, U., Downar, J., & Montague, P. R. (2011). Interoception drives increased rational decision-making in

- meditators playing the ultimatum game. *Frontiers in Neuroscience*, 5, 49.
- Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science*, 342(6157), 482-484.
- Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, 56(1), 185-196.
- Vieira, J. B., Almeida, P. R., Ferreira-Santos, F., Barbosa, F., Marques-Teixeira, J., & Marsh, A. A. (2014). Distinct neural activation patterns underlie economic decisions in high and low psychopathy scorers. *Social Cognitive and Affective Neuroscience*, 9(8), 1099-1107.
- Wei, Z., Zhao, Z., & Zheng, Y. (2013). Neural mechanisms underlying social conformity in an ultimatum game. *Frontiers in Human Neuroscience*, 7, 896.
- Wu, Y., Yu, H., Shen, B., Yu, R., Zhou, Z., Zhang, G., . . . Zhou, X. (2014). Neural basis of increased costly norm enforcement under adversity. *Social Cognitive and Affective Neuroscience*, 9(12), 1862-1871.
- Wu, Y., Zang, Y., Yuan, B., & Tian, X. (2015). Neural correlates of decision making after unfair treatment. *Frontiers in Human Neuroscience*, 9, 123.

另外，研究采用了被试内设计，参考了 Spitzer 等人(2007)和 Ruff 等人(2013)的研究。在上述两篇研究中，研究者同样将惩罚和无惩罚作为被试内变量，但与本研究不同的是，上述两个研究中被试每一个试次面对的互动对象都是随机配对的，即被试每一个试次的互动对象都不一样，这在一定程度上保证了试次之间的独立性。研究结果显示，在无惩罚条件下，提议者会将总金额的 10% 左右分配给接受者。但在本研究中，提议者和接受者是固定角色的，且进行多轮分配，这导致两个实验条件相互有影响，主要体现在相比 Spitzer 等人和 Ruff 等人的研究结果，无惩罚条件下提议者的分配金额（30.27±17 元）提高了 20 元，这可能是因为在无惩罚条件下，提议者的分配行为也受到了潜在的惩罚威胁的影响。但是行为学结果也表明，惩罚条件下提议者的分配金额（41.39±9.28）高于无惩罚条件（30.27±17）， $t(21)=3.91, p<0.001$, Cohen's $d=0.83$, 95%CI: 5.21~17.03，因此，相比无惩罚条件，惩罚条件下明确的惩罚对个体自利行为的约束力更大。相应的，fNIRS 的结果显示，惩罚条件下，提议者和接受者的脑间活动同步性也更强，揭示了惩罚下人际公平形成的脑-脑机制。

- Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science*, 342(6157), 482-484.
- Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, 56(1), 185-196.

意见 2：“自检报告”部分，研究亮点的两点，其实是一点。社会规范由于其本质是一种社会交互，因此需要测量交互中的双人的大脑，而超扫描技术可以解决这一问题。

回应：已将研究亮点合并为一点，“人际公平的形成是个体间互动博弈的结果，只观察单个个体的大脑活动，并不足以揭示由互动双方共同完成的社会认知活动的脑机制。因此，研究结合真实的‘人-人’互动与超扫描技术，通过分析脑-脑互动的变化规律，从群体脑水平上揭示人际公平形成的神经机制。”

意见 3：“摘要”部分，提到“惩罚条件下，提议者的分配金额更高，接受者的惩罚频率降低…显著增强，且从接受者到提议者更强的信息流向性”，更高、降低、增强、更强都应该需要是与某些指标进行对比得出的结论，请表述完整。

回应：“摘要”部分已经根据最新结果进行更新，并根据审稿专家的建议修正表述，使其更加完整，具体如下：“最后通牒博弈任务被广泛用以探究公平行为，以往研究大多集中于对

博弈中某一方决策行为和神经机制的探讨，但是人际公平可能是互动双方重复博弈的结果。因此只考察单个大脑活动，并不足以揭示由互动双方共同完成的社会认知活动的脑机制。因此，本研究结合修改版的最后通牒博弈任务和基于 fNIRS 的超扫描技术，从群体脑水平上考察人际公平形成的脑机制。行为结果显示，相比无惩罚条件，惩罚下提议者的分配金额更高，且惩罚力度越强，分配越趋近公平分配。fNIRS 的结果显示，惩罚下右侧背外侧前额叶皮层、顶下小叶和颞-顶联合区的脑间活动同步性显著强于无惩罚条件，而且两条件的分配金额差异越大，右侧顶下小叶的脑间活动同步性差异也越大。综上，脑间活动同步性可以作为惩罚下人际公平形成的客观脑指标，研究为探讨人际公平的内在机制提供了新的视角。”

意见 4：“引言”部分，从社会规范的重要性、惩罚对于社会规范形成的重要性、提议者的大脑机制（社会规范的遵从）和接收者的大脑机制（社会规范的执行）来提出研究问题，但是忽略了公平这一具体的社会规范的阐述。文中在超扫描技术的普遍应用里提到了基于核磁的相关研究，并未说明本研究为什么采用的是基于近红外的超扫描技术。建议参考阅读相关国内综述文章。

回应：感谢审稿专家的建议，已经在“引言”部分增添“公平规范”的阐述，具体如下：“从古至今，‘不患寡而患不均’的思想始终促使着人们不断追求公平（吴燕，周晓林，2012）。公平作为一种基本的社会准则，在人类生活中起着重要的作用，它不仅保障了个体的基本权利，同时也维护了社会的稳定发展。”

另外我们也增添了选用基于近红外超扫描技术的理由，具体如下：“相比其他的脑功能成像技术，功能近红外光谱技术（functional near-infrared spectroscopy, fNIRS）在成本、对头动的容忍性以及研究的生态效度等方面具有优势(Scholkmann, Holper, Wolf, & Wolf, 2013)。”

意见 5：“方法”部分，“基于对两个时间序列的 WTC 分析”该分析过程是什么？是主观观察吗？任务的时长在分配阶段和反馈阶段是自定的，导致每一轮的 trial 时长不一样长，后续的分析的选取频率段是怎么保证任务相关的频率段呢？“FDR 校正”的方法采取的是哪一种。

回应：小波变换相干（wavelet transform coherence, WTC）的分析过程及任务相关频段的选择方法已增添在“2.5.2 fNIRS 数据”部分。文中采用 Benjamini (1995) 提出的 FDR（False Discovery Rate）校正方法对 p 值进行校正，已在“2.5.2 fNIRS 数据”部分增添了说明。

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289-300.

意见 6：“结果”部分，如何解释行为结果里社会规范形成后的惩罚频率仍高达 50%？“而这一结果没有在无惩罚条件下有所发现， $t(42)=1.81$, $p=0.08$, Cohen's $d=0.54$, 95%CI: -0.002~0.03。”如何解释边缘显著，且效应值也较大？有 fnirs 的激活结果吗？惩罚和无惩罚条件的脑区激活模式上的差异？

回应：结果已更新，上述结果已经删除。按照审稿专家的建议，我们对单脑进行了分析。在 Matlab (2014a) 的操作环境下，使用 NIRS_SPM 对数据进行预处理。采用血液动力响应函数（Hemodynamic Response Function, HRF）和 Wavelet-MDL 进行低通滤波和高通滤波，从而消除仪器产生的噪音和呼吸心跳等引起的生理噪音(Ye, Tak, Jang, Jung, & Jang, 2009)。接着，通过一般线性模型(General Linear Model, GLM)计算基线（任务前 15s），提议者分配阶段和接受者反馈阶段的脑激活（beta 值）。分别将不同条件下提议者和接受者在分配阶段和

反馈阶段的 beta 值与基线进行配对样本 t 检验。

结果显示，在提议者分配阶段，无论在惩罚条件下还是无惩罚条件下，提议者和接受者都没有显著激活的通道（如图 3 所示），惩罚条件：提议者 $ts < 2.5$, $ps > 0.28$, FDR 校正；接受者 $ts < 2.14$, $ps > 0.42$, FDR 校正；无惩罚条件：提议者 $ts < 2.03$, $ps > 0.16$, FDR 校正；接受者 $ts < 1.99$, $ps > 0.99$, FDR 校正。

在接受者反馈阶段，在惩罚条件下，提议者没有显著激活的通道， $ts < 2.86$, $ps > 0.14$, FDR 校正；接受者通道 8 和通道 17 显著激活， $t_8 = 3.12$, $p_8 = 0.04$, $t_{17} = 3.77$, $p_{17} = 0.02$, FDR 校正。无惩罚条件下，提议者和接受者均无显著激活的通道，提议者 $ts < 0.92$, $ps > 0.04$, FDR 校正；接受者 $ts < 1.94$, $ps > 0.86$, FDR 校正。对接受者通道 8 和通道 17 分别进行配对样本 t 检验以考察不同条件下的脑激活是否有差异，结果显示，通道 8 和 17 的激活程度在条件间均无显著的差异， $t_8 = 0.94$, $p_8 = 0.36$; $t_{17} = 1.90$, $p_{17} = 0.07$ 。

因为在本研究中我们主要关注的是人际公平形成的脑-脑机制，因此，没有对提议者和接受者不同条件下不同阶段的脑激活进行更为深入的分析。

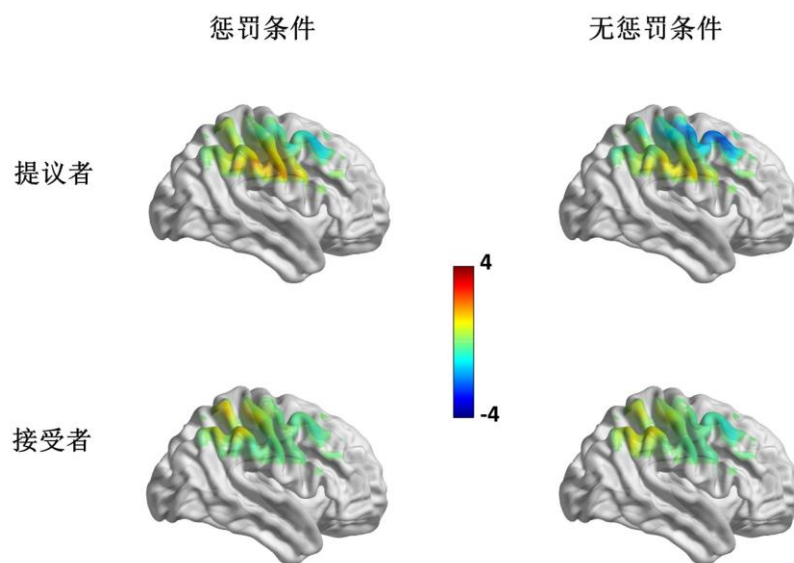


图 3 不同条件下互动双方在提议者分配阶段的脑激活

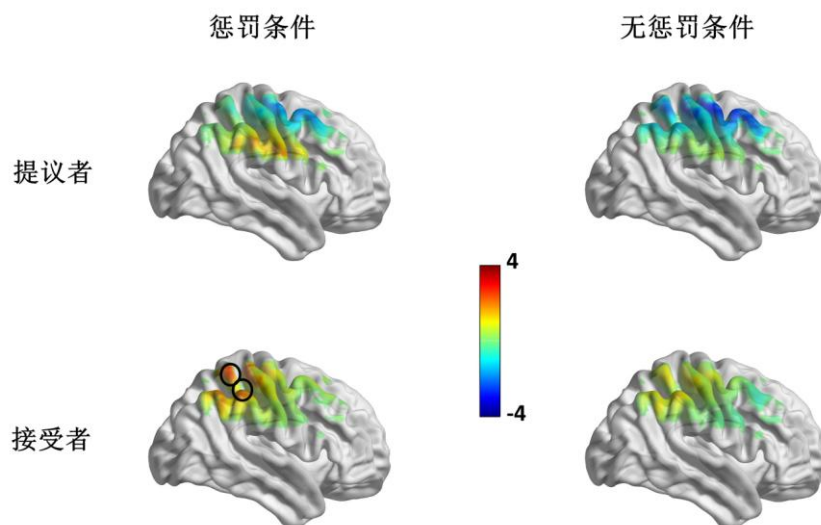


图 4 不同条件下互动双方在接受者反馈阶段的脑激活

意见 7: 文章多处用词不清晰, 指代不清, 让读者难以理解, 如什么叫做“结构良好的互动式经济游戏”, 结构良好的条件是什么? 又比如这种重复的博弈是一种“二合一”的过程难以理解重复和二合一, 重复是指多回合制吗? 二合一是指什么? “该技术已在社会认知神经科学的多个领域有所应用”之后提到的所以文献都是社会经济决策领域, 多个体现在哪里? 表述不准确。又比如“也暗示着社会规范形成下互动双方神经活动的相似性。”这句话的意思是说两种单侧化暗示这双脑大脑活动的相似性? 理由不够充分, 双方的脑区激活都在单侧并不能暗示大脑活动的相似性。“另外, 基于以往研究, 惩罚条件下, 随着社会互动的持续, A 对 B 的分配金额会逐步接近公平分配”, 请引用相关文献。

回应: 对上述没有表述清晰的内容已经进行了修正, 对表述不合理的内容进行了删除, 也对公平规范遵从下个体间在背外侧前额叶皮层存在脑间活动同步性的假设提供了更为充足的假设依据, 具体请见“引言的第 3、4、5 段”。

审稿人 2 意见: 本研究采用近红外脑功能成像技术考察人际互动情境下社会规范的脑机制, 从研究手法上是一个创新。研究的问题亦具有较深刻的理论意义和潜在的应用价值。

意见 1: 对三个问题的讨论可能是至关重要的: 一、本研究对社会规范的理论定义是什么? 操作性定义又是什么? 如果用“分配给他人更多的金钱”或者“遭受更少的惩罚”作为社会规范的形成依据是否合适? 是否有陷入循环论证的倾向? 此外, 用提议-接受模式(最后通牒范式)研究社会规范这样的大问题是否过于简单? 该范式只是一个简单的多轮博弈任务, 在博弈任务中(甚至有研究表明, 在石头剪子布游戏中)脑间活动的同步性就已经存在了。二、正如作者所述, 社会规范有形成和遵守两个不同的过程, “惩罚”作为该研究中唯一操作的变量, 起作用的是形成过程呢, 还是遵守过程? 如果作用的是形成过程, 那么在没有规范的前提下, 惩罚的主体(对象)是什么呢? 如果作用的是遵守过程, 那么将研究结果分为“规范形成前”和“规范形成后”就没有意义了(因为规范早就有了)。此外, 按照作者的观点, 既然规范是宏观的, 渐次形成的, 那么在实验中, 连续不断出现的“惩罚”或“无惩罚”试次是否会干扰或打破社会规范的形成过程, 使得这个研究的生态效度降低? 三、研究没有假设, 且只讨论了 rTPJ, 没有比较脑区, 也是本研究的不足之处之一。

回应: 感谢审稿专家的审阅。根据两位审稿专家的意见, 我们对原来的研究结果重新进行了思考, 发现对数据结果的解读有不妥之处, 因而对数据进行了重新分析, 主要原因有以下 3 点:

(1) 原稿中社会规范是指公平规范, 公平规范是业已存在的一种社会规范, 因此原稿中我们将惩罚条件下的 25 个试次划分成 2 阶段, 命名为“社会规范形成前”和“社会规范形成后”有不妥之处, 以下我们将其改之为“违反阶段”和“遵守阶段”。原稿中我们对转折点的寻找是依据前后两个阶段的平均分配金额和惩罚频率是否存在显著的差异, 但正如审稿专家指出的那样, 不同条件出现转折点的时间点可能不同, 且这样的转折点可能存在多个。按照审稿专家的建议, 我们给出了 24 种阶段划分的具体情况, 如图 1 所示, 第 6~10 个点都满足转折点的要求。但是仅以“分配更多”和“惩罚更少”来定义“违反”和“遵守”并不充足, 并且如审稿专家的解读, 这些转折点可能仅是前后两阶段达成平衡的时间点而已。因此, 通过上述方式定义公平规范的形成并不合适。

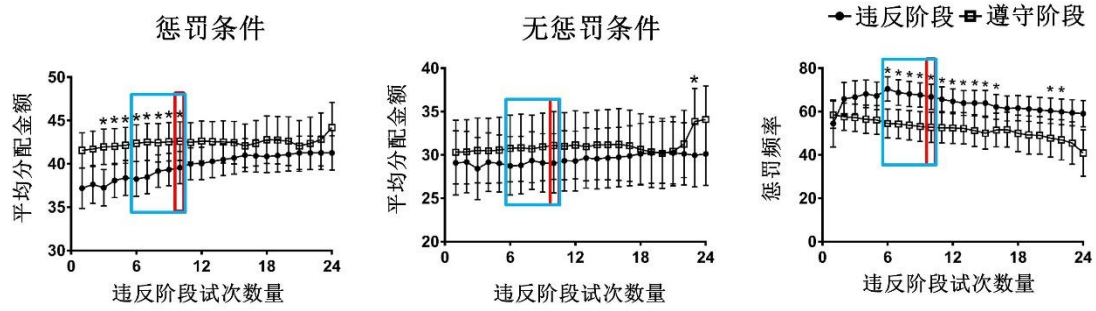
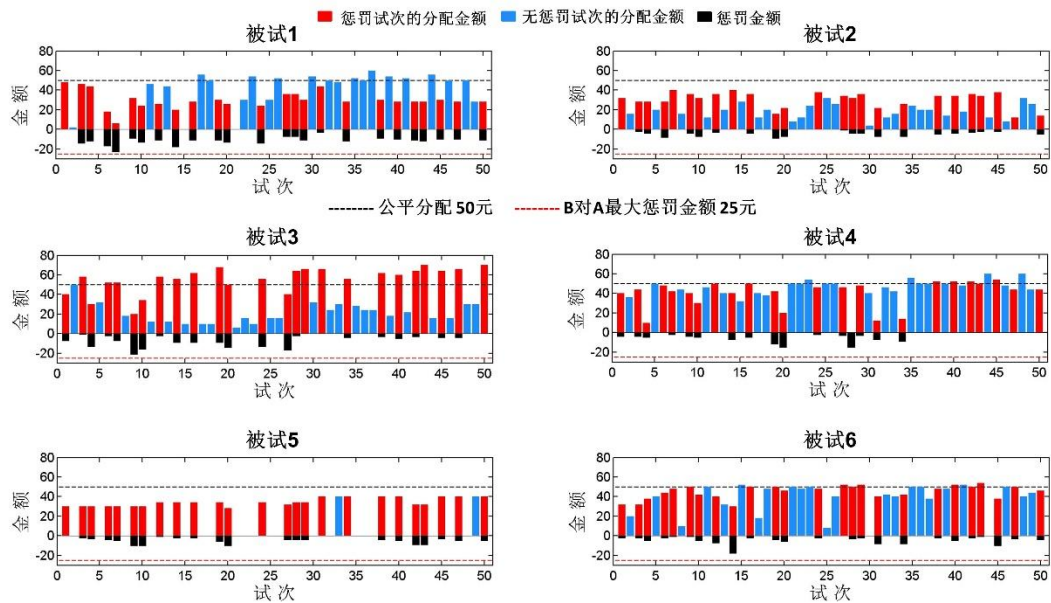


图1 24种划分情况

注：蓝框-满足转折点的第6~10个试次；红框-满足转折点的第10个试次。

(2) 为了更直观地了解每一对被试50个试次的分配和惩罚情况，我们将22对被试任务过程中的互动情况一一作图。从图2中可以看出，对于大部分的被试来说，惩罚条件下提议者对公平规范的遵从与违反存在反复交替出现的现象，违反与遵从两阶段并不明确，所以在修改稿中，我们不再对试次进行阶段的划分。



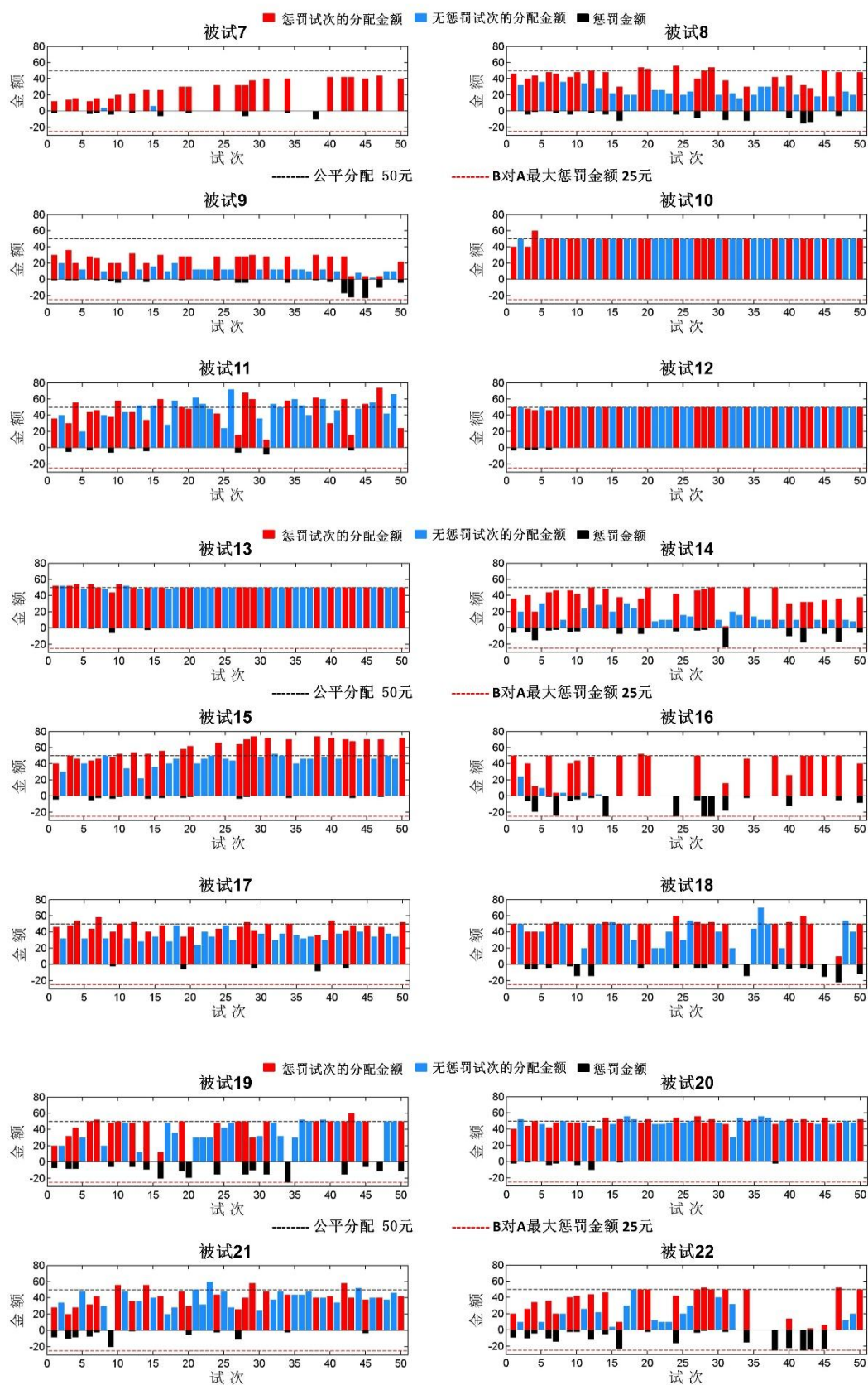


图2 22对被试50个试次的分配金额与惩罚金额

(3) 我们直接比较了惩罚条件和无惩罚条件下的分配金额，发现惩罚条件下提议者的分配金额 (41.39 ± 9.28) 高于无惩罚条件 (30.27 ± 17)， $t(21)=3.91$, $p<0.001$, Cohen's $d=0.83$,

更接近平均分配。已有研究显示,在最后通牒博弈任务中,提议者往往会遵守公平规范,其分配金额一般占总金额的 40%~50%(Ruff, Ugazio, & Fehr, 2013; Spitzer, Fischbacher, Herrnberger, Gron, & Fehr, 2007),而接受者也会接受高于 40%的分配金额(Wu et al., 2014)。另外,鉴于以往研究中,研究者对人际公平的操作定义为分配金额占总金额的 40%~50%(Baumgartner, Knoch, Hotz, Eisenegger, & Fehr, 2011; Halko, Hlushchuk, Hari, & Schürmann, 2009; Kirk, Downar, & Montague, 2011; Vieira et al., 2014; Wei, Zhao, & Zheng, 2013; Wu, Zang, Yuan, & Tian, 2015)。在本研究中,我们同样采取了上述操作定义,因此在修改稿中,我们直接比较惩罚与无惩罚条件下脑间活动同步性的差异,以揭示社会互动下人际公平形成的脑-脑机制。

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience*, 14(11), 1468-1474.

Halko, M.-L., Hlushchuk, Y., Hari, R., & Schürmann, M. (2009). Competing with peers: Mentalizing-related brain activity reflects what is at stake. *Neuroimage*, 46(2), 542-548.

Kirk, U., Downar, J., & Montague, P. R. (2011). Interoception drives increased rational decision-making in meditators playing the ultimatum game. *Frontiers in Neuroscience*, 5, 49.

Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science*, 342(6157), 482-484.

Spitzer, M., Fischbacher, U., Herrnberger, B., Gron, G., & Fehr, E. (2007). The neural signature of social norm compliance. *Neuron*, 56(1), 185-196.

Vieira, J. B., Almeida, P. R., Ferreira-Santos, F., Barbosa, F., Marques-Teixeira, J., & Marsh, A. A. (2014). Distinct neural activation patterns underlie economic decisions in high and low psychopathy scorers. *Social Cognitive and Affective Neuroscience*, 9(8), 1099-1107.

Wei, Z., Zhao, Z., & Zheng, Y. (2013). Neural mechanisms underlying social conformity in an ultimatum game. *Frontiers in Human Neuroscience*, 7, 896.

Wu, Y., Yu, H., Shen, B., Yu, R., Zhou, Z., Zhang, G., . . . Zhou, X. (2014). Neural basis of increased costly norm enforcement under adversity. *Social Cognitive and Affective Neuroscience*, 9(12), 1862-1871.

Wu, Y., Zang, Y., Yuan, B., & Tian, X. (2015). Neural correlates of decision making after unfair treatment. *Frontiers in Human Neuroscience*, 9, 123.

意见 2: p.1. (以下页数均是正文页数, 不包括自检报告) “综上, 惩罚能够促进社会规范的形成, 增加互惠合作, 并伴随着更强的脑间活动同步性”, 这一说法不妥, 建议改成 “通过脑间活动同步性的结果可以推测, 惩罚能够促进社会规范的形成, 增加互惠合作。”

回应: 已根据最新结果对摘要进行了更新, 并按照审稿专家的建议修正了措辞, 具体如下: “最后通牒博弈任务被广泛用以探究公平行为, 以往研究大多集中于对博弈中某一方决策行为和神经机制的探讨, 但是人际公平可能是互动双方重复博弈的结果。因此只考察单个大脑活动, 并不足以揭示由互动双方共同完成的社会认知活动的脑机制。因此, 本研究结合修改版的最后通牒博弈任务和基于 fNIRS 的超扫描技术, 从群体脑水平上考察人际公平形成的脑机制。行为结果显示, 相比无惩罚条件, 惩罚下提议者的分配金额更高, 且惩罚力度越强, 分配越趋近公平分配。fNIRS 的结果显示, 惩罚下右侧背外侧前额叶皮层、顶下小叶和颞-顶联合区的脑间活动同步性显著强于无惩罚条件, 而且两条件的分配金额差异越大, 右侧顶下小叶的脑间活动同步性差异也越大。综上, 脑间活动同步性可以作为惩罚下人际公平形成的客观脑指标, 研究为探讨人际公平的内在机制提供了新的视角。”

意见 3: p.1. “社会规范是基于共享观念, 要求个体在特定情境中遵守的行为标准。”如果按照这一定义, 本研究采用的范式还够不上社会规范, 而只能称为一种简单的博弈任务。

回应: 标题及正文中的“社会规范”实际上指的是“公平规范”。实验心理学和认知神经科学对于公平问题的探讨主要采用行为经济学的博弈范式, 其中一个应用较为广泛的实验范式就是最后通牒博弈任务(Feng, Luo, & Krueger, 2015; Gabay, Radua, Kempton, & Mehta, 2014), 因此本研究通过采用 Spitzer 等人(2007)修改的最后通牒博弈范式从社会互动视角出发探究人际公平形成的脑-脑机制。另外, 已将标题更改为“**社会互动视角下人际公平形成的脑机制**”, 避免研究问题过大。

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.

Gabay, A. S., Radua, J., Kempton, M. J., & Mehta, M. A. (2014). The Ultimatum Game and the brain: a meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews, 47*, 549-558.

意见 4: p.1. “大量研究致力于探讨社会规范形成和维持的影响因素, 较为一致的研究结果表明惩罚是有效制裁违反行为的举措之一。”惩罚是制裁违反社会规范的有效方法, 但不能说惩罚是社会规范形成的途径。

回应: 已对不当表述进行了修正。

意见 5: p.1. “Spitzer 等人曾(2007)采用最后通牒博弈范式, 比较在不同惩罚条件下提议者的分配金额...”此处年代应标记在“等人”之后。

回应: 感谢审稿专家的细心审阅, 已对该处错误进行了修正。

意见 6: p.1. “结果显示, 面对不公平分配时, 接受者会采取惩罚措施来促进接受者对公平规范的遵从。”这个结果语焉不详, 接收者对“公平规范遵从”的操作定义是什么, 如何体现在分配金额上? 需要具体阐述。

回应: 已对上述结果进行具体阐述, 修改为“**结果显示, 无惩罚下提议者的分配金额仅为总金额的 10%, 而惩罚下分配金额达到总金额的 40%, 更接近平均分配 (50%)。**”

意见 7: p.1. “也有研究者采用公共物品博弈范式探究惩罚对于国际合作的影响(Fehr & Gächter, 2000, 2002), 同样发现了惩罚的效用, 惩罚大大减少了‘搭便车’的现象, 提高了群体内的合作水平。”首先, 这个结果语焉不详, 什么是“搭便车”现象没有说明; 其次“提高群体内的合作水平”似乎和社会规范并无关系; 第三, “公共物品博弈”如果是一个经典的范式, 应稍作解释并加注英文。

回应: 对引言内容进行了调整, 着重于以最后通牒博弈为研究范式的研究结果, 因此, 删除了有关公共物品博弈相关的研究结果, 请见“**引言第 2 段**”。

意见 8: p.1. “基于社会规范的互惠合作需要两个要素, 社会规范的执行 (norm enforcement) 与遵从 (norm compliance) ”“执行”和“遵从”不是一回事吗? 它们的内涵有何区别?

回应: 执行与遵从在本文中的含义是一致的, 鉴于提议者与接受者的角色不同, 分别以“执行公平规范”来表达提议者对不公平行为的惩罚, 以“遵从公平规范”来表达接受者进行公平分配。两个动词表达的都是对人际公平的维护, 只是动作的发出者不同, 因而以不同的动词加以表述。为避免审稿专家和读者在阅读时产生困惑, 我们已统一使用“遵从”这个词。

意见 9: p.2. “提议者做出更为公平的分配时, 双侧背外侧前额叶皮层 (dlPFC)、腹外侧前额

叶皮层(vIPFC)、眶额皮层(OFC)、以及双侧尾状核(caudate nucleus)的激活程度增强(Spitzer et al., 2007)”在这里,“更为公平”是如何界定的,因为作者没有提及对比条件,怎么区分这些脑区是因为“惩罚”而激活的,还是因为“公平”而激活的?

回应:已对该研究结果的表述进行了修改,修改为“已有研究显示,相比无惩罚条件,惩罚条件下提议者的分配金额更高,分配过程中双侧背外侧前额叶皮层(dIPFC)、腹外侧前额叶皮层(vIPFC)、眶额皮层(OFC)以及双侧尾状核(caudate nucleus)的激活程度增强(Spitzer et al., 2007),且两条件的分配金额差异越大,右侧背外侧前额叶皮层、左侧眶额皮层、右侧尾状核激活程度的差异也越大。这表明遵守公平规范可以激活与认知控制(Miller & Cohen, 2001)、优势反应(自私自利)抑制(Aron, Robbins, & Poldrack, 2004)、惩罚威胁评估(O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001)以及不确定奖赏与惩罚刺激加工相关的脑区(Delgado, Locke, Stenger, & Fiez, 2003)。”

意见 10: p.2. “提议者的公平决策行为与右侧背外侧前额叶的激活程度紧密相关,增强该脑区的激活程度可显著提升提议者的分配金额,促进社会规范的遵从(Ruff, Ugazio, & Fehr, 2013)。”什么是“公平决策行为”?是提议者分配更多的钱给对方,还是其他?应对此有一个操作性的定义。“增强该脑区的激活程度”的主体是谁?是被试在决策行为中该脑区“自发地”激活,还是通过什么技术手段(如经颅磁刺激技术)激发?如果是前者,那么这句话是无意义的重复论证(“增加分配金额增加了这一脑区的激活,这一脑区激活时提升了提议者的分配金额”)。

回应:感谢审稿专家的细心审阅,对于研究结果的表达确实不够清晰,我们已对其进行了修正,修改为“同时,研究者还发现公平规范遵从的神经通路具有偏侧化的特点,使用经颅直流电刺激(tDCS)增强右侧 dIPFC 的激活程度可显著提升提议者的分配金额,使其更接近平均分配;而抑制该脑区则会降低分配方案的公平性(Ruff, Ugazio, & Fehr, 2013)。”

意见 11: p.2. 关于右侧背外侧前额叶的作用,刺激该区域,提议者分配更多的钱,而抑制该区域,接收者倾向于接受不公平的决定,一方是刺激,一方是抑制,这能“暗示社会规范形成下互动双方的神经活动有相似性吗”?

回应:对第3段中有关 tDCS 研究结果进行补充,增添了抑制右侧背外侧前额叶皮层对提议者公平分配行为的影响,具体修改如下:“同时,研究者还发现公平规范遵从的神经通路具有偏侧化的特点,使用经颅直流电刺激(tDCS)增强右侧 dIPFC 的激活程度可显著提升提议者的分配金额,使其更接近平均分配;而抑制该脑区则会降低分配方案的公平性(Ruff, Ugazio, & Fehr, 2013)。”同时,我们也修改了抑制右侧背外侧前额叶皮层对接受者惩罚行为的影响,具体修改如下:“研究者们同样采用了无创的干预技术探究了接受者在遵从公平规范中 dIPFC 功能的偏侧化,发现当右侧 dIPFC 被抑制时,接受者面对不公平分配时惩罚行为显著减少(Baumgartner, Knoch, Hotz, Eisenegger, & Fehr, 2011; Knoch et al., 2008; Knoch et al., 2006)。”(注:引用的三个研究均抑制了右侧背外侧前额叶皮层的激活,没有探讨增强该脑区的激活对接受者惩罚不公平行为的影响)。对于提议者而言,接近或者进行平均分配是遵从公平规范的表现;对于接受者而言,惩罚不公平行为是遵从公平规范的表现。从以往研究结果来看,无论对提议者还是接受者,抑制右侧背外侧前额叶皮层均会对其公平规范的遵从产生负性影响。因此,我们认为上述研究结果一定程度上暗示了公平规范遵从下提议者和接受者 dIPFC 神经活动具有相似性。

意见 12: p.2. “因而,这种重复的博弈是一种‘二合一’的过程”此处的“二合一”是什么意思?改为“互为因果的决策过程”似乎更为明确。

回应：感谢审稿专家的建议，在原稿中我们使用“二合一”是为了说明最后通牒博弈中提议者和接受者不是两个相互独立毫无关联的个体，而是一个相互影响的整体。在任务中，他们每一轮试次的决策都会将对方的行为和意图纳入考虑，个体的行为结果既是一种反馈，也是其互动对象行为的前提。我们采用审稿专家的建议，修改为“**重复的博弈是一种互为因果的过程，个体的行为结果既是一种反馈，也是其互动对象行为的前提**(Shaw et al., 2018)。”

意见 13: p.3. 在这里作者提及了多种游戏（范式），如最后通牒游戏、囚徒困境游戏、面对面金钱分配游戏、重复通牒博弈任务，而且发现，这些游戏范式下激活的脑区有所差异，如在囚徒困境中，同步性的是额-顶区域和颞-顶联合区，在重复通牒博弈中，是脑岛前部和扣带回中部等等，作者有必要对这些游戏所考查的心理层面的差异以及激活脑区的本质含义加以梳理。

回应：研究着眼于公平，上述提及的有些研究关注的是合作，因此对于有些研究结果进行删除。另外，人际公平的形成是互动双方重复博弈的结果(Spitzer et al., 2007)，而这一过程中涉及对他人心理状态进行解释和预测，即心理理论。因此，引言的第 5 段着重于阐述与“心理理论”相关脑区的 INS 的含义。

意见 14: p.3. “共招募大学生 44 名 ($M=21.3$ 岁, $SD=1.6$ 岁)，其中男生 26 人，女生 18 人，共形成 22 对同性别被试对。”在这里，需要明确招募总体，且认知神经科学实验一般要求有更多的人口统计学信息。

回应：研究招募的总体为华东师范大学在读本科生，按照审稿专家的建议已在“2.1 被试”补充了其他的人口统计学信息，修改为“**共招募在读本科生 44 名 ($M=21.3$ 岁, $SD=1.6$ 岁)，其中男生 26 人，女生 18 人，共形成 22 对同性别被试对。所有被试均为右利手，无脑部疾病或者精神疾病，视力或者矫正视力均正常。实验前，所有被试签署了知情同意。实验后，根据被试在实验过程中的实际表现给与相应的被试费。**”

意见 15: p.3. 本研究采用的是组内设计，5:1 和 0:0 两名被试都能看到吗？两种条件的交替出现而导致的交互影响效应如何在实验中剔除？能否极端地认为，因为 0:0 条件的存在，被试很难形成真正的社会规范？

回应：5:1 和 0:0 两名被试都可以看见，任务过程中电脑屏幕上呈现给互动双方的内容是完全一致的，这一点已经补充在“2.3 实验流程”部分的第 1 段，具体如下“**任务中 A 和 B 电脑屏幕上所呈现的内容一致，即 B 可以实时看见 A 的分配情况，A 也能实时获取 B 的反馈情况。**”

意见 16: p.4. 实验中，被试 B 是否知道被试 A 有多少钱，如果 B 惩罚了 A 提出的分配方案，如何处理？B 知道不知道分配的金额……类似的信息，应该在方法部分详细阐述。

回应：已根据审稿专家的建议对“2.3 实验流程”部分的第 1 段进行了补充，具体如下“**任务中 A 和 B 电脑屏幕上所呈现的内容一致，即 B 可以实时看见 A 的分配情况，A 也能实时获取 B 的反馈情况。**”

意见 17: p.4. 实验的很多参数，如 100 元，5:1，0:0 设置的依据是什么？这些参数水平的控制可能极大地影响实验的结果。

回应：研究选用分配金额 100 元，惩罚比例 5:1，一是参考了 Spitzer 等人(2007)和 Ruff 等人(2013)的研究；二是考虑到较大的分配金额更能诱发出被试真实的行为。

意见 18: p.4. 实验中, 分配者与接收者是否互换地位?

回应: 实验中分配者与接受者的角色是固定的, 这一点已补充在“2.3 实验流程”部分的第 1 段, 具体如下“**实验开始前, 被试双方通过抽签的方式来决定任务角色 (提议者 A 和接受者 B), 任务中角色固定。**”

意见 19: p.4. 图一的绘制看不清, 建议用更好的软件重新绘制, 提高图片质量。

回应: 已按照审稿专家的建议重新绘制流程图。

意见 20: p.5. “如果以某一试次为分界点, 相比第 2 阶段, 第 1 阶段的分配金额较低且惩罚比率较大, 那么该试次为社会规范形成的时间点。”怎么叫“低”, 怎么叫“大”? 应该有更加明确的标准, 按照这个指标, 满足要求的时间点应该有多个, 哪一个表明是“唯一的”形成点?

回应: 新的分析结果不再将试次分为两阶段, 请审稿专家审阅更新的结果。

意见 21: p.5. “使用单样本 t 检验 (FDR 校正) 找到惩罚和无惩罚条件下两阶段中脑间活动同步性显著的通道。”研究应有先验的假设, 如果用这种方式去“发现”, 那么即使数据完全随机, 显著的通道也总会有, 但并不是真实的。这一点涉及到后期数据的分析, 需要给予足够关注。

回应: 感谢审稿专家的建议, 引言部分总结了最后通牒博弈任务中提议者和接受者遵从公平规范的神经机制, 其中右侧背外侧前额叶皮层与公平行为紧密相关。另外, 引言中还总结了采用超扫描技术探究人际互动脑机制的研究结果, 发现“心理理论”相关的脑间活动同步性反映了人际互动水平和共享意图。因此, 研究主要关注右侧背外侧前额叶皮层和“心理理论”相关的脑区, 并根据前人研究结果, 做出了如下假设: “**基于以往研究结果, 惩罚可以促使个体在经济决策中遵从公平规范, 我们预期 (1) 惩罚下分配金额高于无惩罚条件, 且更趋近平均分配。与此同时, 遵从公平规范需要个体抑制优势反应 (自利行为), 且该过程中个体间的交互性更强。因此, 我们预期 (2) 相比无惩罚条件, 惩罚下提议者和接受者右侧背外侧前额叶以及 ‘心理理论’ 相关脑区的 INS 更强。**”

意见 22: p.6. “结果显示, 在惩罚条件下, 前 10 个试次的平均分配金额显著少于后 15 个试次, $t(21)=-2.14$, $p=0.04$, Cohen's $d=0.46$, 95%CI:-6.02~-0.08 (如图 2B 所示)。相应地, 惩罚频率在这两阶段中也存在显著的差异, $t(21)=3.00$, $p=0.007$, Cohen's $d=0.64$, 95%CI:4.31~23.87 (如图 2C 所示)。”这样的“点”应该不是唯一的, 为什么作者要定在 10 和 11 呢? 9 和 10 呢? 8 和 9 呢? 最好也给予报告。

回应: 新的分析结果不再将试次分为两阶段, 请审稿专家审阅更新的结果。

意见 23: p.6. 按照作者的定义, 即使在形成社会规范以后, 也有 50% 以上的惩罚 (见图 2), 如何理解? 作者可能需要进一步提供每一轮的博弈信息。

回应: 新的分析结果不再将试次分为两阶段, 请审稿专家审阅更新的结果。

意见 24: p.6. 应该给出每个通道的含义及对应的脑区, 而作者只大略给出 8, 9 和 18, 其他的没有。

回应: 根据审稿专家的建议, 已在“2.4 数据收集”部分给出了 22 个通道的 MNI 坐标和对应的脑区。

意见 25: p.6.8 和 9 是同一个脑区（颞顶联合区），但结果有不同，如何解释？
回应：新的结果已经更新，该结果已经删除。

第二轮

审稿人 1 意见：

意见 1：总的意见：本研究采用近红外脑功能成像技术考察人际互动情境下社会规范的脑机制，具有一定的创新性。且此次大修后的论文，已对上次本人提出的建议做了相应的调整和补充。此稿现在，无论是在组织上的逻辑性，还是在语言表述上的严谨性都有了很大的进步。建议发表。下面还有一点疑问：对于行为结果的分析，所采用的这个计算公式，需解释一下为什么是这样计算的。

回应：感谢审稿专家对于修改稿的评阅以及对修改内容的认可。计算惩罚条件下接受者 B 对提议者 A 不公平分配（分配金额小于 50 元）的惩罚力度，我们采用了如下公式：

$$\frac{5 \times B \text{ 对 A 的惩罚金额}}{50 \times \text{不公平试次数量} - A \text{ 对 B 不公平分配的总金额}}$$

- (1) **5**：惩罚比例，接受者每拿出 1 元，就相应地减少提议者 5 元。
- (2) **B 对 A 的惩罚金额**：不公平分配试次下（即分配金额低于 50 元）接受者惩罚金额的总数。任务过程中，每一轮试次开始前，被试两人都先获得 25 元，这保证了惩罚条件下，即便 A 分配给 B 0 元，B 也有资本对 A 进行惩罚，因此每一轮试次中 B 对 A 的惩罚金额在区间[0, 25]中。
- (3) **不公平试次数**：惩罚条件的 25 个试次中，提议者分配金额低于 50 元的试次数。
- (4) **A 对 B 不公平分配的总金额**：惩罚条件的 25 个试次中，提议者分配金额低于 50 元试次的分配金额总数。

因此，公式中的分子是不公平试次（即分配金额小于 50 元）中的总惩罚金额；而分母是不公平试次中提议者亏欠接受者的总金额（以 50 元为基准）。当惩罚力度为 1 时，意味着提议者亏欠多少金额，接受者相应地惩罚了多少金额。当惩罚力度小于 1 时，意味着接受者的惩罚金额小于提议者的亏欠金额。当惩罚力度大于 1 时，意味着接受者的惩罚金额大于亏欠金额，可能表现为接受者对提议者进行惩罚直至双方最后的收益相近或相等。总之，比值越大，说明惩罚力度越大。

在修改稿“2.5.1 行为数据”中也增加了对公式的解释。

审稿人 2 意见：

针对一稿，作者除了对原建议做了详细回复和修改以外，还对全文的论述结构做了较大幅度的修改。同意作者对一稿建议的大部分答复，这些局部修改使得实验细节的叙述更加到位。也认可作者关于论述结构的大的修改，相比一稿，这样的论述更加清晰而简洁。针对改动后新稿（由于作者的修改，一稿中提及的一些问题已被删除），有一些新的问题同作者商榷：

回应：感谢审稿专家对于修改稿的评阅以及对修改内容的认可。我们也对审稿专家提出的问题进行逐一回复，具体如下：

意见 1：本文的主要因变量是同步性（INS），这个指标是否有基线的参照？双方的某一脑区“都不激活”或者“都有很低的激活”时，“同步性”应该也挺高。

回应：在以往的采用近红外超扫描的研究中，研究者往往会将任务状态下和休息状态下的脑间活动同步性（INS）相减，然后再对“增加的脑间活动同步性”进行单样本 t 检验以检验是否存在 INS，或进行差异检验以比较不同条件下 INS 的差异。但是，休息状态下与任务相关的脑区激活可能不强，但是互动双方的特定脑区都处于“很低激活”的时候，脑间同步性也可能很高。已有研究显示，对于单脑研究而言，休息状态并不总是为一个合理的基线(Stark & Squire, 2001)。对于多脑研究而言，存在同样的问题。休息和任务状态是截然不同的两种状态，无论脑区是否显著激活，都可能存在脑间活动同步性。因此，本研究中参照以往研究(Reindl, Gerloff, Scharke, & Konrad, 2018)，没有将任务状态和休息状态的 INS 相减，而是直接通过比较惩罚与无惩罚条件下 INS 的差异来探究人际公平形成的脑-脑机制。

Reindl, V., Gerloff, C., Scharke, W., & Konrad, K. (2018). Brain-to-brain synchrony in parent-child dyads and the relationship with emotion regulation revealed by fNIRS-based hyperscanning. *Neuroimage*, 178, 493-502.

Stark, C. E., & Squire, L. R. (2001). When zero is not zero: the problem of ambiguous baseline conditions in fMRI. *Proceedings of the National Academy of Sciences*, 98(22), 12760-12766.

意见 2：继续前一个问题，在比较同步性这一因变量之前，应有一个全脑激活水平的结果。（注意到作者在给审稿人的答复意见中有，这个结果应该加入正文）。

回应：感谢审稿专家的建议，但是我们还是没有将单脑的结果放入文中，主要基于以下两个原因：**(1)** 本文的创新之处在于将超扫描技术和修改版的最后通牒博弈任务相结合，从群体脑出发，通过计算脑间活动同步性来揭示社会互动过程中人际公平形成的脑-脑机制，因此，本文的关注点在于双脑关系而非提议者或接受者做出社会决策的单脑激活。**(2)** 虽然单脑激活为探究双脑机制时脑区的选择提供了参考（由于通道数的限制，基于近红外的超扫描研究一般无法做到全脑覆盖），但是单脑的激活程度和个体间脑间活动同步性仍是不同层面的神经指标，二者所揭示的内容也截然不同。鉴于本文以双脑关系为研究点，若将单脑激活的相关结果也放入文中，内容上则会显的冗余。

意见 3：实验中，如果被试彼此熟悉，会大大弱化实验结果的解释，在该实验中，如何确保被试彼此不熟悉？

回应：审稿专家担忧的这一问题的确在实验设计时我们也考虑到了。因此在实验前，为确保被试之间相互陌生，我们会收集被试的姓名、学院和专业，在进行配对的时候，向被试确认是否认识另外一名被试，如果双方均不认识对方，就进行配对；如果双方有一方认识，就换其他被试，直到双方均不认识对方。配对完后，再告知被试实验时间，而后按照约定的时间完成实验任务。我们已在“2.1 被试”中增加“**每组被试在实验前互不认识。**”

意见 4：该实验程序是否有不完善之处，例如，惩罚额度让提议者的金额少于 0 怎么处理？

（这是可能的。如在某轮分配中提议者自己分 80，给接受者 20，接受者认为不公平，用 25 元“惩罚”）

回应：感谢审稿专家的细心审阅。在程序设计中，当接受者让提议者的金额已经为 0 时，即便接受者还有金额可以进行惩罚，接受者也不能再进行惩罚了，因此惩罚试次中，提议者最终金额最小为 0，不会为负数。我们已在“2.3 实验流程”中增加“**需要注意的是，惩罚过程中，接受者仅能减少提议者的金额直至为 0 元。**”

意见 5：“频段”是否有其生理意义，是否应该自上而下地确定？目前选择的两个频段，是发现有显著了，再来分析，这样的方式是否是一种循环论证？

回应：感谢审稿专家的细心审阅，对于选频段的问题，我们也进行了深入的思考，采用目前的方式进行频段的选择主要基于以下思考：

采用小波变化相干性（WTC）计算脑间活动同步性（INS），目前已发表的论文主要采用 4 种形式选择频段：（1）根据单个试次的时长并结合 wtc 图(Cui, Bryant, & Reiss, 2012; Hu, Hu, Xianchun, Pan, & Cheng, 2017; Osaka et al., 2015; Tang et al., 2016)；（2）根据前人研究结果(Cheng, Li, & Hu, 2015; Pan, Cheng, Zhang, Li, & Hu, 2017)；（3）选择避免全局生理信号的较大频段，例如 0.01-0.1Hz(Jiang et al., 2015; Jiang et al., 2012)；（4）在（3）的基础上，对频段内的每一频逐一计算，从而得出与任务相关的频段范围。

本研究虽然也是 event 设计，可以根据单个试次时长来选择相应的频段。但是该方法一般用于较为简单的实验范式，例如合作按键任务(Cui et al., 2012)或者时间估计任务(Hu et al., 2017)，对于复杂任务，可能涉及多个频段。再者，目前采用 wtc 计算复杂社会互动中脑间同步性的研究还较少，探究的问题和采用的范式也不同，因此，无法直接采用已有研究中报告的频段来计算本研究中的 INS。另外，也有研究选择了较大的频段范围，例如选择 0.01-0.1Hz 计算对话过程中的 INS(Jiang et al., 2012)，但是已有研究显示，该频段中仅有部分频段与对话任务有关(Nozawa, Sasaki, Sakaki, Yokoyama, & Kawashima, 2016)，无法明确其余的频段对结果产生的影响。

近年来涉及到复杂的人际互动的超扫描研究，往往会采用逐频计算的方式即本文的方法来选择与任务相关的频段，这些研究涉及师生教学互动(Liu et al., 2019; Pan, Novembre, Song, Li, & Hu, 2018; Zheng et al., 2018)、团体头脑风暴(Lu, Xue, Nozawa, & Hao, 2018; Xue, Lu, & Hao, 2018)、心理咨询(Zhang, Meng, Hou, Pan, & Hu, 2018)等。本研究探讨的是社会互动中人际公平形成的脑-脑机制，如之前的研究一样，也涉及到复杂的高级社会认知过程，因此，参考了以往的研究，采用了方法 4 来选择频段计算 INS，并发现两个频段与人际公平的形成相关，分别是 0.03~0.04Hz (28.02s~33.32s) 和 0.13~0.2Hz (5.24s~7.86s)（如图 1 所示）。任务中惩罚条件下 trial 的平均时长为 29.39s \pm 6.20s，无惩罚条件下 trial 的平均时长为 21.65 \pm 2.49s，所有 trial 的平均时长为 25.52 \pm 3.94s。0.03~0.04Hz 这一频段与 trial 的平均时长相接近。另外，已有研究发现教学互动与频段 0.16~0.31Hz (3.2s~6.4s) 的 INS 相关(Liu et al., 2019)，对话交流中出现频段 0.14~0.20Hz (4.95s~7s) 的 INS(Nozawa et al., 2016)，这些研究结果可能说明频段 0.1~1Hz (1s~10s) 与高级的社会认知过程紧密相关。需要注意的是，在频段分析中，没有发现无惩罚条件下哪一频段的 INS 高于惩罚条件，这一结果也说明了人际公平的形成伴随着更强的 INS。

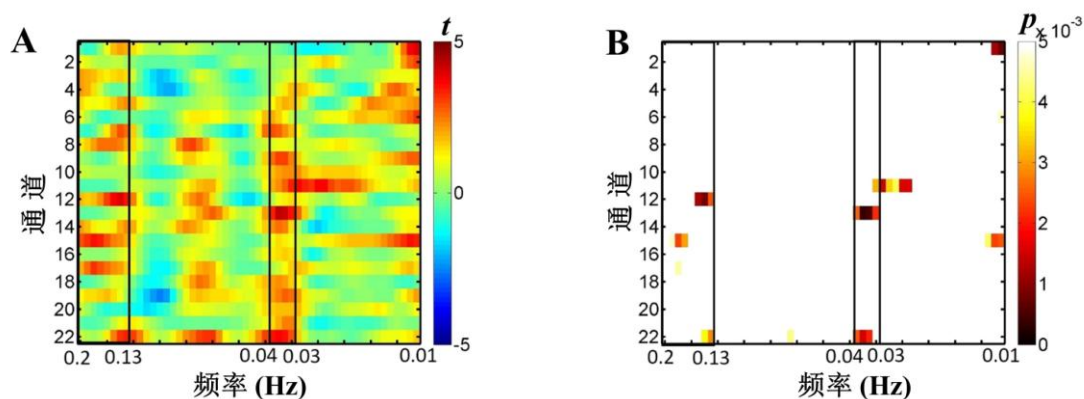


图 1 频段的选取

Cheng, X., Li, X., & Hu, Y. (2015). Synchronous brain activity during cooperative exchange depends on gender of partner: A fNIRS-based hyperscanning study. *Human Brain Mapping, 36*(6), 2039-2048.

Cui, X., Bryant, D. M., & Reiss, A. L. (2012). NIRS-based hyperscanning reveals increased interpersonal

- coherence in superior frontal cortex during cooperation. *Neuroimage*, 59(3), 2430-2437.
- Hu, Y., Hu, Y., Li, X., Pan, Y., & Cheng, X. (2017). Brain-to-brain synchronization across two persons predicts mutual prosociality. *Social Cognitive and Affective Neuroscience*, 12(12), 1835-1844.
- Jiang, J., Chen, C., Dai, B., Shi, G., Ding, G., Liu, L., & Lu, C. (2015). Leader emergence through interpersonal neural synchronization. *Proceedings of the National Academy of Sciences*, 112(14), 4274-4279.
- Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., & Lu, C. (2012). Neural synchronization during face-to-face communication. *Journal of Neuroscience*, 32(45), 16064-16069.
- Liu, J., Zhang, R., Geng, B., Zhang, T., Yuan, D., Otani, S., & Li, X. (2019). Interplay between prior knowledge and communication mode on teaching effectiveness: Interpersonal neural synchronization as a neural marker. *Neuroimage*, 193, 93-102.
- Lu, K., Xue, H., Nozawa, T., & Hao, N. (2018). Cooperation Makes a Group be More Creative. *Cerebral Cortex*.
- Nozawa, T., Sasaki, Y., Sakaki, K., Yokoyama, R., & Kawashima, R. (2016). Interpersonal frontopolar neural synchronization in group communication: An exploration toward fNIRS hyperscanning of natural interactions. *Neuroimage*, 133, 484-497.
- Osaka, N., Minamoto, T., Yaoi, K., Azuma, M., Shimada, Y. M., & Osaka, M. (2015). How Two Brains Make One Synchronized Mind in the Inferior Frontal Cortex: fNIRS-Based Hyperscanning During Cooperative Singing. *Frontiers in Psychology*, 6, 1811.
- Pan, Y., Cheng, X., Zhang, Z., Li, X., & Hu, Y. (2017). Cooperation in lovers: An fNIRS-based hyperscanning study. *Human Brain Mapping*, 38(2), 831-841.
- Pan, Y., Novembre, G., Song, B., Li, X., & Hu, Y. (2018). Interpersonal synchronization of inferior frontal cortices tracks social interactive learning of a song. *Neuroimage*, 183, 280-290.
- Tang, H., Mai, X., Wang, S., Zhu, C., Krueger, F., & Liu, C. (2016). Interpersonal brain synchronization in the right temporo-parietal junction during face-to-face economic exchange. *Social Cognitive and Affective Neuroscience*, 11(1), 23-32.
- Xue, H., Lu, K., & Hao, N. (2018). Cooperation makes two less-creative individuals turn into a highly-creative pair. *Neuroimage*, 172, 527-537.
- Zhang, Y., Meng, T., Hou, Y., Pan, Y., & Hu, Y. (2018). Interpersonal brain synchronization associated with working alliance during psychological counseling. *Psychiatry Research: Neuroimaging*, 282, 103-109.
- Zheng, L., Chen, C., Liu, W., Long, Y., Zhao, H., Bai, X., . . . Lu, C. (2018). Enhancement of teaching outcome through neural prediction of the students' knowledge state. *Human Brain Mapping*, 39(7), 3046-3057.

意见 6: 在 3.2 fNIRS 结果中，并未直接报告有颞顶联合区（rTPJ）的相关通道激活，但在讨论中却针对它的激活有详细的分析，结果和讨论似乎无法对应。

回应: 在讨论中我们未对颞顶联合区的激活进行详细的分析，仅提及“rTPJ 也是心理理论加工的重要脑区，大量的研究证据表明 rTPJ 不仅与推断他人信念有关(Perner, Aichhorn, Kronbichler, Staffen, & Ladurner, 2006)，还涉及表征他人行为意图、整合意图与行为结果等(Koster-Hale, Saxe, Dungan, & Young, 2013; Young & Saxe, 2008)。”惩罚条件下，互动双方为利益最大化，进行着人际博弈，根据反馈推断互动对象的行为意图，从而调整分配金额或者惩罚金额。因此，惩罚条件下颞顶联合区增强的脑间活动同步性可能反映了人际公平形成中互动双方更强的心理交互。

意见 7: 在讨论中，作者认为，惩罚条件下的效应可能有两种原因导致，即“人们对不公平的普遍厌恶”和“惩罚的效应”，可实际上，前者在无惩罚条件下也应该有，这一部分的效应的脑区激活可能被减掉了。

回应：提及“人们对不公平的普遍厌恶”是用于解释“惩罚下提议者和接受者最终的平均收益没有显著的差异”这一结果，由于个体对优势不公平和劣势不公平的厌恶，接受者面对不公平的分配方案时会实施惩罚措施，但是惩罚不会过度，惩罚结果往往使双方的最终收益相等或者相近。

意见 8：p.2. （该页码指的是正文页码，不包括“修改人意见回复”和“自检报告”，下同）：“Makwana 等人（2015）使用了相同的实验范式，研究结果与上述保持一致。”如果在实验设计和结果上均无不同，该句可删除，只加引用即可。

回应：已根据审稿专家的建议删除该句，改为引用。

意见 9：p.2. “且两条件的分配金额差异越大，右侧背外侧前额叶皮层、左侧眶额皮层、右侧尾状核激活程度的差异也越大”，该句话叙述不清晰，是比条件还是比金额？

回应：已对上述表述进行了修改，具体为“已有研究显示，相比无惩罚条件，惩罚条件下提议者的分配金额更高，分配过程中双侧背外侧前额叶皮层（dlPFC）、腹外侧前额叶皮层（vlPFC）、眶额皮层（OFC）以及双侧尾状核（caudate nucleus）的激活程度增强(Spitzer et al., 2007)，且惩罚条件下右侧背外侧前额叶皮层、左侧眶额皮层、右侧尾状核激活程度的增强与公平规范遵从程度的提升（惩罚下分配金额减去无惩罚下分配金额）呈正相关。”

意见 10：p.2. 对于提议者，抑制 dlPFC 会“降低分配方案的公平性”，而对于接受者，抑制 dlPFC 会“导致惩罚行为减少”。作者认为该脑区“负责自利反应的抑制”。如果按这个解释，以上结果是矛盾的。实际上，dlPFC 可能仅反映的是对公平的感知。

回应：感谢审稿专家提出的建议。Ruff 等人(2013)为探讨右侧背外侧前额叶皮层与公平规范遵从的因果关系，将经颅直流电刺激（tDCS）和修改版的最后通牒博弈任务相结合。实验中对被试右侧 dlPFC 施以阳极刺激（增强激活）或阴极刺激（抑制激活）或伪刺激（不干预），而后被试作为提议者完成多轮金钱分配任务。结果显示，增强右侧 dlPFC 的激活程度可显著提升提议者的分配金额，使其更接近平均分配；而抑制该脑区则会降低分配方案的公平性。为了揭示行为差异背后的心理机制，研究者要求被试完成任务后对不同分配方案的公平性、接受者面对不同分配方案的愤怒情绪进行评分，以及预测接受者面对不同分配方案的惩罚金额。研究者发现增强或抑制右侧 dlPFC 的激活程度并不会影响个体对于公平的信念和感知，以及对不公平行为后果的预测。因此，dlPFC 更有可能反应的是认知控制，这种 dlPFC 功能的解释也受到了很多研究者的认可。在上述任务中 dlPFC 认知控制的功能具体表现为“自利反应的抑制（分配给自己更多的金钱）”，因而抑制提议者 dlPFC 的激活则会阻碍其抑制自利反应，从而降低分配方案的公平性。研究者同样采用了无创的干预技术（eg. TMS, tDCS）探究了抑制右侧 dlPFC 对接受者公平感知的影响(Baumgartner, Knoch, Hotz, Eisenegger, & Fehr, 2011; Knoch et al., 2008; Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006)。实验任务中，要求被试与多名互动对象完成一次性的最后通牒博弈任务。研究者发现当右侧 dlPFC 被抑制时，接受者面对不公平分配时惩罚行为显著减少即接受更多的不公平分配。在一次性的最后通牒博弈任务中，惩罚行为并不会影响后续试次的分配方案，因此，无论分配多少全部接受是自利行为的表现。因此，抑制 dlPFC 导致惩罚行为减少这一结果也能反应“dlPFC 负责自利反应的抑制”。另外，研究者还发现抑制和非抑制两组被试对不同分配方案公平程度的评价没有显著差异，说明抑制 dlPFC 并不影响个体对于公平的感知，为 dlPFC 认知控制的解释提供了进一步的支持。基于以上研究结果，我们还是保留了修改稿中对于右侧 dlPFC 功能的理解。

Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and

- ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience*, 14(11), 1468-1474.
- 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.
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, 314(5800), 829-832.
- Ruff, C. C., Ugazio, G., & Fehr, E. (2013). Changing social norm compliance with noninvasive brain stimulation. *Science*, 342(6157), 482-484.

意见 11: p.6. “2.5.1 行为数据”的分析方法似可删除，在结果中直接表述即可。

回应: 感谢审稿专家提出的建议。因为“2.5.1 行为数据”中涉及到 t 检验效应量大小的说明以及惩罚力度的计算公式，为使行为结果的报告更加简洁明晰，我们还是决定在方法中保留该部分。

意见 12: p.8. 全体被试互动情况的描述性统计结果是否有必要列出？

回应: 虽然通过比较惩罚和无惩罚条件下的分配金额，可以发现惩罚能够促进人际公平的形成，但是社会互动是一个动态变化的过程，平均数和标准差不能展现这一过程。而列出全体被试互动情况，一是可以让读者看到这一动态过程，二是揭示了人际公平的形成是互动双方相互博弈的结果，为“脑间活动同步性可能反应了博弈中互动双方更强的心理交互”提供了支持。

第三轮

编委复审: 经两轮审稿和修改，作者已很好地回答了审稿人的意见，并做了仔细、恰当地修改。

主编终审: 本研究使用最后通牒博弈任务的变式，利用超扫描技术，探讨了社会互动中人际公平形成的脑机制。研究最终发现，惩罚促进人际公平的形成与互动双方前额叶、颞顶联合区等部位更强的脑间活动同步有关。本研究为理解人际公平的内在机制提供了新的视角。论文研究方法使用恰当，研究结果可信，研究结论具有一定理论意义。