《心理学报》审稿意见与作者回应

题目：重复信任博弈的决策过程与结果评价
作者：王益文　张振　郭丰波　原胜　敬一鸣

第一轮

审稿人 1 意见：

意见 1：科学假设不清晰。在引言部分第四段，作者对 P2，N2，FRN，P300 等 4 个 ERP 成分进行了综述，但是并未阐明为何这四个成分可能与信任过程存在联系，每个成分所反映的心理过程对信任过程可能存在怎样的影响。换言之，该研究并未对“信任决策脑机制的时间进程”这一问题提出明确的科学假设，只是描述了一些实验预期。而且这些实验预期与文中的理论综述缺乏逻辑联系。对 P2 和 N2 成分，引言提到它们分别与注意选择和控制加工以及认知控制相关，并提到“而且个体在信任博弈中往往存在较高的合作倾向”。这与接下来提出的预期“决策阶段中不信任选择比信任选择诱发更大的 P2 和 N2”并不存在任何明显的逻辑联系。而对于 FRN，“损失反馈比获益反馈诱发更负的 FRN”已经被大量研究所证明过，该研究并未提供任何新的证据或能对现有理论进行任何补充。对于 P300，“获益反馈比损失反馈诱发更正的 P300”这一预期并不能从前文提到的任何理论或证据中推断出来。

回应：非常感谢审稿人对实验假设引入所提的宝贵意见。我们已经据此对初稿引言部分进行了重新修改和补充，详尽描述了相关研究假设的理论基础与先前研究结果，具体说明如下：

*对于博弈决策阶段，目前关于“个体为何选择信任或不信任，以及两种选项的优势性问题”的理解存在两种相悖的理论观点。背叛厌恶理论认为信任涉及到策略不确定性，为回避背叛风险个体倾向于选择不信任行为(Bohnet, Greig, Hermann, & Zeckhauser, 2008)；而强制规范理论则强调信任是受道德规范制约的行为，个体为了显示对他人品质的尊重，更倾向于选择信任以维持他人值得信赖的社会矫饰(Dunning, Anderson, Schlösser, Ehlebracht, & Fetchenhauer, 2014)。鉴于目前关于信任互动情境下博弈决策过程的 ERP 研究较为缺乏，为了从神经电生理水平上探讨信任决策中两项选项的优势性问题，我们基于风险决策和侧抑制任务的 ERP 研究(Gajewski, Stoerig, & Falkenstein, 2008; Gajewski, & Falkenstein, 2013)，提出了通过 P2 和 N2 两种成分检验上述问题的研究思路。鉴于研究者关于博弈决策中不同行为选择的优势性问题尚存争议，本研究提出假设一：如果决策博弈阶段信任选择为优势选项(与**强制规范理论**相一致)，则个体会更多的选择信任，而且不信任选择比信任选择诱发更大的 P2 和 N2 成分；反之，如果不信任选择为优势选项(与**背叛厌恶理论**相一致)，则个体会更多的选择不信任，同时信任选择比不信任选择诱发更大的 P2 和 N2 成分。*

对于结果评价阶段，鉴于当前采用 TG 范式探讨结果评价过程的脑电研究上不充足(Chen et al. 仅涉及结果评价中的 FRN 成分)，而且同时考察 TG 范式中博弈决策阶段与结果评价阶段的脑电研究也十分匮乏，因此当前研究试图以 FRN 和 P300 为指标，补充相关方面的研究数据。另外，依据结果评价的独立编码理论及相关研究结果，FRN 和 P300 分别独立负责编码或加工结果的效价与大小(Sato et al., 2005; Yeung & Sanfey, 2004)，即 FRN 反映了一种基于结果的好坏或预期目标是否实现的二分法所完成的早期自动化评价过程，而 P300 则反映了一种基于动机/情感意义或者注意资源分配的晚期控制评价过程(Yu & Sun, 2013)；并且已有研究发现 P300 对金额大小的数量加工存在参照点效应，即以赌注为基准的相对数量加工(Xiang, Wang, Zhang, & Yuan, 2008)。因此，我们对 FRN 和 P300 的研究假设进行了调整，即研究假设二：损失反馈比获益反馈诱发更负的 FRN，损失反馈与获益反馈诱发的 P300 波幅

无显著差异。

引言部分已按照上述思路进行修改，详见文中引言部分绿色字体，请审稿人审查。。

意见 2：实验情境设置可能存在问题。文中在对“重复匿名互动方式”进行描述时提到：被试被告知在实验者所面对的受托者的决策策略是事先采集的， 采集的方式是要求受托者描述“在重复匿名互动条件下，如果其被信任且给予了 30 点，那么他们如何对此情况进行反应”。这一对指导语的描述并不清晰。其中“重复”是如何定义的?为何只描述被信任之后的一轮反应?在信任游戏中，一个相当重要的前提是存在人际互动过程，并通过在互动过程中对于对方行为模式的学习来调整自身的行为策略，以求获得最高的期望收益。而在该研究所描述的情境中，这一学习与调整的过程是缺乏的。根据目前文中所描述的指导语，被试并不能知道选择信任或者不信任对方是否会对对方之后的行为产生影响，因此实验情境的真实性和有效性可能受到影响。

回应：常感谢审稿人对实验情境操纵所提供的富有建设性的意见。确实，初稿对实验情境的描述过于简洁，略显模糊。现对实验指导语所述内容进行补充说明，以期提高实验情境的真实性与有效性：*受托者的决策策略是事先采集的，具体而言，脑电被试被告知脑电实验前实验者从社会上选择了 400 名具有社会代表性的成人被试，要求他们扮演受托者与另一名实验助手在网络上完成 30 回合的重复性 TG 任务，然后对其在这些回合中的行为选择进行分析与学习，采用计算机程序模拟这 400 名被试的反应策略，并贮于计算机当中；在脑电被试进行信任博弈时，计算机会从中随机选择一个被试的反应策略以完成所有回合的游戏*。详见文中任务与程序部分绿色字体。

需要强调的是，重复复互动过程非常复杂多变, 难以在一个实验中探索明晰, 需要系列实验研究才可探讨. 本实验操纵是一种预先固定式的重复信任操纵, 不存在重复互动过程. 实际任务中采用预先编制的、结果强化率为 50%的电脑程序来操纵受托者的行为，因此脑电被试可能无法有效地从互动过程中获得对方策略，或者依据结果反馈调整随后的行为决策。这种缺陷是实验室实验与真实社会互动过程之间必然存在的差距，因此，作为探讨信任博弈互动的初步研究，当前研究无法探究重复互动过程中的强化学习过程。在此对这种不足进行说明，相关问题详见问题 3 的解答. 在进一步的研究中将专门探讨信任双方的策略互动过程。详见文中方法部分绿色字体，请审稿人审查。

意见 3：数据分析可能存在问题。如前所述，信任的建立是个强化学习的过程，因此随着决策次数的增加，被试的心理过程和行为模式都可能存在变化。在行为数据分析层次上，作者只给出了信任和不信任的平均比例，而并未描述出这一比例是否随时间增加而发生变化。而如果这一比例会发生变化，例如随着时间的增加信任比例降低（因为学习到了实验设置的反馈强化率是 50%），那么在对 ERP 数据进行叠加的时候信任条件的试次将更多来自实验前半段，而不信任条件更多来自后半段，这可能造成信噪比差异等问题。再者，在神经活动层次上，强化学习理论指出随着信任的建立，决策相关神经元（或脑区）的活动将会提前(King-Casas et al., 2005)。这会导致在 ERP 叠加上的严重问题。因为对 EEG 信号进行叠加平均提取 ERP 信号的前提是：ERP 成分对相应的事件是锁时锁相的。而信任过程可能会对锁时前提造成破坏，因此导致无法有效地提取到决策过程相关的 ERP 成分。而且，学习过程可能产生试次间的影响，例如前一个试次的损伤反馈可能导致下次决策的不信任，这一点在该研究中并未被考虑到。

回应：非常感谢审稿人对数据分析与结果解释部分的提醒。具体说明如下所述：

1） 依据审稿人的建议，我们将所有试次分为前后各 75 试次，以分析前后两段中被试信任率的变化。配对 t 检验表明前后 75 试次中信任率的不存在显著差异，$t(19)=1.75$, $p=0.096$，

即前 75 试次的信任率(72.3±10.3%)与后 75 试次的信任率(67.4±12.3%)基本一致，不存在前后阶段信任率差异对 ERP 数据信噪比的影响。

2） King-Casas et al.研究主要分析了受托者意图增加信任(即受托者当前回合返还的金钱要高于之前回合返还的金钱，此时能够向信任者传递受托者具有较高的合作意愿)时的神经活动，这种互动模式下受托者的返还比例是递增的，容易促使双方形成良好的互动声誉与预期。当前研究中受托者的返还比例(即反馈强化率)为 50%，相比较而言其所形成的“互动声誉与预期”应该是接近中性的，且前后 75 试次的信任率不存在显著差异，表明被试在任务中存在较弱的强化学习过程。因此，扮演信任者的被试大脑活动应该不会随着互动而发生较大的变化，所提取的决策阶段 ERP 成分是可信的。

3） 为了检验当前互动过程中学习过程可能产生试次间的影响，例如前一个试次的损伤反馈可能导致下次决策的不信任，我们对不同反馈(选择信任后的获益反馈<20>、不信任后的反馈<10>、选择信任后的损失反馈<0>)后的信任选择率进行了单因素重复测量方差分析，结果表明不同反馈后的信任率不存在显著差异，$F(2,38)=2.78$，$p=0.076$，$\eta^2=0.13$[不同反馈后的信任率分别为：获益=76.6±3.5%; 中性=64.4±3.2%; 损失=64.6±4.6%]；单独对获益或损失反馈后信任率的配对 t 检验结果也没有发现两者存在显著差异，$t(1,19)=1.93$, $p=0.068$。另外，依据已有研究(Bai et al., 2014; Gehring & Willoughby, 2002)，我们也尝试了不同反馈(获益/损失)后信任率与相对应反馈诱发的 FRN/P300 波幅之间的相关，结果表明：损失反馈后的信任率与损失反馈诱发的 FRN(r=-0.16, p=0.51)和 P300(r=-0.11, p=0.65)相关不显著；获益反馈后的信任率与获益反馈诱发的 FRN(r=-0.16, p=0.50)和 P300(r=-0.36, p=0.12)相关不显著。简言之，这些结果表明当前实验当中可能存在较弱的强化学习过程，并不存在试次间的互动影响，这可能是由于当前实验中结果反馈率为中性的 50%所造成的。

另外，仍需说明的是，当前研究只采用几率水平(50%)的结果反馈率，其目的就是为了初步探索信任博弈决策过程中大脑加工的动态过程，这种实验操纵使得我们无法探讨互动过程中的强化学习过程或者声誉形成过程(前后 75 试次的信任率以及不同反馈后信任率的分析也支持这种理解)，同时这种过程也并非当前探索性研究的主要目的。关于互动过程中声誉形成过程以及反应策略的探讨，我们也会在后续研究中进行逐步的、有条理的分析与探索。请审稿人审查。

意见 4：ERP 数据分析描述不全面。1）数据是否进行了滤波？2）对反馈进行事件相关分割时候的锁时点分别在哪里？基线选取多长？3）基线矫正是如何进行的? 4）各条件叠加次数的差异，尤其是信任和不信任条件下的差异是否会对数据分析产生影响？5）各 ERP 成分的时间窗和感兴趣电极是如何选定？有什么证据表示必须选择这些时间窗和电极？尤其是 P2 成分，在 PZ 处其总平均的峰值点几乎处在该时间窗的边缘。这对其峰值和潜伏期的提取是否有影响? 6）对 P2 和 N2 进行统计的时候，2×5×3 的 ANOVA 是否适用于只有 20 个被试的样本？7) 对 FRN 和 P300 的分析时，选择信任时的反馈是如何处理的？8）对 sLORET 对象的描述的“评估不信任决策与信任决策所诱发 ERP 差异成分的 3D 激活源”，但是结果部分只有 P2 没有 N2。

回应：非常感谢审稿人对数据分析部分所提的细致入微的建议。我们已经据此对文稿的内容进行了逐一的修改和补充，具体说明如下：

1） “*所得 ERP 波形进行 30Hz 的低通滤波*”，该信息已添加到。

2） “*反馈阶段分析时程为试次结果反馈前 200 ms(作为基线)到呈现后 1000 ms*”已添加。

3） 基线矫正是依据 NeuroScan 软件的模块化程序完成的。

4） 叠加次数对 ERP 数据的信噪比可能存在较大影响，为评估决策阶段信任选择与不信任选择叠加次数对数据分析的影响，我们从两种条件的总平均文档内导出了所分析电极点的信

噪比(SNR)数据，并进行配对 t 检验。统计结果表明两种条件下信噪比没有显著差异，$t(14)=1.65$，$p=0.122$。这就表明两种条件下叠加次数对 ERP 数据的影响是可以接受的。

5） 对于决策阶段而言，由于目前探讨信任决策阶段的 ERP 研究几乎没有，可借鉴性文章较少，因此我们选择前后分布(Fz、FCz、Cz、CPz、Pz)和左右分布(左(3)、中(z)、右(4))以试图更为全面的了解决策阶段 ERP 数据的特性；P2 和 N2 时窗选择是依据对原始波形图的视觉检测，并考虑两种成分的临近性而选取的。对于 Fz 点而言，P2 的平均潜伏期为 217±6ms，基本处于所分析时窗内，因此所选取的分析时窗能够较大程度反映 P2 峰值与潜伏期。对于反馈阶段而言，电极点的选择主要依据已有文献表明脑中线 5 电极点上 FRN 和 P300 成分最大；分析时窗则依据当前研究中 ERP 波形图特性以及两种成分的经典分析时程进行确定的(Gehring & Willoughby, 2002; Leng & Zhou, 2010; Yeung et al., 2004)。

6） 对 P2 和 N2 完成的 2×5×3 的 ANOVA 中三种因素均为被试内因素(后两者为电极分布因素，以探测脑电获得的边侧化效应或前后分布效应)，已有的脑电研究中大部分均采用小样本被试(一般小于 20 名被试)，因此依据已发表的脑电研究来看(Wang et., 2011; Wu & Zhou, 2012)，基于小样本被试对脑电成分进行 2×5×3 的重复测量方差分析是可行的。

7） 在实验前对整个信任博弈程序的反馈进行编码，具体为 150 试次中损失反馈和收益反馈各 75 次，然后进行随机排列。在某个试次中，当被试选择信任时则依据反馈编码给以损失或收益反馈。基于反馈编码（损失或反馈）是随机排列的，因此理论上可以实现被试选择信任时损失反馈与收益反馈各半的情况，而且当前研究结果中损失反馈(48±7)与收益反馈(49±8)的叠加次数基本一致，也进一步表明这种反馈设置的有效性。

8） 由于 sLORETA 是一种以 ERP 数据为基础的逆向溯源分析，其前提就是所关注的 ERP 成分必须存在条件主效应，否则对无条件差异的 ERP 成分进行溯源分析是没有任何意义的，因此结果部分只呈现了 P2 的溯源结果。为排除前后描述的不一致，对 sLORETA 对象的描述已修改为“当 ERP 成分存在显著的条件差异时，采用 sLORETA 评估不信任决策与信任决策所诱发 ERP 差异成分的 3D 激活源”。

请审稿人审查。

意见 5：讨论阶段对前人文献综述过于宽泛，结构松散，并未体现出与该研究结果的明显逻辑关系。

回应：非常感谢审稿人对初稿讨论部分所提的细致深入的意见。为避免讨论部分结构松散，我们对初稿讨论部分的框架进行了创新整理，设置了四个小部分：“遵循强制规范的信任行为”、“P2 反映了博弈决策中信息整合引起的冲突”、“FRN 反映实际结果与事先预期之间的偏差”、“P300 反映金钱奖励的动机/情感意义加工”，竭力依据本研究结果开展讨论，增强讨论部分的逻辑关系。详见修改稿中讨论部分绿色字体，请审稿人审查。

意见 6：次要问题：

1. 引言部分很多信息与研究本身并未直接关系。如其他因素对信任过程的影响，决策过程的三阶段，P2 和 N2 与冲突加工等。
2. 图和对实验设计描述时采用了英文词汇，如要在中文期刊上发表则需进行仔细修订。
3. ERP 总平均叠加次数没有标准差。
4. 结果描述是有效数字位数不正确：对反应时和潜伏期的描述达到了.01 ms，这是超出了采样周期和统计需要的。
5. ERP 波形图格式不一致：图 2 的时间轴是在底部，图 3 则在 0 μV 处。
6. 图 2 差异波在当前尺度下完全看不出来。
7. 图 3 小标题错误：图 3B 并不只是 FRN，而是 FRN 和 P3 的共同体现；图 3C 不是 P300，

而是 300~600 ms 时间窗内每 100 ms 的地形图。

8. 为何要呈现 300~600 ms 内每 100 ms 的地形图? 文章对这一问题并没有任何统计分析或讨论。

9. 图上只标注了 P2 和 FRN 的时间窗，没有标注 N2 和 P300。

10. 为何决策选项呈现 2000 ms 而相关 ERP 只分析到决策选项后 600 ms？？

回应：非常感谢审稿人所提的宝贵意见，我们已经据此对文稿的内容进行了逐一的修改和补充，具体说明如下：

1. 引言部分已经进行发幅度的整理与修改，删减了其他部分与研究本身相关不高的信息，如其他因素对信任过程的影响，决策过程的三阶段。鉴于当前研究主要来探讨决策过程中不同选择优势性的问题，试图从冲突检测与认知抑制方面检验决策优势性当中的认知加工过程，因此我们仍然保留了 P2、N2 以及冲突加工部分，并对上述两方面内容进行了重新整理以突显该部分内容与本研究主题的密切关系。

2. 已经对实验流程图中的英文词汇进行了相应修改，以符合中文期刊发表要求。

3. 已添加了 ERP 总平均叠加次数的标准差。

4. 对结果部分反应时和潜伏期的描述已调整至整数。

5. 决策阶段与反馈阶段的 ERP 波形图格式已统一。

6. 已对图 2 进行了调整使得差异波更为明显。

7. 图 3 小标题已进行调整，修改为“不同反馈条件引起的 ERP 波形图（A）和地形图(B)”

8. 已将 300~600 ms 内每 100 ms 的 P300 地形图删去，调整为 300~600ms 时窗内 P300 地形图，进而确保与 FRN 地形图的一致。

9. 决策阶段和反馈阶段中 ERP 波形图中已标明相对应 ERP 成分的时间窗口。

10. 虽然决策选项呈现时间为 2000ms，但鉴于被试的反应时大约为 500ms，即此时被试的行为选择已经完成，对之后的数据进行分析并不恰当，因此，我们选择分析决策选项后 600ms 的 ERP 数据。

请审稿人审查。

审稿人 2 意见：

意见 1：论文题目要求简洁、清晰、逻辑性强，反应研究的领域及问题。本文题目建议修改。

回应：引言部分已按照上述思路进行修改非常感谢审稿人对论文题目所提的宝贵建议。题目已修改为“投资任务中信任博弈的决策过程与结果评价”。请审稿人审查。

意见 2：摘要的第一句话，言过其词。个体每天多少时间处于这种困境？怎么是经常面临的呢？

回应：非常感谢审稿人对论文摘要所提的意见。依据已有文献(Delgado, 2008)，该句已修改为“‘信任他人或不信任他人’是一种社会困境，直接影响着个体的社会生活。”详见摘要部分绿色字体，请审稿人审查。

意见 3：信任固然重要，如何证明“在人类社会的繁荣发展中起至关重要的作用”？本人认为摘要中的每一句话都应该是有理有据的，不是一种想法、说法、感想。

回应：非常感谢审稿人所提的宝贵建议。一些关于国际调查问卷的研究已经发现信任与国家或城市的经济增长呈现显著正相关，对人类社会的经济发展起着主要的推动作用(Algan & Cahuc, 2010, 2013; Knack & Keefer, 1997; Tabellini, 2010; Zak & Knack, 2001)，因此文章认为“信任在人类社会的繁荣发展中起至关重要的作用”。经过审稿人建议之后，为确保言辞的

准确性与客观性，该句已修改为“促进社会经济的发展”。详见摘要部分绿色字体，请审稿人审查。

意见 4： 本研究的发现，现实意义是什么？从 P2，FRN，P300 的变化，我们可以借鉴什么呢？讨论中应该详细说明。
回应：非常感谢审稿人对当前研究结果的理论与实际意义所提的检验。为此，我们已在讨论部分倒数第二段增添了相应的内容，即*本研究对人类信任行为领域的理论与实际意义在于：第一，首次采用 ERP 技术探讨决策博弈过程中大脑加工的动态时程变化，弥补了现有脑电研究主要关注结果评价阶段的局限；第二，为信任行为的强制规范理论提供了神经电生理方面的初步证据，即博弈决策中不信任选择比信任选择诱发更大的 P2 成分，一定程度上支持了“信任行为遵循强制规范”的观点；第三，为结果评价的独立编码模型提供了进一步的支持，即在信任互动这种特异的社会情境中，FRN 和 P300 分别独立的负责编码或加工结果的效价与大小*。请审稿人审查。

意见 5： 被试右利手如何选择的？使用的问卷吗？请介绍一下。
回应：非常感谢审稿人对被试筛选部分所提的意见。被试的基本情况(包括视力、精神疾病、右利手等信息)是依据其自我报告完成的，并没有采用任何问卷，请审稿人审查。

意见 6： 信任游戏的流程图，建议修改成汉字。Decision tree, Fixation, 等等。
回应：非常感谢审稿人对实验流程图呈现所提的意见。实验流程图中的英文已修改为中文，以符合国内期刊发表的要求。请审稿人审查。

意见 7： 如何确定被试的按键真的代表他信任呢？如果被试一直不信任，一直按不信任键，被试自己也会觉得不对，那就随机按吧。你如何判断是那种情况呢？此时的信任不信任叠加，意义在哪里？
回应：非常感谢审稿人所提的细致入微的意见。为了检验个体的行为选择是否属于随机按键，我们将被试的信任选择率与随机水平(50%)进行了单样本 t 检验，结果发现被试的信任率(69.8±2.1%)显著高于随机水平，$t(19)=9.46$，$p<0.001$。同时，我们将所有试次分为前后各 75 试次，以分析前后两段中被试信任率与随机水平(50%)之间的差异。单样本 t 检验表明，前 75 试次的信任率(72.3±2.3%)显著高于随机水平，$t(19)=9.67$，$p<0.001$；后 75 试次的信任率(67.4±2.7%)显著高于随机水平，$t(19)=6.31$，$p<0.001$。因此，上述结果表明实验中被试的按键远超过随机按键水平，其按键确实代表其真实意图与意愿。请审稿人审查。

意见 8： 信任 100 次，不信任 43 次，在结果处理上，如何排除影响？文中补充。
回应：非常感谢审稿人对数据分析部分所提的宝贵建议。叠加次数对 ERP 数据的信噪比可能存在较大影响，为评估决策阶段信任选择与不信任选择叠加次数对数据分析的影响，我们从两种条件的总平均文档内导出了所分析电极点的信噪比(SNR)数据，并进行配对 t 检验。统计结果表明两种条件下信噪比没有显著差异，$t(14)=1.65$，$p=0.122$。这就表明两种条件下叠加次数对 ERP 数据的影响是可以接受的。请审稿人审查。

意见 9： 行为结果发现，两种选择反应时没有显著差异。就是说按键没有冲突？还是按什么都一个心态？博弈过程中，选择信任与选择不信任都一样的反应时？行为数据反应信任行为的真实度，影响脑电结果的解释。
回应：非常感谢审稿人对结果部分所提的宝贵建议。两种选择反应时没有显著差异可能是由

于一下几方面造成的：首先，重复性社会互动中随着互动次数增加，被试可能会产生习惯化或采用固定策略进行反应。我们将所有试次进行了前后均分并比较了前后两个阶段中不同选择的反应时，结果发现前部互动阶段中信任选择的反应时(533±22ms)显著大于不信任选择(505±25ms)，而后部互动阶段中信任选择(484±24ms)与不信任选择的反应时差异不显著(494±27ms)，这种结果支持上述假定。其次，为将被试行为选择的动作准备效应最小化，以及降低其反应时的变异性，当前研究的指导语要求“被试看到决策选项图之后再进行反应”，这种操纵可能会导致两种选择的反应时没有显著差异(Boudreau, McCubbins, & Coulson, 2009)。最后，需要说明的是，已有 ERP 研究中也存在行为结果与脑电成分不一致的情况，这种相分离的结果可能是因为脑电指标比行为结果更敏感，或者脑电成分只反映信息加工的某个阶段，而行为结果则是整个信息加工过程的最终输出结果，进而导致了两者无法一一对应(Kounios & Holcomb, 1992; Wu & Zhou, 2012; Wang, Huang, Zhang, Song, & Bai, 2014)。综上所述，我们认为当前研究中反应时差异可能是由于实验指导语和重复性互动的特性造成的，前半程互动中不同选择的反应时差异则部分程度上支持了研究结果的有效性，并不影响脑电结果的解释。请审稿人审查。

意见 10：讨论 1 对本研究的结果讨论不多，都是说别人的研究及观点。这部分的讨论目的是什么呢？想解决什么问题呢？

回应：非常感谢审稿人对初稿讨论部分所提的宝贵意见。我们对初稿进行了大幅度的修改与调整，修改稿中讨论 1 部分试图从强制规范理论角度来解释当前的行为结果(为什么信任选择率显著高于几率水平)。详见修改稿讨论 1 部分的内容：*信任对于文明社会与人际关系的建立与发展是十分重要的，然而目前关于“个体为何选择信任或不信任，以及两种选项的优势性问题”的理解尚存在较大争议：背叛厌恶理论认为信任涉及到策略不确定性，为回避背叛风险个体倾向于选择不信任行为；而强制规范理论则强调信任是受道德规范制约的行为，个体为了显示对他人品质的尊重，更倾向于选择信任以维持他人值得信赖的社会矫饰。考虑到当前研究中结果强化率仅为 50%，实验中被试会体验到较高强度的背叛水平，有人可能会预期信任选择率可能会随着互动过程的发展而逐渐降低，总体上也应接近或低于几率水平。但是，研究结果却发现前后互动阶段中信任率没有显著变化，而且总的信任率更是显著高于几率水平，因此当前研究结果与强制规范理论相一致(Dunning et al., 2014；Dunning, Fetchenhauer, & Schlösser, 2012)。换言之，信任博弈中的信任行为在某种程度上是一种表达性行为，个体更在意信任行为本身及其所代表的某种心理意义；人们选择信任是为了遵循一种强制规范，即个体借此向他人的品格表示尊重，维持其关于他人值得信赖与诚信善意的社会矫饰，即使个体私下并不相信如此*。请审稿人审查。

意见 11：讨论 2 博弈决策过程什么机制呢？这部分只是讨论了不信任冲突更大，不信任涉及更多信息更新与整合。这只是推测，也没有讨论出机制是什么。

回应：非常感谢审稿人对讨论 2 内容所提的深刻宝贵建议。依据审稿人一的建议“避免讨论过于宽泛松散，忽视与该研究结果的明显逻辑关系”，我们对初稿讨论部分的框架重新进行了整理，讨论 2 已修改为“P2 反映了博弈决策中信息整合引起的冲突”，旨在依据已有研究结果详尽解释决策过程中不同选择导致的 P2 波幅差异。详见修改稿讨论 2 部分的内容：*在决策选项呈现后 200ms 左右，不信任选择比信任选择诱发了更大的 P2 波峰。作为冲突检测的早期成分，有研究认为 P2 与注意选择和控制有关，反映了决策启动的评估过程(Martin & Potts, 2004; Potts, 2004)。在侧抑制任务中，不兼容刺激诱发更大的 P2 波幅，可能反映了刺激评估和冲突检测过程(Nikolaev et al., 2008; Gajewski et al., 2008; Nikolaev et al., 2008)。在当前研究中被试的信任行为显著高于几率水平，表明遵循强制规范的信任行为是其优先或默认*

*选项，因此选择不信任行为可能会引起较大的认知冲突，进而导致更大的P2 波幅。与本研究结果相一致，Wiswede 等(2011)发现当个体遭遇对方挑衅并选择给予惩罚时会引起P2 波幅的增大，认为P2 对个体随后的趋近或撤回行为起着重要作用，在早期阶段即可反映个体惩罚对方的决策。另外，P2 差异波的偶极子定位于左侧额中回(middle frontal gyrus, BA 46)，属于背外侧前额叶皮层(dlPFC)。大多数研究者认为额中回负责工作记忆的信息激活与维持过程，也参与认知控制加工，如执行最优决策、过滤不相干信息等(Goel & Dolan, 2004; Picton et al., 2007; Polosan et al., 2011)。同时，奖赏学习的相关研究也发现 dlPFC 能够编码过去选择及其收益，为奖赏预期更新提供相关信号，进而引导有意识的目标指向性选择与适宜行为，保证个体做出最优决策(Barraclough, Conroy, & Lee, 2004; Paulus, Hozack, Frank, & Brown, 2002)。源定位结果表明不信任选择涉及更多的信息更新或整合过程，在一定程度上也支持 P2 成分的解释，即这种信息更新或整合过程较多涉及到先前行为选择及其收益，其过程越复杂繁琐，越容易引起更多的冲突。综上所述，我们认为当前研究中个体做出不信任选择时与内化的强制规范(信任他人)存在较强烈的冲突，因此 P2 反映了决策过程中信息整合引起的冲突检测加工。*请审稿人审查。

意见 12：本研究的不足分析很勉强，为了分析不足找不足。而且也不是本研究无法克服的，如果说是不足，就是没有去做。情绪评定，难道会干扰本研究的脑电波吗？

回应：非常感谢审稿人对初稿中研究不足部分所提的真挚宝贵的意见。经过再次对当前研究框架的分析，我们对此部分内容进行了重新整理与修改。详见修改稿最后一段内绿色内容：*其次，由于当前研究中结果强化率为几率水平(50%)，脑电被试可能无法有效地从互动过程中获得对方声誉信息，或者依据结果反馈调整随后的行为决策，因此当前研究无法有效的探究真实人际互动中信任的强化学习过程或者声誉形成过程，及其对决策与反馈阶段中脑电成分的影响。*请审稿人审查。

审稿人 3 意见：

意见 1：引言部分缺乏对自己的实验设计、操纵变量的描述。

回应：非常感谢审稿人对实验假设引入所提的宝贵意见。我们已经据此对初稿引言部分进行了重新修改和补充，详尽描述了相关研究假设的理论基础与先前研究结果，具体说明如下：

对于博弈决策阶段，目前关于“个体为何选择信任或不信任，以及两种选项的优势性问题”的理解存在两种相悖的理论观点。背叛厌恶理论认为信任涉及到策略不确定性，为回避背叛风险个体倾向于选择不信任行为(Bohnet, Greig, Hermann, & Zeckhauser, 2008)；而强制规范理论则强调信任是受道德规范制约的行为，个体为了显示对他人品质的尊重，更倾向于选择信任以维持他人值得信赖的社会矫饰(Dunning, Anderson, Schlösser, Ehlebracht, & Fetchenhauer, 2014)。鉴于目前关于信任互动情境下博弈决策过程的ERP 研究较为缺乏，为了从神经电生理水平上探讨信任决策中两项选项的优势性问题，我们基于风险决策和侧抑制任务的ERP 研究(Gajewski, Stoerig, & Falkenstein, 2008; Gajewski, & Falkenstein, 2013)，提出了通过 P2 和 N2 两种成分检验上述问题的研究思路。鉴于研究者关于博弈决策中不同行为选择的优势性问题尚存争议，本研究提出假设一：如果决策博弈阶段信任选择为优势选项(与强制规范理论相一致)，则个体会更多的选择信任，而且不信任选择比信任选择诱发更大的P2 和N2 成分；反之，如果不信任选择为优势选项(与背叛厌恶理论相一致)，则个体会更多的选择不信任，同时信任选择比不信任选择诱发更大的P2 和N2 成分。

对于结果评价阶段，鉴于当前采用TG 范式探讨结果评价过程的脑电研究上不充足(Chen et al. 仅涉及结果评价中的FRN 成分)，而且同时考察TG 范式中博弈决策阶段与结果评价阶

段的脑电研究也十分匮乏，因此当前研究试图以 FRN 和P300 为指标，补充相关方面的研究数据。另外，依据结果评价的独立编码理论及相关研究结果，FRN 和P300 分别独立负责编码或加工结果的效价与大小(Sato et al., 2005; Yeung & Sanfey, 2004)，即 FRN 反映了一种基于结果的好坏或预期目标是否实现的二分法所完成的早期自动化评价过程，而 P300 则反映了一种基于动机/情感意义或者注意资源分配的晚期控制评价过程(Yu & Sun, 2013)；并且已有研究发现 P300 对金额大小的数量加工存在参照点效应，即以赌注为基准的相对数量加工(Xiang, Wang, Zhang, & Yuan, 2008)。因此，我们对 FRN 和P300 的研究假设进行了调整，即研究假设二：损失反馈比获益反馈诱发更负的 FRN，损失反馈与获益反馈诱发的 P300 波幅无显著差异。

引言部分已按照上述思路进行修改，详见文中引言部分绿色字体，请审稿人审查。

意见 2：方法部分中作者告知脑电被试“每一回合的互动对象均是同一个体”，之后又需要保证“信任选择后正性反馈的比例为 50%”，又说是“400 名代表性被试”，到底怎么做的？被试是如何理解实验程序的？如果真是理解成同一个对家，对于被试来说，同一名对家的反应策略为何会在不同试次间变化？另外，为什么是“信任后的正性反馈”？从图上看，你有三种反馈，又是如何分布的？你后面的 FRN 处理怎么考虑这些分布的？

回应：非常感谢审稿人对方法部分所提的宝贵意见。我们已经据此对文稿的内容进行了逐一的修改和补充，具体说明如下：

1) 初稿对实验情境的描述过于简洁，略显模糊。现对实验指导语所述内容进行补充说明，以期提高实验情境的真实性与有效性：“*受托者的决策策略是事先采集的，具体而言，脑电被试被告知脑电实验前实验者从社会上选择了 400 名具有社会代表性的成人被试，要求他们扮演受托者与另一名实验助手在网络上完成 30 回合的重复性 TG 任务，然后对其在这些回合中的行为选择进行分析与学习，采用计算机程序模拟这 400 名被试的反应策略，并贮于计算机当中；在脑电被试进行信任博弈时，计算机会从中随机选择一个被试的反应策略以完成所有回合的游戏。*”详见文中任务与程序部分绿色字体。需要强调的是，重复复互动过程非常复杂多变, 难以在一个实验中探索明晰, 需要系列实验研究才可探讨. 本实验操纵是一种预先固定式的重复信任操纵, 不存在重复互动过程. 实际任务中采用预先编制的、结果强化率为 50%的电脑程序来操纵受托者的行为，因此脑电被试可能无法有效地从互动过程中获得对方策略，或者依据结果反馈调整随后的行为决策。这种缺陷是实验室实验与真实社会互动过程之间必然存在的差距，因此，作为探讨信任博弈互动的初步研究，当前研究无法探究重复互动过程中的强化学习过程，在此对这种不足进行说明。

2) “信任后的正性反馈”是指脑电被试选择信任对方后，程序选择平均分配进而使得被试获得 20 点的反馈。实验流程图中确实存在三种反馈，但三种反馈是以被试的信任选择为基础而分布。对于被试选择信任后，电脑程序以 50%的正性结果反馈强化操纵而言，我们是这样设置的：在实验前对整个信任博弈程序的反馈进行编码，具体为 150 试次中损失反馈和收益反馈各 75 次，然后进行随机排列。在某个试次中，当被试选择信任时则依据反馈编码给以损失或收益反馈。基于反馈编码（损失或反馈）是随机排列的，因此理论上可以实现被试选择信任时损失反馈与收益反馈各半的情况，而且当前研究结果中损失反馈(48±7)与收益反馈(49±8)的叠加次数基本一致，也进一步表明这种反馈设置的有效性。

请审稿人审查。

意见 3：结果部分中作者关注决策阶段的脑电反应，却只进行了刺激锁定的 ERP 反应(P2/N2)，而没有进行反应锁定的 ERP 分析；如果进行反应锁定的 ERP 分析，即把被试按键的时间点作为叠加平均的起始点，考察按键前后脑电成分的差异，结果是怎样的？文中的潜伏期是否

指峰值潜伏期？作者需要报告各个条件最终用于叠加平均的试次数均值和标准差。另外，对反馈部分的脑电处理根本就没有考虑决策者当初的“信任”、“不信任”的决定，不可接受。如果考虑了，不同条件下的 trial 如何？数据模式又如何？最后，“损失”和“收益”无法体现信任博弈中对家是否背叛被试的信任，可以考虑重新命名，以更好地反映作者所操纵的心理变量。

回应：非常感谢审稿人对结果部分所提的宝贵意见。我们已经据此对文稿的内容进行了逐一的修改和补充，具体说明如下：

1） 在决策阶段，被试的行为反应确实具有一定的随机性，反应时存在一定的差异性，采用反应锁定方式叠加脑电活动也是一种比较合适与常用的方法(如侧抑制任务或 go/no-go 任务)。但是，由于被试做出反应后决策过程就已终止，以反应锁定叠加脑电活动可能无法完整的刻画整个决策过程，而且我们参考了两篇关于反应性攻击范式的论文(Taylor Aggression Paradigm, TAP;该范式与 TG 博弈相似，包含决策阶段和反馈阶段两个部分)，两者均采用刺激锁定方法考察了决策阶段的脑电成分，因此，我们认为采用刺激锁定的叠加方式是有效的和可取的。当然，为了从其他侧面更精确的刻画个体在决策阶段的大脑加工模式，我们在今后研究中也会努力尝试进行反应锁定的脑电分析，以期发展并丰富当前研究的结果。同时，文中已补充不同条件下叠加次数的均值与标准差。

2） 如前所述，本实验操纵是一种预先固定式的重复信任操纵, 不存在重复互动过程. 实际任务中采用预先编制的、结果强化率为 50%的电脑程序来操纵受托者的行为，因此脑电被试可能无法有效地从互动过程中获得对方策略，或者依据结果反馈调整随后的行为决策。为了检验当前互动过程中学习过程可能产生试次间的影响，例如前一个试次的损伤反馈可能导致下次决策的不信任，我们对不同反馈(选择信任后的获益反馈<20>、不信任后的反馈<10>、选择信任后的损失反馈<0>)后的信任选择率进行了单因素重复测量方差分析，结果表明不同反馈后的信任率不存在显著差异，$F(2,38)=2.78$，$p=0.076$，$\eta^2=0.13$[不同反馈后的信任率分别为：获益=76.6±3.5%; 中性=64.4±3.2%; 损失=64.6±4.6%]；单独对获益或损失反馈后信任率的配对 t 检验结果也没有发现两者存在显著差异，$t(1,19)=1.93$, $p=0.068$。另外，依据已有研究(Bai et al., 2014; Gehring & Willoughby, 2002)，我们也尝试了不同反馈(获益/损失)后信任率与相对应反馈诱发的 FRN/P300 波幅之间的相关，结果表明：损失反馈后的信任率与损失反馈诱发的 FRN(r=-0.16, p=0.51)和 P300(r=-0.11, p=0.65)相关不显著；获益反馈后的信任率与获益反馈诱发的 FRN(r=-0.16, p=0.50)和 P300(r=-0.36, p=0.12)相关不显著。简言之，这些结果表明当前实验当中可能存在较弱的强化学习过程，并不存在试次间的互动影响。

3） 关于反馈条件(损失 vs. 收益)的命名，鉴于当前研究未设置纯粹金钱得失的控制条件，因此我们认为仍采用这种命名较为妥当，详见修改稿末段当前研究不足部分。

请审稿人审查。

意见 4：讨论部分中，“互惠偏好”跟你的研究有什么关系？你说了一堆无关的话！另外，你的数据分析不可靠，关于 FRN 的讨论也没有什么意义。

回应：非常感谢审稿人对讨论部分所提的宝贵意见。我们已经据此对文稿的内容进行了逐一的修改和补充，具体说明如下：

1） 我们对初稿进行了大幅度的修改与调整，修改稿中讨论 1 部分试图从强制规范理论角度来解释当前的行为结果(为什么信任选择率显著高于几率水平)。详见修改稿讨论 1 部分的内容：*信任对于文明社会与人际关系的建立与发展是十分重要的，然而目前关于“个体为何选择信任或不信任，以及两种选项的优势性问题”的理解尚存在较大争议：背叛厌恶理论认为信任涉及到策略不确定性，为回避背叛风险个体倾向于选择不信任行为；而强制规范理论则强调信任是受道德规范制约的行为，个体为了显示对他人品质的尊重，更倾向于选择信任以*

维持他人值得信赖的社会矫饰。考虑到当前研究中结果强化率仅为 50%，实验中被试会体验到较高强度的背叛水平，有人可能会预期信任选择率可能会随着互动过程的发展而逐渐降低，总体上也应接近或低于几率水平。但是，研究结果却发现前后互动阶段中信任率没有显著变化，而且总的信任率更是显著高于几率水平，因此当前研究结果与强制规范理论相一致(Dunning et al., 2014；Dunning, Fetchenhauer, & Schlösser, 2012)。换言之，信任博弈中的信任行为在某种程度上是一种表达性行为，个体更在意信任行为本身及其所代表的某种心理意义；人们选择信任是为了遵循一种强制规范，即个体借此向他人的品格表示尊重，维持其关于他人值得信赖与诚信善意的社会矫饰，即使个体私下并不相信如此。

2） 如前所述，本实验操纵是一种预先固定式的重复信任操纵, 不存在重复互动过程. 实际任务中采用预先编制的、结果强化率为 50%的电脑程序来操纵受托者的行为，因此脑电被试可能无法有效地从互动过程中获得对方策略，或者依据结果反馈调整随后的行为决策。因此，当前研究中结果评价阶段个体对不同反馈(损失 vs.收益)的加工过程是可信且可靠的。

请审稿人审查。。

审稿人 4 意见：

意见 1：作者选择 P2 和 N2 成分的理由不够成分。从作者的描述中，看不成为何风险决策和抑制任务的研究结果能提供证据，使作者关注 P2 和 N2 两种成分，而不是其他成分。“对于决策阶段而言，基于风险决策和抑制任务的研究结果，我们主要关注 P2 和 N2 两种成分”。

回应：非常感谢审稿人对实验假设引入所提的宝贵意见。我们已经据此对初稿引言部分进行了重新修改和补充，详尽描述了相关研究假设的理论基础与先前研究结果，具体说明如下：*对于博弈决策阶段，目前关于“个体为何选择信任或不信任，以及两种选项的优势性问题”的理解存在两种相悖的理论观点。背叛厌恶理论认为信任涉及到策略不确定性，为回避背叛风险个体倾向于选择不信任行为(Bohnet, Greig, Hermann, & Zeckhauser, 2008)；而强制规范理论则强调信任是受道德规范制约的行为，个体为了显示对他人品质的尊重，更倾向于选择信任以维持他人值得信赖的社会矫饰(Dunning, Anderson, Schlösser, Ehlebracht, & Fetchenhauer, 2014)。鉴于目前关于信任互动情境下博弈决策过程的 ERP 研究较为缺乏，为了从神经电生理水平上探讨信任决策中两项选项的优势性问题，我们基于风险决策和侧抑制任务的 ERP 研究(Gajewski, Stoerig, & Falkenstein, 2008; Gajewski, & Falkenstein, 2013)，提出了通过 P2 和 N2 两种成分检验上述问题的研究思路。*请审稿人审查。

意见 2：作者如何得出两个研究假设？缺乏合理的理论推导。“本研究假设一：决策阶段中不信任选择比信任选择诱发更大的 P2 和 N2 成分。对于反馈阶段而言，依据已有研究及结果评价的两阶段理论，FRN 反映了反馈结果与事先预期之间的偏离程度，P300 对反馈结果的效价和金额大小敏感，本研究假设二：损失反馈比获益反馈诱发更负的 FRN，获益反馈比损失反馈诱发更正的 P300”。并且，作者未说明为何要研究获益和损失的不同。

回应：非常感谢审稿人对实验假设引入所提的宝贵意见。我们已经据此对初稿引言部分进行了重新修改和补充，详尽描述了相关研究假设的理论基础与先前研究结果，具体说明如下：

对于博弈决策阶段，目前关于“个体为何选择信任或不信任，以及两种选项的优势性问题”的理解存在两种相悖的理论观点。背叛厌恶理论认为信任涉及到策略不确定性，为回避背叛风险个体倾向于选择不信任行为(Bohnet, Greig, Hermann, & Zeckhauser, 2008)；而强制规范理论则强调信任是受道德规范制约的行为，个体为了显示对他人品质的尊重，更倾向于选择信任以维持他人值得信赖的社会矫饰(Dunning, Anderson, Schlösser, Ehlebracht, & Fetchenhauer, 2014)。鉴于目前关于信任互动情境下博弈决策过程的 ERP 研究较为缺乏，为

了从神经电生理水平上探讨信任决策中两项选项的优势性问题，我们基于风险决策和侧抑制任务的ERP 研究(Gajewski, Stoerig, & Falkenstein, 2008; Gajewski, & Falkenstein, 2013)，提出了通过 P2 和 N2 两种成分检验上述问题的研究思路。鉴于研究者关于博弈决策中不同行为选择的优势性问题尚存争议，本研究提出假设一：如果决策博弈阶段信任选择为优势选项(与强制规范理论相一致)，则个体会更多的选择信任，而且不信任选择比信任选择诱发更大的P2 和N2 成分；反之，如果不信任选择为优势选项(与背叛厌恶理论相一致)，则个体会更多的选择不信任，同时信任选择比不信任选择诱发更大的P2 和N2 成分。

对于结果评价阶段，鉴于当前采用TG 范式探讨结果评价过程的脑电研究上不充足(Chen et al. 仅涉及结果评价中的FRN 成分)，而且同时考察TG 范式中博弈决策阶段与结果评价阶段的脑电研究也十分匮乏，因此当前研究试图以FRN 和P300 为指标，补充相关方面的研究数据。另外，依据结果评价的独立编码理论及相关研究结果，FRN 和P300 分别独立负责编码或加工结果的效价与大小(Sato et al., 2005; Yeung & Sanfey, 2004)，即FRN 反映了一种基于结果的好坏或预期目标是否实现的二分法所完成的早期自动化评价过程，而P300 则反映了一种基于动机/情感意义或者注意资源分配的晚期控制评价过程(Yu & Sun, 2013)；并且已有研究发现 P300 对金额大小的数量加工存在参照点效应，即以赌注为基准的相对数量加工(Xiang, Wang, Zhang, & Yuan, 2008)。因此，我们对FRN 和P300 的研究假设进行了调整，即研究假设二：损失反馈比获益反馈诱发更负的FRN，损失反馈与获益反馈诱发的P300 波幅无显著差异。

引言部分已按照上述思路进行修改，详见文中引言部分绿色字体，请审稿人审查。

意见 3：既然作者在前沿部分已经说明有前人研究决策反馈阶段，为何作者还要研究这个阶段，其目的何在？请作者在引言部分说明。

回应：非常感谢审稿人所提的宝贵意见。对于结果评价阶段，鉴于当前采用 TG 范式探讨结果评价过程的脑电研究上不充足(Chen et al. 仅涉及结果评价中的 FRN 成分)，而且同时考察 TG 范式中博弈决策阶段与结果评价阶段的脑电研究也十分匮乏，因此当前研究试图以 FRN 和 P300 为指标，补充相关方面的研究数据。请审稿人审查。

意见 4：任务和程序部分，150 回合的游戏，被试都只面临同样的点数进行决策么？如果是这样的话，如何保证被试做出的反应是有效的？

回应：非常感谢审稿人所提的细致入微的意见。为了检验个体的行为选择是否属于随机按键，我们将被试的信任选择率与随机水平(50%)进行了单样本 t 检验，结果发现被试的信任率(69.8±2.1%)显著高于随机水平，$t(19)=9.46$，$p<0.001$。同时，我们将所有试次分为前后各 75 试次，以分析前后两段中被试信任率与随机水平(50%)之间的差异。单样本 t 检验表明，前 75 试次的信任率(72.3±2.3%)显著高于随机水平，$t(19)=9.67$，$p<0.001$；后 75 试次的信任率(67.4±2.7%)显著高于随机水平，$t(19)=6.31$，$p<0.001$。因此，上述结果表明实验中被试的按键远超过随机按键水平，其按键确实代表其真实意图与意愿。请审稿人审查。

意见 5：结果部分，为何对 P2 和 N2 两种成分的分析考虑了电击左右分布这个因素，而对 FRN 和 P300 的分析不考虑该因素？

回应：非常感谢审稿人对结果部分数据分析方法所提的宝贵意见。对于决策阶段而言，由于目前探讨信任决策阶段的 ERP 研究几乎没有，可借鉴性文章较少，因此我们选择前后分布(Fz、FCz、Cz、CPz、Pz)和左右分布(左(3)、中(z)、右(4))以试图更为全面的了解决策阶段 ERP 数据的特性。对于反馈阶段而言，电极点的选择主要依据已有文献表明脑中线 5 电极点上 FRN 和 P300 成分最大进行确定的(Gehring & Willoughby, 2002; Leng & Zhou, 2010;

Yeung et al., 2004)。

意见 6：讨论部分，作者认为 P2 和 N2 成分与冲突检测相关。一般来说，冲突大小可由反应时长短表现出来，但是信任与不信任的反应时并无显著差异。作者对脑电结果的解释均缺乏有力的数据支持。

回应：非常感谢审稿人对讨论部分所提的宝贵建议。两种选择反应时没有显著差异可能是由于一下几方面造成的：首先，重复性社会互动中随着互动次数增加，被试可能会产生习惯化或采用固定策略进行反应。我们将所有试次进行了前后均分并比较了前后两个阶段中不同选择的反应时，结果发现前部互动阶段中信任选择的反应时(533±22ms)显著大于不信任选择(505±25ms)，而后部互动阶段中信任选择(484±24ms)与不信任选择的反应时差异不显著(494±27ms)，这种结果支持上述假定。其次，为将被试行为选择的动作准备效应最小化，以及降低其反应时的变异性，当前研究的指导语要求“被试看到决策选项图之后再进行反应”，这种操纵可能会导致两种选择的反应时没有显著差异(Boudreau, McCubbins, & Coulson, 2009)。最后，需要说明的是，已有 ERP 研究中也存在行为结果与脑电成分不一致的情况，这种相分离的结果可能是因为脑电指标比行为结果更敏感，或者脑电成分只反映信息加工的某个阶段，而行为结果则是整个信息加工过程的最终输出结果，进而导致了两者无法一一对应(Kounios & Holcomb, 1992; Wu & Zhou, 2012; Wang, Huang, Zhang, Song, & Bai, 2014)。综上所述，我们认为当前研究中反应时差异可能是由于实验指导语和重复性互动的特性造成的，前半程互动中不同选择的反应时差异则部分程度上支持了研究结果的有效性，并不影响脑电结果的解释。请审稿人审查。

意见 7：“研究者认为 mPFC 和 rTPJ 主要参与心理推理加工，如推测对方的信念、意图或互动策略，而 dlPFC 的激活则促使个体克服短期自私利益的诱惑，进而选择互惠利他的合作行为。”这句话缺乏参考文献支持。

回应：非常感谢审稿人的提醒，已补充此句的参考文献支持(Declerck, Boone, & Emonds, 2013)。请审稿人审查。

意见 8：缺乏引用相关文献：“Boudreau, C., McCubbins, M. D., & Coulson, S. (2008). Knowing when to trust others: An ERP study of decision making after receiving information from unknown people. Social cognitive and affective neuroscience, nsn034.”。

回应：非常感谢审稿人的提醒，修改稿中已添加了此处文献，请审稿人审查。

第二轮

审稿人 1 意见：

意见 1： 详细意见：

1、关于假设的论述与修改，很好，具有逻辑性。 本研究提出假设一：如果决策博弈阶段信任选择为优势选项(与强制规范理论相一致)，则个体会更多的选择信任，而且不信任选择比信任选择诱发更大的 P2 和 N2 成分；反之，如果不信任选择为优势选项(与背叛厌恶理论相一致)，则个体会更多的选择不信任，同时信任选择比不信任选择诱发更大的 P2 和 N2 成分。

2、引言部分的其他内容进行了修改，说服力有很大提高。

3、指导语的修改很好，这个修改是为了提高情境的真实性与有效性。疑问是：指导语的修改是为了实验过程中被试理解任务等内容。现在的修改，不能改变已经做完的实验。略有困

惑。
4、有关数据分析可能存在问题的分析与回答，非常认真，全面，符合逻辑。
5、ERP 数据分析的补充很全面。
6、讨论的修改很好！并且对不同波形的意义做了认真说明。
7、论文题目也做了认真修改。摘要的修改也有提高。也补充了文献。
8、论文不足的修改很好。
总之，四位审稿人的意见都认真的进行了研究，论文相应部分做了认真修改。论文质量有很大提升。
回应：非常感谢审稿人对本稿件所做的工作。

对于第三个问题而言。实际实验开始之前，为了帮助被试更好地理解并完成实验任务，我们是采用了较为详尽的语言来描述实验情境(即现修改稿中的指导语)，但在初稿的指导语撰写时则犯了过分简洁与缩略的问题，忽视了实验情境操纵的表达准确性. 现在修改稿中的指导语是对初稿中简略指导语的补充与说明，是被试参与实验前确实知晓的相关信息。请审稿人审查。

审稿人 2 意见：
意见 1： 2 幅插图均为彩图，如果用灰度图发表，请注意换用不同的线条（波形图）和灰度地形图。
回应：非常感谢审稿人的细心提醒，我们重新制作了图 2 和图 3，波形图换用了不同的线条,灰度地形图将在印刷版中使用,彩色地形图在电子版中使用,请审稿人审查。

意见 2：篇幅过长，建议适度压缩。
回应：非常感谢审稿人的宝贵意见，文章篇幅已进行了适度缩减，请审稿人审查。