《心理学报》审稿意见与作者回应

题目：机器人与人类来源的多轮正负反馈对信任与合作的动态影响
作者：宣泓舟，朱珈莹，赖乾恩，何贵兵

第一轮

审稿人 1 意见：

本研究围绕“机器人与人类来源反馈对协作信任动态演变的差异性”展开，选题新颖，实验设计规范，讨论深入，整体结构完整，语言表达清晰。尤其是从动态视角揭示了人机信任与人际信任的演变差异，具备一定的理论贡献和应用价值。部分内容仍存在改进空间，以下提出具体问题和建议。

意见 1：该研究在理论上具有一定价值，探讨了被试在时间维度上对人类与机器人队友信任的差异，对于揭示人机信任动态演变规律有一定的贡献。在应用层面，作者提出可利用机器人传递负面信息，具备一定的现实参考意义。然而，被试的反应在一定程度上也可能体现为对队友语言反馈的情绪响应，但作者只从信任视角加以解读。因此，建议作者对理论框架与应用价值部分进行更清晰、严谨的梳理与界定。

回应：非常感谢审稿专家对本文的肯定以及提出的建设性意见。您关于情绪响应维度的意见极具启发性，为我们更全面地审视本研究中的反馈效应提供了重要的新视角。我们根据您的建议，对理论框架和应用价值部分进行了认真修订与完善。

在理论框架层面，我们在前言部分（1.2-1.3 节）进一步阐明了本研究聚焦反馈与信任的理论依据。同时，我们在 1.6 节补充讨论了反馈对情绪影响的相关文献，以完善理论框架的完整性。

具体修改内容如下：“信任对于团队效能和团队凝聚力至关重要(DeRosa et al., 2004; Jones & George, 1998; Marks et al., 2001; Nicholson et al., 2001)，它不仅存在于人际之间(Mayer et al., 1995; Ozaras & Abaan, 2018)，在人机互动中也同样关键(Gkinko & Elbanna, 2023; Komiak & Benbasat, 2006; Kopp et al., 2023; Vance et al., 2008; 齐玥 等, 2024)”；

“正性反馈指赞扬性表达，主要涉及对他人贡献的认可或钦佩，通常引发积极情绪，因此更容易被接受(Higgins et al., 2001)。相比之下，负性反馈是一种可能威胁自尊的表达方式，尽管有助于规范个体行为，但可能让接收者感到不适或尴尬(Malle et al., 2014)，并引发强烈的消极情绪和负面反应(Weiss et al., 1999)。因此，正负性反馈对信任与合作可能存在不同的影响”

此外，结合专家意见，我们在讨论部分（5.1 节）新增了对情绪潜在作用的讨论。具体修改内容如下：“此外，本研究在不同反馈来源和效价条件下观察到的信任动态差异也可能部分源于被试对反馈的情绪响应差异。机器人反馈相比于同样内容的人类反馈，通常被认为“情绪性”较低，进而使被试能够更理性地对待反馈信息；而人类反馈可能立即触发较大的情绪反应（如羞愧、愤怒等），进而影响信任与合作。这种“反馈→情绪→信任/合作意愿”的影响路径值得未来研究进一步探索”

我们在研究局限部分（5.4 节）同样进行了补充说明。具体修改内容如下：“最后，本研究主要从信任角度探索被试对反馈的反应模式，未直接测量可能的机制性变量，如情绪、归因、感知能动性、反馈接受度等。因此，无法通过中介分析直接验证潜在机制。这一局限源

于研究设计时的权衡：（1）作为率先在人机场景下探讨多轮次反馈与动态信任关系的研究，我们选择优先聚焦于现象揭示，以便为后续机制研究奠定实证基础；（2）本研究需要进行多时点反馈操作以及对动态信任的重复测量，被试的认知负荷较重，若增加对机制性变量的多次测量，可能损害实验过程的流畅性，并且多次测量的内容本身可能带来干扰效应。因此，未来研究应设计出适用于更多变量动态测量的实验范式，以探明多轮次反馈与动态信任之间的潜在心理机制”

在应用价值层面，我们对实践建议的表述也进行了相应调整（5.3 节），强调组织应从反馈设计、情绪调节、认知训练等多角度采用综合手段增进人机信任和协作意愿。

意见 2：在实验设计部分，被试对自身和队友的任务表现及整体任务信息几乎不了解，建议作者进一步阐明这一设计的原因。

回应：感谢审稿专家对实验设计提出的非常专业的问题。我们就此进行了解释说明，并在修订稿的 3.2.3 实验材料部分对采用任务不确定性设计的原因进行了补充。

本实验的任务设计改编自广泛使用的 TNO 信任范式，前人在应用该范式时采用了类似的设计思路，即适度增加个体任务表现的不确定性来创造适宜的信任测量情境(De Visser et al., 2016, 2020; Kulms & Kopp, 2019; van Dongen & van Maanen, 2013)。根据信任的脆弱性理论(Mayer et al., 1995)，信任的本质是在不确定性条件下对他人的积极预期和依赖意愿。人们在适度不确定性的信息环境下，才能更好体现出信任的作用，因此不确定环境也是有效测量信任的必要条件。

具体修改内容如下（见正文第 10 页）：

“本研究有意限制了被试对自身和队友个人任务表现的了解，主要考虑是：不呈现个人具体绩效信息，而只呈现组队的排名信息以及队友间的评价反馈，有助于避免个体客观绩效数据可能产生的干扰效应，防止其掩盖反馈本身和反馈来源的作用，更好地观测较纯净的反馈效价和反馈来源的效应”

意见 3：文章将采纳队友建议的程度直接定义为行为信任，并在数据分析中沿用该表述，这种用法似乎并不严谨。尽管作者引用的 de Visser et al. (2016) 也提到该指标有时被视为行为信任，但该文献本身将该指标命名为 Compliance。

回应：非常感谢审稿专家提出这一重要的概念辨析问题。为能更好说明“行为信任”与“依从性”（Compliance）这两个概念的关系，我们进行了系统的文献梳理，现总结如下：

（1）在人机交互和团队协作研究领域，采纳他人建议的程度被广泛用作信任的客观行为测量指标(Kulms & Kopp, 2019; Logg et al., 2019; Mucha et al., 2021; Poursabzi-Sangdeh et al., 2021)。值得注意的是，Kulms 和 Kopp(2019)在其研究中明确将该指标命名为“行为信任”(Behavioral Trust)，这也是我们使用该术语的文献依据。

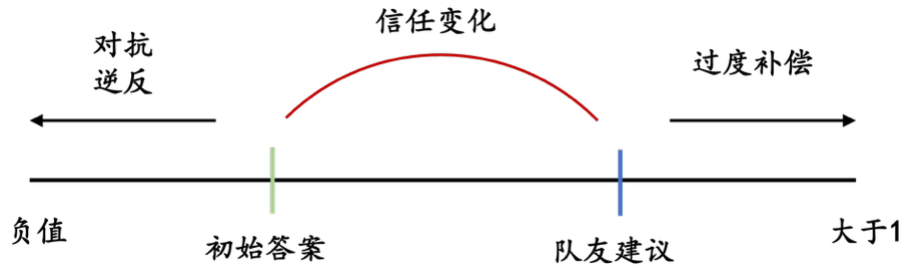
（2）正如您所指出的，de Visser 等人（2016）将该测量指标标注为“依从性”，但其研究的核心理论框架和讨论仍聚焦于“信任”。综合相关研究者的观点(Mohseni et al., 2021; van Dongen & van Maanen, 2013)，依从性本质上是信任程度的外显行为表现，即信任的操作化定义。因此，本研究使用“行为信任”这一术语是适当的，与 de Visser 的用词在语义上也无本质矛盾。

意见 4：行为信任的计算采用了绝对值处理，也就是说，即使个体的行为明显偏离了队友的建议（行为信任原始值为负值），计算结果也会被视为向队友建议方向的接近。此外，该指标被限定在 0 到 1 之间。建议作者进一步说明采取这种处理方式的理论依据与实际考虑。

回应：非常感谢审稿专家对行为信任指标计算方式提出的专业意见。我们已在 3.2.3 实验材

料部分对这一方法学问题做了进一步说明。具体修改内容可见修订稿第 11 页，我们用红色进行标注以便您审阅。

参考前人广泛采用的温泽化处理方法(Winsorization)，如果被试的最终答案超出初始答案和队友建议所构成的区间，则该试次数据将被舍弃(Lanz et al., 2024; Logg et al., 2019)。这种处理方式在信任和决策研究领域已获得广泛认可，其核心目的是减少极端异常值对结果的不当影响。如下文示意图所示，负值（即反向调整，背离队友建议）或大于 1 的值（即过度调整，超越队友建议）往往反映的是对抗、逆反、过度补偿等其他心理过程，而非本研究所关注的信任行为本身。但在本研究的实际操作中，未出现最终答案超出区间的情况。



从理论角度，行为信任的核心内涵是个体对建议来源的依赖程度。在协同决策情境中，信任的本质体现为被试在多大程度上接受并采纳队友的建议，即向队友建议靠拢的倾向。这种 0-1 标准化处理不仅与经典的 TNO 信任测量范式的逻辑保持一致(De Visser et al., 2016, 2020; Kulms & Kopp, 2019; van Dongen & van Maanen, 2013)，也更符合信任作为“依赖意愿”的操作化要求。

意见 5：建议补充一些描述性指标，如实验 1 在行为信任指标上报告了反馈来源的主效应是否显著，但是不同反馈来源的行为信任的均值和标准差等没有报告。

回应：感谢审稿专家的建议。我们已根据您的建议在修订稿中系统补充了各条件下的描述性统计指标。具体而言，在第 3.3 节和第 4.3 节中，我们补充报告了：（1）不同反馈来源条件下行为信任和未来合作意愿的均值和标准差；（2）各互动阶段（T0-T4）的描述性统计数据。

具体修改内容可见修订稿第 13 页-第 15 页、第 18 页，我们用红色进行标注以便您审阅。

意见 6：实验 1 和实验 2 的结果在一些指标上是否可以合并分析，如行为信任指标？目前只合并分析了“机器人来源的正负反馈对协作信任整体调整幅度的影响”。

回应：非常感谢审稿专家的宝贵建议。您提出的问题对于全面理解反馈来源和反馈效价的交互作用具有重要意义。我们在修订稿的第 4.3.6 节对实验 1 和实验 2 的数据进行了系统性合并分析，为本研究的核心发现提供了重要支持。具体结果如下：

为系统检验反馈来源和反馈效价对信任与合作动态演变的影响，本研究将实验 1 和实验 2 的数据进行合并分析，采用 2（反馈来源）×2（反馈效价）×5（互动阶段）的三因素混合方差分析，其中反馈来源和反馈效价为被试间变量，互动阶段为重复测量因素。

行为信任的分析结果显示，互动阶段的主效应显著($F(4, 432) = 6.100$, $p < 0.001$, $partial\ \eta^2 = 0.053$)，行为信任随互动进程显著下降；反馈来源的主效应不显著($F(1, 108) = 0.805$, $p = 0.371$)；反馈效价的主效应显著($F(1, 108) = 18.948$, $p < 0.001$, $partial\ \eta^2 = 0.149$)，负性反馈条件下的行为信任水平($M = 0.46$, $SD = 0.20$)显著低于正性反馈条件($M = 0.55$, $SD = 0.15$)。互动阶段与反馈来源的交互效应边缘显著($F(4, 432) = 2.361$, $p = 0.053$, $partial\ \eta^2 = 0.021$)，表明反馈来源（机器人 vs 人类）对行为信任的跨阶段变化模式产生了一定影响；互动阶段与反馈效价的交互效应显著($F(4, 432) = 3.235$, $p = 0.021$, $partial\ \eta^2 = 0.029$)，表明反馈效价（正性 vs

负性）显著影响了行为信任的跨阶段变化模式。然而，互动阶段、反馈来源以及反馈效价的三阶交互效应未达到显著水平($F(4, 432) = 0.779, p = 0.512$)。

未来合作意愿的分析结果显示，互动阶段的主效应显著($F(4, 432) = 31.712, p < 0.001, partial \eta^2 = 0.227$)，未来合作意愿随互动进程显著下降；反馈来源的主效应不显著($F(1, 108) = 2.607, p = 0.109$)；反馈效价的主效应显著($F(1, 108) = 38.182, p < 0.001, partial \eta^2 = 0.261$)，负性反馈条件下的未来合作意愿水平($M = 4.10, SD = 1.82$)显著低于正性反馈条件($M = 5.52, SD = 1.04$)。互动阶段与反馈来源的交互效应显著($F(4, 432) = 2.971, p = 0.037, partial \eta^2 = 0.027$)，说明反馈来源对未来合作意愿的跨阶段变化模式存在显著影响；互动阶段与反馈效价的交互效应显著($F(4, 432) = 40.774, p < 0.001, partial \eta^2 = 0.274$)，表明反馈效价对未来合作意愿的跨阶段变化模式产生了显著影响。互动阶段、反馈来源以及反馈效价的三阶交互效应未达到显著水平($F(4, 432) = 2.307, p = 0.083$)。

修改内容可见修订稿第 21 页，我们用红色进行标注以便您审阅。

意见 7：根据文章的描述，被试既无法获取自身的具体绩效数据，也无法获知队友的实际表现信息。换言之，绩效这一客观指标并未为被试提供评估队友能力或可靠性的有效线索。因此，文章中的行为信任指标是否仅仅反映了被试对队友的信任程度，尚需进一步讨论和明确。建议作者补充分析实验中行为信任指标的内涵。

回应：感谢审稿专家对行为信任指标内涵提出的专业性意见。我们已针对该问题在第 3.2.3 节进行了补充说明。

（一）关于反馈中对绩效作模糊处理的原因和依据。本实验的任务设计改编自广泛使用的 TNO 信任范式，前人在应用该范式时采用了类似的设计思路，即适度增加个体任务表现的不确定性来创造适宜的信任测量情境(De Visser et al., 2016, 2020; Kulms & Kopp, 2019; van Dongen & van Maanen, 2013)。根据信任的脆弱性理论(Mayer et al., 1995)，信任的本质是在不确定性条件下对他人的积极预期和依赖意愿。人们在适度不确定性的信息环境下，才能更好体现出信任的作用，因此不确定环境也是有效测量信任的必要条件。

具体修改内容如下（见正文第 10 页）：

“本研究有意限制了被试对自身和队友个人任务表现的了解，主要考虑是：不呈现个人具体绩效信息，而只呈现组队的排名信息以及队友间的评价反馈，有助于避免个体客观绩效数据可能产生的干扰效应，防止其掩盖反馈本身和反馈来源的作用，更好地观测较纯净的反馈效价和反馈来源的效应”

（二）关于行为信任指标的内涵。本研究中的行为信任指标主要反映在信息不确定条件下，被试愿意在多大程度上依赖队友的判断并据此调整自己的决策，即个体对队友的依赖程度和向其建议靠拢的倾向。在绩效信息模糊的情况下，队友的评价反馈传递了关于整体互动关系和合作态度的信息，而被试对这些信息的加工和响应正是信任动态演变的内在原因和行为表现。

修改内容可见修订稿第 10 页-第 11 页，我们用红色进行标注以便您审阅。

意见 8：建议作者进一步补充和明确实验过程的具体细节。例如，被试在团队任务中是否具有最终决策权？人类演员是通过何种方式向被试传递反馈信息，是否为直接的口头反馈？机器人是使用机械臂按键盘将建议值输入系统吗？此外，请说明实验中最终呈现给被试的个人排名结果设置情况。

回应：非常感谢审稿专家的建议。我们已根据您的建议在 3.2.4 节（实验流程部分）对实验过程进行了更为详尽的描述和澄清：

（1）关于决策权分配：在首轮任务开始前，被试通过一个虚假的抽签小游戏被指定为“团

队主导者”，拥有最终决策权。这一设计确保了被试对队友建议的采纳程度能够真实反映其信任水平。

（2）关于人类演员的反馈方式：人类演员严格按照预设脚本提供直接的口头反馈。所有演员在实验前均接受了统一培训，以确保反馈的语气、语调等表达方式在不同被试间保持一致性。

（3）关于机器人的操作方式：机器人 Pepper 在收到系统生成的随机建议值后，会模拟短暂的“思考”过程（通过细微的头部动作来增强真实感），然后以语音形式报告该数值。为增强实验的可信度，我们预先布置了数据线连接机器人和电脑，并告知被试机器人会通过该数据线将建议值传入系统。实际上，数据的传输由实验程序自动完成。

（4）关于排名结果的处理：实验中，仅在第一阶段是独立投资任务后呈现了虚假的团队排名和个人得分（“团队排名第 6；你超越了 68%的先前被试；你的队友超越了 77%的先前被试”）。“第 6”没有达到前五名获得额外奖金的标准，但非常接近，同时队友成绩设定为略好于被试，这一设置能够同时为后续的正性和负性反馈提供依据，让反馈更加真实。在第二阶段协同投资任务中，为防止绩效信息产生干扰效应，掩盖多轮次反馈的作用，我们未呈现个人排名和得分。此外，为避免最终排名结果可能对被试造成的负面心理影响，实验结束时也不会向被试呈现具体的个人排名。所有被试均被告知“排名需要综合所有参与者的数据后统一计算”，并获得全额实验报酬（40 元人民币）。

具体修改内容可见修订稿第 12 页-第 13 页，我们用红色进行标注以便您审阅。

意见 9：文章将行为信任与未来合作意愿作为协作信任的两个维度，但在结果分析中可以看出，实验设计对这两个变量的影响存在差异。而且，作者在某些部分将协作信任和行为信任混用。因此，建议作者进一步阐明这两个变量是如何共同构成协作信任的，以及为何可以将其整合为协作信任的综合指标。

回应：非常感谢审稿专家对行为信任与未来合作意愿这两个测量维度所提出的宝贵意见。您的意见和建议促使我们更深入地思考这两个变量的理论关系及其整合方式。我们对该问题的说明和修订如下：

协作信任(collaborative trust)是一个多维度构念，包含认知、情感和行为等多个成分(Lewis & Weigert, 1985; McAllister, 1995)，其至少包含行为信任（行为上依赖合作对象的倾向）和关系信任（建立和维持长期合作关系的意愿）这两个核心维度(Getha-Taylor et al., 2019; Kożuch & Sienkiewicz-Małyjurek, 2022; Lee et al., 2011; Whitehair & Berdanier, 2017)。我们原本希望通过多指标测量来捕捉协作信任的完整内涵：通过测量实际依赖行为来反映行为信任，并通过主观量表测量未来合作意愿以反映关系信任。

但是，协作信任与行为信任、未来合作意愿的关系似可类比为绩效与任务绩效、周边绩效的关系。员工激励制度对任务绩效和周边绩效的影响可能不同，如果直接将其概括为“员工激励制度对绩效的影响”确实存在不妥，容易造成误解。正如您所指出的，反馈对行为信任与未来合作意愿的影响也存在一定差异，我们将其概括为“反馈对协作信任的影响”也同样存在不妥。因此，为了便于读者理解，避免造成概念混淆并提高表述的准确性，我们进行了以下重要调整：

（1）将论文标题修改为《机器人与人类来源的多轮正负反馈对信任与合作的动态影响》，文中相关的表述也一并进行了修改。

（2）在结果呈现和讨论部分，我们明确区分“行为信任”和“未来合作意愿”，避免使用笼统的“协作信任”表述。

（3）在结果讨论部分（3.4 节）增加了对两个维度差异的深入讨论。例如，实验 1 中，行为信任和未来合作意愿虽然整体趋势一致，但在某些时间点的变化幅度和显著性存在细微

差异。我们指出，这种维度间的差异恰恰体现了协作关系的复杂性：行为层面的信任调整可能更为渐进和审慎，而态度层面的合作意愿变化可能更为敏感和快速。这种差异为我们理解信任的多维度特性提供了重要证据。

具体修改内容可见修订稿第 17 页，我们用红色进行标注以便您审阅。

意见 10：文章讨论部分认为研究将静态信任研究扩展到动态信任领域，但学界对于动态信任的研究也很多，建议重新修改；其次，文章在讨论部分认为研究突破了传统人机信任研究中“人-机-环境”的特征框架，但实验控制的还是与机器人相关的特征因素，并没有突破这一框架，建议修改这一论述。

回应：非常感谢审稿专家指出理论定位和贡献阐述中的表述不当之处。我们已对相关内容进行了审慎的修订：

（1）关于动态信任研究的定位。我们已将“将静态信任研究扩展到动态信任领域”这一欠妥的表述修改为“丰富了人机动态信任的研究”（5.3 节），并明确强调本研究的独特价值在于揭示了人机信任与人际信任的差异化动态演变模式。

（2）关于“人-机-环境”框架的表述。我们已将“突破了传统人机信任研究中‘人-机-环境’的特征框架”改为“重点关注强调互动过程在信任形成中的关键作用”（5.3 节），并明确说明本研究是在该框架基础上侧重动态交互过程。

具体修改内容可见修订稿第 24 页，我们用红色进行标注以便您审阅。

意见 11：文章中存在一些笔误，可再进行检查。具体如：摘要部分“而来自人类队友的多轮次负性反馈则则会导致协作信任呈现”中“则”字重复，参考文献引用格式“Lv et al, 2024”中缺少英文句点。

回应：非常感谢审稿专家细致的审阅工作。针对您指出的笔误问题，我们已对全文进行了仔细检查和修订，修正了摘要中的重复字词、参考文献引用格式等问题，并对全文内容进行了校对，确保论文的规范性。再次感谢您的宝贵意见。

……

审稿人 2 意见：

该论文研究围绕人机协作情境下“多轮次（动态）反馈”如何影响协作信任的演化展开，聚焦“反馈来源（机器人 vs. 人类） × 反馈效价（正性 vs. 负性） × 时间（多轮交互）”的动态过程，回应了当前人机信任研究中“静态快照”范式的不足。作者提出并命名了两类差异化动态模式：“滞后累积式”（机器人负性反馈）与“即时响应式”（人类负性反馈），并进一步揭示正负反馈的非对称性（负性反馈驱动的差异 > 正性反馈）。总体而言，选题契合“心理学前沿 + 智能时代人机互动”这一交叉方向，具备现实意义和一定理论潜力。然而，当前论文仍存在一些有待商榷的地方，供作者参考。其中，建议需要重点关注的地方：

意见 1：文献综述篇幅较长但各部分间缺乏有机整合：建议修改调整为：(1) 信任动态研究缺口；(2) 反馈互动作为动态触发机制；(3) 机器人来源反馈的特殊性（归因与意图性）；(4) 负性信息非对称的必要性假设；以形成更清晰递进的阐述逻辑。

回应：非常感谢审稿专家对文献综述结构提出的宝贵建议。我们已根据您的建议对前言部分进行了调整，形成了更清晰的递进逻辑：

1.3 节 人机信任动态的研究缺口：明确阐述现有研究的“快照”视角局限，建立从静态转向动态研究的必要性和理论价值。

1.4 节 反馈互动作为信任动态演变的触发机制：聚焦反馈互动这一具体的动态触发机

制，论述其在人机协作中的关键作用，并指出现有研究在多轮次反馈动态影响方面的不足。

1.5 节 机器人来源反馈的特殊性：通过“媒体不等同”效应和归因理论，推导假设：机器人与人类来源反馈影响下的信任与合作可能具有不同动态演变模式。

1.6 节 正负性反馈的非对称效应：基于正负反馈的差异和消极偏差理论，提出反馈效价的可能发挥的调节作用，并推导假设：在多轮次负性反馈情境下，机器人与人类来源反馈将对信任与合作产生差异化的动态影响；而在多轮次正性反馈情境下，不同反馈来源对信任与合作的动态影响差异较小。

具体修改内容可见修订稿第 5 页-第 7 页，我们用红色进行标注以便您审阅。

意见 2：理论机制的验证不足，作者提出了归因/意图性感知/媒体不等同性等机制，却缺乏直接测量或中介检验：主观感知测量（如归因维度：能力/意图/公平性/善意、感知中立性 vs. 人格化）中介/调节验证，例如：感知能动性是否调节首轮反馈冲击幅度？接受度是否中介负性反馈 → 信任轨迹斜率？

回应：衷心感谢审稿专家对理论机制验证提出的专业而深刻的建议。我们完全认同理论机制验证的重要性，本团队目前也正在其他研究中重点探索相关机制。针对本研究缺乏机制的直接测量或中介检验的问题，我们作如下说明：

本研究着重对多轮次反馈如何影响人机和人际信任动态演变特征进行探索，在实验设计时我们优先聚焦于现象揭示，当然，我们也曾考虑对反馈归因、情绪、意图感知、反馈接受度、感知能动性等可能的机制进行探究。然而，本实验设计涉及 5 个时间点的重复测量，每轮还包含 5 个试次的任务操作、反馈互动和合作意愿评估，整体实验时长接近 1 小时，被试的认知负担较大。如果再在每个时间点增加归因维度、感知能动性、感知中立性、反馈接受度等主观感知量表，将显著加重被试的认知负担，可能引发疲劳效应和响应偏差，还可能破坏人机互动的自然性，频繁的主观评估本身也可能构成干扰变量。因此权衡再三，我们选择先聚焦动态反馈和信任动态变化的现象揭示。

我们在研究设计的权衡时考虑的另一个因素是，本团队近期完成并已发表的一项单轮次反馈效应的研究中（Xuan & He, 2025, *Journal of Business Research*），已对部分理论机制进行了初步探索。我们发现，反馈来源对信任与合作存在显著影响，并且这种影响受到反馈接受度的中介作用和反馈效价的调节作用。具体而言，相较于人类来源的负性反馈，个体对机器人来源的负性反馈表现出更高的接受度，这进一步促成了对机器人队友更强的行为信任与未来合作意愿。这也是本研究暂未考虑深入探索机制的部分原因。

不过，您所提及的问题确实是本研究的欠缺之处，因此我们在研究不足部分（5.4 节）增加了相应内容，说明了缺乏理论机制验证的不足，并提出了相关未来研究方向，包括具体的测量变量（归因维度、感知能动性、感知公平性、反馈接受度等）、验证方法（中介/调节分析）以及整合生理指标测量（皮肤电、心率变异性等）的方案。具体修改内容可见修订稿第 26 页，我们用红色进行标注以便您审阅。

意见 3：研究设计与构想之间存在可以优化的结构性替代（例如为何不使用 2×2 的效价 × 来源的单一实验框架，以便直接比较交互效应）。

回应：非常感谢审稿专家对实验设计提出的意见和建议。您建议的确实是一个更为简洁和统计上更优的设计方案。我们完全理解您的考量，现就实际采用分离实验设计的原因如实说明如下：

本研究实际上经历了一个渐进探索的过程。基于本团队前期的单轮次反馈研究发现 (Xuan & He, 2025)，我们观察到相较于人类来源的负性反馈，个体对机器人来源的负性反馈表现出更高的接受度，进而导致对机器人队友更强的行为信任与未来合作意愿；然而，就正

性反馈而言，无论来源于机器人还是人类队友，并未导致个体反应上的显著差异。这一发现提示我们，负性反馈情境更可能触发反馈来源的差异效应。基于这一认识，我们在设计多轮次反馈研究时，优先独立开展了负性反馈条件的实验（即实验 1），希望能够集中资源和精力深入、细致地刻画机器人与人类来源的多轮次负性反馈影响下信任与合作的动态演变轨迹特征。

在实验 1 发现了“滞后累积式”与“即时响应式”两种截然不同的信任动态演变模式后，为了完善实验逻辑的完整性并检验反馈效价的调节作用，我们决定补充进行正性反馈条件的实验（即实验 2）。换言之，从时间顺序上，我们确实是先进行了负性反馈条件的实验，后补充了正性反馈条件的实验，而非一开始就设计为单一的 2×2 框架。这确实是当时考虑不周。

尽管实际数据收集采用了分离设计，但我们非常认同您关于直接比较交互效应的建议（实际上审稿人 1 也提出了类似建议）。因此，我们在修订稿中增加了实验 1 和实验 2 数据的合并分析（第 4.3.6 节），采用 2（反馈来源）×2（反馈效价）×5（互动阶段）的三因素混合方差分析，系统检验了各因素的主效应、两两交互效应和三阶交互效应。具体修改如下：

为系统检验反馈来源和反馈效价对信任与合作动态演变的影响，本研究将实验 1 和实验 2 的数据进行合并分析，采用 2（反馈来源）×2（反馈效价）×5（互动阶段）的三因素混合方差分析，其中反馈来源和反馈效价为被试间变量，互动阶段为重复测量因素。

行为信任的分析结果显示，互动阶段的主效应显著($F(4, 432) = 6.100$, $p < 0.001$, $partial\ \eta^2 = 0.053$)，行为信任随互动进程显著下降；反馈来源的主效应不显著($F(1, 108) = 0.805$, $p = 0.371$)；反馈效价的主效应显著($F(1, 108) = 18.948$, $p < 0.001$, $partial\ \eta^2 = 0.149$)，负性反馈条件下的行为信任水平($M = 0.46$, $SD = 0.20$)显著低于正性反馈条件($M = 0.55$, $SD = 0.15$)。互动阶段与反馈来源的交互效应边缘显著($F(4, 432) = 2.361$, $p = 0.053$, $partial\ \eta^2 = 0.021$)，表明反馈来源（机器人 vs 人类）对行为信任的动态演变模式产生了一定影响；互动阶段与反馈效价的交互效应显著($F(4, 432) = 3.235$, $p = 0.021$, $partial\ \eta^2 = 0.029$)，表明反馈效价（正性 vs 负性）显著影响了行为信任的动态演变模式。然而，互动阶段、反馈来源以及反馈效价的三阶交互效应未达到显著水平($F(4, 432) = 0.779$, $p = 0.512$)。

未来合作意愿的分析结果显示，互动阶段的主效应显著($F(4, 432) = 31.712$, $p < 0.001$, $partial\ \eta^2 = 0.227$)，未来合作意愿随互动进程显著下降；反馈来源的主效应不显著($F(1, 108) = 2.607$, $p = 0.109$)；反馈效价的主效应显著($F(1, 108) = 38.182$, $p < 0.001$, $partial\ \eta^2 = 0.261$)，负性反馈条件下的未来合作意愿水平($M = 4.10$, $SD = 1.82$)显著低于正性反馈条件($M = 5.52$, $SD = 1.04$)。互动阶段与反馈来源的交互效应显著($F(4, 432) = 2.971$, $p = 0.037$, $partial\ \eta^2 = 0.027$)，说明反馈来源对未来合作意愿的动态演变存在显著影响；互动阶段与反馈效价的交互效应显著($F(4, 432) = 40.774$, $p < 0.001$, $partial\ \eta^2 = 0.274$)，表明反馈效价对未来合作意愿的动态演变产生了显著影响。互动阶段、反馈来源以及反馈效价的三阶交互效应未达到显著水平($F(4, 432) = 2.307$, $p = 0.083$)。

这些合并分析结果为本研究的核心结论提供了更为稳健和全面的支持。

意见 4：行为信任指标构造及其分布特性、潜在混淆变量控制不充分。主观信任（未来合作意愿）量表：仅 2 个题项，需报 α 或 McDonald’s ω；如 α<0.70，应说明信度并考虑增加条目或进行潜变量模型。“(最终 − 初始)/(建议 − 初始)”的分母若接近 0（建议值与初始值距离很小）会产生不稳定比值；是否剔除阈值试次？是否检查偏态/离群？建议将“差异幅度（|建议 − 初始|）”作为协变量 / 分层因素，因为采纳概率随差值大小非线性变化（参考社会影响–从众函数）；是否可补充“采纳二元变量（是否跨过中点采纳）”的稳健性逻辑回归分析。同时，未控制或测量的个体差异变量（如：自动化一般信任倾向、反馈取向、负性情绪敏感性、技术熟悉度、人格维度）也可能解释部分反应差异。

回应：非常感谢审稿专家对测量指标和混淆变量控制提出的专业意见和建议。您的建议涉及多个重要的方法学问题，我们已根据您的意见进行了系统性的补充分析和完善，逐一回应如下：

（1）关于主观信任量表的信度问题，我们已在正文中补充报告了未来合作意愿量表的 Cronbach's α 系数，实验 1 为 0.952（见 3.2.3 节），实验 2 为 0.903（见 4.2.3 节），均达到良好信度标准，表明两个题项具有良好的内部一致性。

（2）针对差异幅度的问题，我们此前已经剔除了差异幅度为 0 的试次（共 27 个，占总试次 1.9%）。在修订稿中，我们进一步将差异幅度（|队友建议-初始答案|）作为协变量纳入重复测量方差分析中（见 3.3.1 节和 4.3.1 节），控制其影响后，实验结果与此前一致。具体修改内容可见修订稿第 13-14 页、第 19 页。

（3）关于采纳二元变量的稳健性检验。按照您的建议，我们构建了“是否跨过中点采纳队友建议”的二元因变量（建议采纳），采用广义线性混合效应模型（Generalized Linear Mixed Model, GLMM）进行分析，并将差异幅度纳入模型作为协变量。模型的固定效应包括反馈来源、互动阶段、差异幅度及其交互项，随机效应为被试 ID 的随机截距。采用最大似然估计法进行参数估计，并通过卡方检验评估各固定效应的显著性（第 3.3.3 节和第 4.3.3 节）。可见修订稿第 15-16 页、第 20 页，详细内容如下：

实验 1 的分析显示：反馈来源的主效应不显著($\chi^2(1) = 0.43$，$p = 0.512$)；互动阶段的主效应显著($\chi^2(4) = 102.14$，$p < 0.001$)；关键的是，反馈来源与互动阶段的交互效应显著($\chi^2(4) = 20.93$，$p < 0.001$)。描述性统计显示，人类队友组的建议采纳率从 T0 的 68.1%下降至 T1 的 46.0%，随后维持在较低水平（T2 至 T4 阶段分别为 38.8%、45.0%、39.9%），呈现急剧下降后趋于稳定的模式；而机器人队友组的采纳率从 T0 的 76.5%逐步下降至 T1、T2、T3、T4 阶段的 68.7%、49.6%、36.8%、31.2%，呈现渐进式下降趋势。这一结果与行为信任测量的模式高度一致，进一步证实了主分析结果的稳健性。

实验 2 的类似分析未发现显著的交互效应，与正性反馈条件下反馈来源无差异影响的主分析结论一致。

（4）关于个体差异变量的控制问题。首先，我们非常认同审稿专家关于个体差异变量可能对结果产生影响的观点。但是，由于潜在的个体变量实在太多，难以完全进行控制，并且过多测量可能干扰实验互动或暴露实验目的，本研究只能通过设置一定的被试招募条件（来自与心理学、机器人学无关的专业，并且此前都没有与机器人密切互动的经验）来尽量减小个体差异变量的影响。然而，我们也充分认识到，上述措施并不能完全排除个体差异的影响，因此，我们在“研究局限与未来方向”部分（见 5.4 节）补充讨论了这一局限性，并明确指出未来研究应深入探讨这些个体因素对机器人反馈效果的影响作用。具体可见修订稿第 25 页。

意见 5：统计分析停留在传统混合 ANOVA 水平，未充分体现“动态”分析的现代方法学优势；建议考虑增补线性/二次/分段混合效应模型（lme / nlme / lmer），比较 AIC / BIC，估计不同来源条件下的斜率差异与变化节点。

回应：非常感谢审稿专家提出的专业建议。您提出的混合效应模型分析确实体现了动态数据分析的现代方法学优势，能够更精细地刻画信任与合作的动态演变模式。我们已根据您的建议在传统混合 ANOVA 分析的基础上增补了混合效应模型分析，相关内容详见第 3.3.2 节和第 4.3.2 节。具体可见修订稿第 15 页、第 19-20 页，详细内容如下：

3.3.2 信任与合作的分阶段演变模式

为全面揭示信任与合作的动态演变模式，本研究采用混合效应模型进行统计分析。具体而言，我们构建并比较了三类模型：（1）线性模型，假设信任随时间线性变化；（2）二次模

型，允许信任呈现曲线型演变；（3）分段模型，根据方差分析结果和演变过程图将互动过程划分为前期（T0-T1）与后期（T2-T4）两个阶段，分别估计不同反馈来源条件下各阶段的变化斜率。所有模型均以反馈来源和互动阶段为预测变量，以行为信任和未来合作意愿为因变量，并且都包含被试层面的随机截距和随机斜率，以捕捉个体差异。模型拟合采用最大似然估计法，并通过赤池信息准则（AIC）和贝叶斯信息准则（BIC）进行模型比较(Burnham & Anderson, 2004)。所有分析使用 R 语言(Version 4.4.0)的 *lme 4* 包和 *lmerTest* 包完成。

行为信任的分析结果表明，分段混合效应模型表现最优(AIC = -181.33, BIC = -148.62)，相较于线性混合效应模型(AIC = -175.76, BIC = -146.69)和二次混合效应模型(AIC = -177.75, BIC = -141.40)有一定优势。未来合作意愿的分析结果表明，分段混合效应模型同样表现最优(AIC = 806.99, BIC = 839.70)，相较于线性混合效应模型(AIC = 834.31, BIC = 863.39)和二次混合效应模型(AIC = 809.14, BIC = 845.48)有较大提升。这表明将互动过程分为前后两个阶段能够更准确地刻画信任与合作的动态演变。

基于分段混合效应模型，我们进一步分析了不同反馈来源条件下信任与合作在多轮次负性反馈前期(T0-T1)与后期(T2-T4)的具体变化轨迹和斜率。在机器人来源反馈条件下，行为信任在前期未出现显著变化(β = -0.038，SE = 0.030，p = 0.209)，但在后期呈现显著的渐进式下降(β = -0.048，SE = 0.015，p = 0.001)。相比之下，在人类来源反馈条件下，行为信任在前期即出现显著的急剧下降(β = -0.111，SE = 0.030，p < 0.001)，而后期变化趋于平稳(β = 0.001，SE = 0.015，p = 0.940)。未来合作意愿的分析结果呈现出与行为信任高度一致的模式。在机器人来源反馈条件下，前期变化不显著(β = -0.499，SE = 0.175，p = 0.005)，但后期持续显著下降(β = -0.581，SE = 0.080，p < 0.001)。在人类来源反馈条件下，前期出现剧烈下降(β = -1.403，SE = 0.175，p < .001)，后期下降趋缓(β = -0.214，SE = 0.080，p = 0.009)。

4.3.2 信任与合作的分阶段演变模式

为进一步探索信任与合作在多轮次正性反馈下的动态演变模式，本研究同样采用了线性、二次以及分段混合效应模型进行统计分析，并通过 AIC 和 BIC 指标对比了这三类模型。所有模型均以反馈来源和互动阶段为预测变量，以行为信任和未来合作意愿为因变量，并且都包含被试层面的随机截距和随机斜率。所有分析使用 R 语言(Version 4.4.0)的 *lme 4* 包和 *lmerTest* 包完成。

行为信任的分析结果表明，线性混合效应模型表现最优(AIC = -303.05, BIC = -273.97)，相较于分段混合效应模型(AIC = -298.79, BIC = -266.08)和二次混合效应模型(AIC = -299.13, BIC = -262.78)均有一定优势。未来合作意愿的分析结果表明，线性混合效应模型(AIC = 638.16, BIC = 667.24)和二次混合效应模型(AIC = 635.09, BIC = 671.44)略优于分段混合效应模型(AIC = 635.57, BIC = 668.28)，但差异很小。根据简约原则，同样选择线性混合效应模型。

进一步检验线性混合效应模型的固定效应，对于行为信任指标，反馈阶段的主效应不显著(β = -0.006，SE = 0.006，p = 0.374)，反馈来源的主效应不显著(β = 0.061，SE = 0.032，p = 0.063)，反馈阶段和反馈来源的交互效应也不显著(β = -0.013，SE = 0.013，p = 0.336)。对于未来合作意愿指标，反馈阶段的主效应不显著(β = 0.020，SE = 0.034，p = 0.569)，反馈来源的主效应不显著(β = 0.382，SE = 0.247，p = 0.128)，反馈阶段和反馈来源的交互效应也不显著(β = -0.046，SE = 0.069，p = 0.501)。

意见 6：“两种模式”的命名与论断带有描述性解释过度色彩，缺少形式化检验；具体而言，“滞后累积式 vs 即时响应式”属于归纳标签，但尚未通过参数化模型（如随机斜率线性/非线性增长模型、分段线性模型、变化点分析、Growth Mixture Model）证明两类轨迹在统计学上存在结构性模式差异。

回应：非常感谢审稿专家的问题和建议。在修订稿的第 3.3.2 节和第 4.3.2 节我们补充了混合效应模型分析，为这两种模式的命名与论断提供了更精确的量化证据。具体可见修订稿第 15 页、第 19-20 页，详细内容如下：

在实验 1 中，模型比较结果显示，分段模型（将互动过程划分为前期 T0-T1 与后期 T2-T4）在拟合优度上显著优于线性模型和二次模型，表明信任与合作演变确实存在阶段性转折。更重要的是，斜率估计揭示了两种来源反馈影响的时间差异性：在机器人来源反馈条件下，行为信任在前期未出现显著变化(β = -0.038，SE = 0.030，p = 0.209)，但在后期呈现显著的渐进式下降(β = -0.048，SE = 0.015，p = 0.001)。相比之下，在人类来源反馈条件下，行为信任在前期即出现显著的急剧下降(β = -0.111，SE = 0.030，p < 0.001)，而后期变化趋于平稳(β = 0.001，SE = 0.015，p = 0.940)。未来合作意愿的分析结果呈现出与行为信任高度一致的模式。在机器人来源反馈条件下，前期变化不显著(β = -0.499，SE = 0.175，p = 0.005)，但后期持续显著下降(β = -0.581，SE = 0.080，p < 0.001)。在人类来源反馈条件下，前期出现剧烈下降(β = -1.403，SE = 0.175，p < .001)，后期下降趋缓(β = -0.214，SE = 0.080，p = 0.009)。

综合来看，人类来源反馈在前期阶段产生的影响强度是机器人来源的近 3 倍（行为信任：2.9 倍；未来合作意愿：2.8 倍），体现出高敏感性；而机器人来源反馈在后期阶段仍持续产生显著负面效应，体现出累积性特征，人类来源的影响则趋于饱和。我们认为，这一结果能够在一定程度上表明“滞后累积式”与“即时响应式”这两类轨迹在统计学上存在结构性模式差异。

意见 7：推论层面将“人机信任”与“人际信任”差异上升为“本质区别”仍显仓促。

回应：非常感谢审稿专家提出的宝贵意见。针对您的意见，我们对全文进行了系统性修订：将原文中“根本差异”、“本质区别”等强断言性表述修改为“显著差异”、“差异性特征”等更加审慎的表述；在多处增加了“初步”、“在特定条件下”、“在实证层面”等限定语；在讨论部分明确指出本研究观察到的差异反映的是特定情境下的不同表现特征，关于两者是否存在本质性区别还需要更多维度、更长时间跨度和更多样化情境的研究加以验证。修改后，本研究的核心贡献界定为：在实证层面揭示了人机信任与人际信任在多轮次负性反馈情境下的差异性演变特征，为理解两者关系提供了初步的经验证据。

意见 8：机制与操纵检验缺失：未测“反馈接受度”“归因类型”“感知意图/善意”“感知能动性”“被冒犯/受威胁感”。导致解释链条停留在假设层；建议回溯（若数据尚存且含有交互录像）进行双盲编码：反馈被感知强度、情绪反应、表情变化；或在修订版中补做附加在线研究验证归因机制。

回应：衷心感谢审稿专家对理论机制验证提出的专业而深刻的建议。我们完全认同理论机制验证的重要性，本团队目前也正在其他研究中重点探索相关机制。针对本研究缺乏机制的直接测量或中介检验的问题，我们作如下说明：

本研究着重对多轮次反馈如何影响人机和人际信任动态演变特征进行探索，在实验设计时我们优先聚焦于现象揭示，当然，我们也曾考虑对反馈归因、情绪、意图感知、反馈接受度、感知能动性等可能的机制进行探究。然而，本实验设计涉及 5 个时间点的重复测量，每轮还包含 5 个试次的任务操作、反馈互动和合作意愿评估，整体实验时长接近 1 小时，被试的认知负担较大。如果再在每个时间点增加归因维度、感知能动性、感知中立性、反馈接受度等主观感知量表，将显著加重被试的认知负担，可能引发疲劳效应和响应偏差，还可能破坏人机互动的自然性，频繁的主观评估本身也可能构成干扰变量。因此权衡再三，我们选择先聚焦动态反馈和信任动态变化的现象揭示。

我们在研究设计的权衡时考虑的另一个因素是，本团队近期完成并已发表的一项单轮次

反馈效应的研究中（Xuan & He, 2025, *Journal of Business Research*），已对部分理论机制进行了初步探索。我们发现，反馈来源对信任与合作存在显著影响，并且这种影响受到反馈接受度的中介作用和反馈效价的调节作用。具体而言，相较于人类来源的负性反馈，个体对机器人来源的负性反馈表现出更高的接受度，这进一步促成了对机器人队友更强的行为信任与未来合作意愿。这也是本研究暂未考虑深入探索机制的部分原因。

不过，您所提及的问题确实是本研究的欠缺之处，因此我们在研究不足部分（5.4 节）增加了相应内容，说明了缺乏理论机制验证的不足，并提出了系统的未来研究方向，包括具体的测量变量（归因维度、感知能动性、感知公平性、反馈接受度等）、验证方法（中介/调节分析）以及整合生理指标测量（皮肤电、心率变异性等）的方案。具体修改内容可见修订稿第 26 页，我们用红色进行标注。

意见 9：多重比较与效应量：简单效应多次 t 检验未见多重校正（Bonferroni / Holm / FDR），当前 p=0.05 附近的阈值解释需更谨慎。建议为所有主效应与交互提供 95% CI，并明确效应量（partial η^2 → 转换为 generalized η^2 或提供 ω^2 以增强跨研究可比性）。

回应：非常感谢审稿专家对统计细节的严格把关。我们已经根据您的建议对统计分析进行了全面修订，具体可见第 3.3.1 节和 4.3.1 节。首先，我们已对所有简单效应分析采用 Bonferroni 法进行多重比较校正，并修正了结果报告。同时，我们按照专家的要求为所有主效应与交互提供 95% CI，并补充报告了 generalized η^2。

具体修改内容可见修订稿第 13 页-第 14 页、第 18 页，我们用红色进行标注以便您审阅。

意见 10：样本量与功效：事前功效分析参数（效应量=0.25）依据何处先验？交互效应实际 partial η^2 报告为 0.19（疑似过大或书写错误：0.19 对应中上到大型），需核对。是否对正性反馈实验也独立进行功效评估？若假设正性反馈来源差异较小，实际功效可能不足。

回应：非常感谢审稿专家对功效分析的细致审查。针对您提出的问题，我们已作如下修改和补充分析：

（1）关于效应量先验依据。我们在修订稿第 3.2.1 节明确说明了事前功效分析是参考类似人机信任动态研究中所采用的功效参数设置(Roesler et al., 2024)，并且效应量 0.25 符合 Hancock 等人(2011)元分析所报告的领域平均效应量（实验分析 $\bar{d}$ = +0.67，中到大效应），该参数属于保守但现实的估计。具体可见修订稿第 9 页。

（2）关于效应量的问题。根据上一条意见的内容，我们已经将 partial η^2 统一转换为 generalized η^2，并进行了仔细核对。

（3）关于实验 2 的功效评估。您指出的问题非常关键。我们已按您的建议在第 4.2.1 节补充了系统的功效分析：敏感性分析显示，在 α=0.05、功效为 0.80 的条件下，当前样本量(N=56)能够检测到的最小效应量 f 为 0.148，属于小到中等效应。事后功效分析表明，本研究对中小效应量(f=0.15)的检验功效为 0.814。这表明本研究具有充分的敏感度检测中小程度的效应。具体可见修订稿第 18 页。

意见 11：生态效度：投资任务与实际团队情境中“绩效反馈—协作任务”耦合程度有限（真实后果、风险、相互依赖度偏低）；应讨论其对外推的限制。

回应：非常感谢审稿专家对实验生态效度提出的中肯意见。我们已在 5.4 节“研究局限与未来方向”中进行了详细的补充和讨论：我们强调了由于被试群体和任务场景的限制，关于研究成果的理论鲁棒性和应用可推广性仍需未来研究进一步检验，并且指出实验设计在真实后果和风险水平、任务相互依赖度、环境复杂性等方面的不足，同时提出了针对性的未来研究方向，包括开展现场研究、设计高相互依赖任务、系统操纵情境因素等。

具体修改内容可见修订稿第 25 页，我们用红色进行标注以便您审阅。

第二轮

审稿人 1 意见：

感谢作者根据首轮审稿意见对稿件所做的认真、细致修改。通读修订稿后可以看出，作者对审稿意见回应充分、态度严谨。本文聚焦于人机协作情境下，机器人与人类多轮次正负反馈对信任与合作动态演变的差异化影响，研究问题具有鲜明的现实意义与一定的前沿价值。目前存在以下几点需要进一步修订之处：

意见 1：Winsorization 通常指对极端值进行替换（截尾处理），而非直接剔除数据。当前文中表述与操作之间可能存在不一致之处。若实际操作中有数据舍弃（而非替换），建议在方法部分进行具体说明，如说明被舍弃数据的比例。

回应：非常感谢审稿专家对我们前期工作的肯定以及提出的宝贵建议。我们根据您的建议在方法部分对数据处理进行了更详细的补充说明。首先，实验 1 和实验 2 中均未发现超出理论区间的不合理极端值，因此我们实际上并未对数据进行 Winsorization 截尾处理。其次，关于数据剔除的情况，有极少量试次因队友建议与被试初始答案完全相同（即公式分母为 0，导致行为信任无法计算）而被舍弃，其中实验 1 有 1.93%的试次数据被剔除，实验 2 则有 2.00%的试次数据被剔除。

具体修改内容如下（见正文第 9 页）：

“行为信任根据被试在接收队友建议后调整自身初始答案的程度来进行评估，计算公式为：

$$行为信任 = \frac{最终答案 - 初始答案}{队友建议 - 初始答案}$$

该指标取值范围理论上为[0,1]，其中 0（最终答案 = 初始答案）代表完全不信任队友，1（最终答案 = 队友建议）代表完全信任队友。参考前人的处理方法，如果行为信任超出理论区间，则需要对该试次数据进行截尾处理(Lanz et al., 2024; Logg et al., 2019)。值得注意的是，实验 1 没有发现超出理论区间的不合理极端值，仅有 1.93%的试次（27 个试次，涉及 14 个被试）因队友建议与其初始答案完全相同（即公式分母为 0）而被剔除。”

意见 2：虽然作者在回复中提到本研究的实际操作中，未出现最终答案超出区间（0-1）的情况，但信任存在方向性，且引用的 Logg et al. (2019)也未标明使用绝对值。因此建议作者将行为信任计算公式中的绝对值符号去掉，以免给其他读者造成误解。

回应：非常感谢审稿专家的中肯建议。我们完全认同您关于信任具有方向性的观点，使用绝对值会丢失这一重要的方向性信息，并可能导致对信任行为的误读。

我们已根据您的建议对方法部分进行了修订，删去了原有行为信任计算公式中的绝对值符号，并进行了具体说明。

具体修改内容如下（见正文第 9 页）：

“行为信任根据被试在接收队友建议后调整自身初始答案的程度来进行评估，计算公式为：

$$行为信任 = \frac{最终答案 - 初始答案}{队友建议 - 初始答案}$$

该指标取值范围理论上为[0,1]，其中 0（最终答案 = 初始答案）代表完全不信任队友，1（最终答案 = 队友建议）代表完全信任队友。参考前人的处理方法，如果行为信任超出理论区间，则需要对该试次数据进行截尾处理(Lanz et al., 2024; Logg et al., 2019)。值得注意的是，实验 1 没有发现超出理论区间的不合理极端值，仅有 1.93%的试次（27 个试次，涉及 14 个被试）因队友建议与其初始答案完全相同（即公式分母为 0）而被剔除。”

意见 3：文中存在统计量与文字表述不一致的地方，请再仔细检查，如：

（1）“在机器人来源反馈条件下，前期变化不显著(β = -0.499，SE = 0.175，p = 0.005)”。p = 0.005 显然是统计显著的，与“不显著”的文字描述矛盾。

（2）“互动阶段的主效应显著($\chi^2(4) = 5.89$，p = 0.001)，建议采纳随互动进程发生了显著变化”。以 df = 4 的卡方分布，$\chi^2 = 5.89$ 不会对应 p = 0.001，请检查是否存在错误录入。

回应：非常感谢审稿专家细致的审阅工作。针对您指出的表述问题，我们已对全文的统计结果和文字表述进行了逐一核对，并对发现的错误进行了全面修正。具体说明如下：

（1）关于该处统计数据“β = -0.499，SE = 0.175，p = 0.005”是正确的，但文字表述“不显著”属于笔误。我们在撰写时误将 p = 0.005 读成了 p = 0.05，导致了这一明显的矛盾。该处已修正为“在机器人来源反馈条件下，未来合作意愿在前期阶段(β = -0.499, SE = 0.175, p = 0.005; $\beta_{standardized}$ = -0.13, 95% CI [-0.23, -0.04])以较缓速率下降”。

（2）该处确实存在数据错误。正确的统计结果应为 $\chi^2(4) = 9.86$，p = 0.043。此错误产生的原因是：在论文撰写过程中，我们从其他文段复制了相似格式的统计报告语句，修改数据后未及时保存，导致最终版本中仍保留了错误数据。我们已在修订稿中对该部分数据进行了修正和核对。

我们对这些疏漏深表歉意，也再次感谢您细致入微的审阅，帮助我们发现并纠正了这些错误。

意见 4：研究结论中“本研究首次揭示了机器人与人类来源反馈对信任与合作动态演变过程的差异化影响”的表述，建议作者在措辞上更加准确：

（1）当前表述未体现研究中的“正负反馈”信息；

（2）之前已有很多人与机器博弈的研究，这个“首次”是否恰当？

回应：非常感谢审稿专家对研究结论表述的精准指正。您的两点意见都非常中肯，我们已对相关内容进行了审慎的修订，使整体表述更加准确。我们做了如下调整：1）我们明确指出研究聚焦于“多轮次正负反馈情境下”的差异化影响，并进一步说明这种差异主要体现在“负性反馈条件下”；2）我们将“首次揭示”改为“揭示了”，更客观地定位本研究的学术贡献。

具体修改内容如下（见正文第 25 页）：

“本研究揭示了多轮次正负反馈情境下，机器人与人类来源反馈对信任与合作动态演变过程的差异化影响模式。研究发现，这种差异化影响主要体现在负性反馈条件下：当反馈来源是机器人队友时，个体的信任与合作表现出“滞后累积式”的演变模式，具有延迟反应和累积效应的特点；而当反馈来源是人类队友时，个体的信任与合作则表现出“即时响应式”的演变模式，具有高敏感性和强适应性的特征。”

再次感谢您的细致审查和宝贵建议！

……

审稿人 2 意见：

本次修改可以看出作者团队投入了扎实而细致的工作。引言按“动态缺口—反馈互动—来源特殊性—负性非对称”的脉络重新排布，使问题链条更清晰；标题与行文也更明确地区分了“行为信任（采纳/依从）”与“未来合作意愿”，避免了此前“协作信任”一词过于笼统带来的混淆。方法与统计方面，增补的合并分析同时考察了来源 × 效价 × 时间的整体效应，对行为信任的稳健性进行了二元采纳的对照检验，并纳入了建议差值这一关键协变量；描述性统计、信度与效应量的补报亦使文本更合规。语言与结论的收敛同样到位，明确承认机制未测与生态外推的边界，整体朝着完善研究迈出了实质性一步。在此基础上，仍有若干关键环节可进一步收束，使实证证据与结论更为“对齐”：

意见 1：动态轨迹的建模与呈现。作者已补充线性/二次/分段模型比较，这是重要进展。但读者仍需要更明确的模型“骨架”。建议在方法或附录中给出完整的模型公式与参数化说明，包括：随机效应结构、分段模型的断点设定依据、因子编码与参考水平。配套展示随时间的边际效应条件估计曲线（按来源、按效价分图），并附 95%置信区间信息。若断点结合了理论与探索性程序确定，可在附录补充变化点搜索的敏感性分析结果；即便结果为“未发现显著优于固定断点的证据”，也能提升说服力。

回应：非常感谢审稿专家对本文前期修改工作的肯定及进一步提出的建设性意见。我们完全认同您关于需要更清晰展示“模型骨架”的建议。针对您提出的具体要求，我们已在修订稿附录部分进行了系统性补充，具体说明如下：

1、完整的模型公式与参数化说明（附录 A1-A2）：我们在附录中系统呈现了三类混合效应模型的完整数学形式，包括：（1）线性模型的一般结构与参数定义（附录 A2.1）；（2）二次模型的扩展形式（附录 A2.2）；（3）分段模型的完整参数化方案（附录 A2.3）。所有模型均明确标注了参数含义。

2、分段模型的断点设定依据（附录 A2.3）：我们在附录中详细说明了将 T1 作为断点的理论与实证依据：理论上，互动初期被认为是信任快速调整与预期重估的关键阶段（Dirks et al., 2009）；实证上，描述性统计与混合方差分析结果均显示 T0–T1 与后续阶段在变化趋势上存在差异。

3、断点选择的敏感性分析（附录 A2.4）：遵循您的建议，我们补充了系统性的断点敏感性分析。具体而言，我们比较了将 T1、T2、T3 分别作为断点时的模型拟合表现。结果显示，当前断点（T1）在 AIC 和 BIC 指标上均优于其他备选断点，增强了该断点选择的透明度与稳健性。

4、边际效应的条件估计与可视化（附录 A3）：我们已绘制分反馈来源的边际均值变化曲线（图 A1、图 A2），并附带 95%置信区间信息，直观展示不同条件下信任与合作随时间的动态变化模式。

意见 2：三重交互的不显著与“非对称性”的表述。在合并分析中，来源 × 效价 × 时间的三重交互未达统计显著，这意味着“来源影响时间轨迹”的效应并未在正负两种情境下形成统计意义上的条件性差异。文本目前强调“负性情境下差异更显著”的事实观察并无不当，但建议在结果与讨论中明确写出三重交互不显著的边界含义，避免形成“全局条件化成立”的印象。相应段落可收敛为：“在本范式与当前样本下，负性条件呈现更清晰的来源差异趋势”，以与证据强度相匹配。

回应：非常感谢审稿专家的专业建议。您指出的三重交互效应统计不显著及相关表述的问题非常重要，我们完全认同您的观点，即在三重交互效应未达统计显著的情况下，需要谨慎界定“非对称性”表述的边界条件与证据强度。据此，我们已对全文相关部分进行了系统性修订，

具体如下：

1、结果部分（见正文第 20 页）。我们对非对称效应进行了更精准化的描述，采用了“在本研究范式与样本条件下，负性反馈情境中呈现出更清晰的反馈来源差异趋势，提示反馈效价可能在信任与合作动态演变中起到调节作用，正负反馈可能存在非对称效应”的表达方式，并增加了“需要注意的是，整合分析中三重交互效应（互动阶段×反馈来源×反馈效价）在行为信任上未达显著，在未来合作意愿上仅达边缘显著，这可能部分源于当前样本量对于检测复杂交互效应的统计检验力有限，因此关于反馈来源影响模式在正负情境下的差异，仍需后续研究提供更多更强有力的证据。”的说明。

2、讨论部分（见正文第 22 页）。我们进行了更审慎的解读，采用“正负反馈对信任与合作影响的潜在非对称性效应”的保守说法，并增加了对三重交互统计结果的明确说明，强调这一发现“应被视为在当前范式与样本条件下的趋势性发现。未来研究需要通过更大样本量、多种任务情境以及不同类型机器人的系统验证，以明确这一非对称效应的稳健性和边界条件”。

3、研究局限部分（见正文第 24 页）。我们新增了一段关于统计检验力局限的说明，指出“当前样本量虽然满足单个实验的统计效力要求，但对于检测复杂三重交互效应的统计检验力可能有限，未来需要更大样本量和多范式多情境验证来增强证据强度”。

4、结论部分（见正文第 25 页）。修改了第(2)点的表述，采用“在当前范式与样本条件下”、“可能存在非对称性效应”等更为谨慎的措辞。

意见 3：多重比较与效应量报告的统一口径。多处涉及时间点的简单效应与两两比较，建议统一采用 Holm 或 Tukey 进行校正，并在表格中直接标示校正后的 p 值。ANOVA 情境下可保留 partial η²，同时在附录补充 generalized η² 或 ω²，便于跨研究对照；混合效应模型建议以边际均值差及其 95% CI 呈现，并配合标准化系数或 OR，使效应大小更直观可比。另请逐条核对图示与正文数值，避免出现数字与图形不一致的情况。

回应：非常感谢审稿专家对结果呈现提出的系统性建议。我们已按照您的建议进行了全面修订，具体说明如下：

1、关于简单效应与两两比较校正：我们已统一采用 Holm 法($\alpha = 0.05$)对所有的简单效应分析和事后两两比较进行校正，并在文中直接呈现了校正后的 p 值（标注为 p_{Holm}）。

2、关于效应量报告：针对 ANOVA 分析，我们在正文中保留了 partial η^2 作为主要效应量指标，同时在附录 B 中补充了完整的效应量对比表，包括 generalized η^2、ω^2以及 95%置信区间，便于读者进行跨研究比较。

3、关于混合效应模型呈现：我们在附录 A3 中补充了边际均值变化曲线（图 A1、图 A2），并附带了 95%置信区间信息。同时我们在文中补充了标准化回归系数($\beta_{\text{standardized}}$)，使效应大小更加直观可比。

4、关于数据一致性核查：我们已逐一核对全文所有图表与正文描述的数值，确保了图示、表格与正文叙述的一致性。

意见 4：行为信任指标的稳健性细化。作者已将建议偏离幅度纳入协变量，并补充了“是否跨中点”的二元检验，方向正确。建议进一步明确：1）“分母接近 0”的处理阈值与规则，报告剔除比例，并给出“剔除/不剔除”两套结果的一致性对照；2）附录提供原始分布（直方图/QQ 图）并说明温泽化是否触发；如未发生剔除，可在文中简要说明，以免读者误解为进行了较多预处理。

回应：非常感谢审稿专家提出的细致建议。我们已根据您的建议在方法部分对数据处理进行了更详细的补充说明。首先，关于数据剔除，仅有极少量试次因队友建议与被试初始答案完

全相同（即公式分母为 0，导致行为信任无法计算）而被舍弃，其中实验 1 有 1.93%的试次数据被剔除，实验 2 则有 2.00%的试次数据被剔除。其次，实验 1 和实验 2 中均未发现超出理论区间的不合理极端值，因此我们实际上并未进行温泽化处理。

具体修改内容如下（见正文第 9 页）：

“行为信任根据被试在接收队友建议后调整自身初始答案的程度来进行评估，计算公式为：

$$行为信任 = \frac{最终答案 - 初始答案}{队友建议 - 初始答案}$$

该指标取值范围理论上为[0,1]，其中 0（最终答案 = 初始答案）代表完全不信任队友，1（最终答案 = 队友建议）代表完全信任队友。参考前人的处理方法，如果行为信任超出理论区间，则需要对该试次数据进行截尾处理(Lanz et al., 2024; Logg et al., 2019)。值得注意的是，实验 1 没有发现超出理论区间的不合理极端值，仅有 1.93%的试次（27 个试次，涉及 14 个被试）因队友建议与其初始答案完全相同（即公式分母为 0）而被剔除。”

意见 5：机制未测的补强与旁证。作者对未测机制的说明与后续规划较为清楚。若条件允许，建议增补一个小型、低负担的在线附加验证：在相同或近似话术下采集“反馈接受度”“感知能动性”“被冒犯/威胁感”“意图/善意归因”等简短量表，验证来源 × 效价对关键机制变量的方向是否与正文解释一致。若补测不便，可对现有实验录像进行独立双盲的情绪强度/负性表情编码并报告一致性系数，作为“反馈 → 情绪 → 信任/合作”的旁证。即便为探索性结果，也能增强解释链条的可信度。

回应：非常感谢审稿专家对机制验证提出的建设性建议。我们高度重视这一意见，并从多个角度进行了尝试,具体说明如下：

首先，我们按照您的建议设计了一个在线补充实验（128 人），采用与原实验近似的情景设置，采集反馈接受度、情绪、能力感知、善意感知、感知公平性等关键机制变量。然而，在数据收集时，我们发现相较于实验室实验，在线被试普遍反映“队友身份”和“队友反馈”的代入感较弱。同时，事后开放式问题显示，有相当比例的被试猜到了实验目的和假设。此外，由于测量量表较多，有约 37%的被试存在随意作答的情况，在后几个反馈的测量中一致选择量表极端值。经过审慎考虑，我们认为这一补充实验的质量难以达到学术严谨性要求，因此未将其纳入文中。

为尽可能回应您的关切，我们转而深入挖掘了原有实验中已收集的数据。我们注意到，原本用于操纵检验的反馈评价题目（“你在多大程度上能够接受刚才队友的反馈”，7 点量表，1 = 完全不能接受，7 = 完全能够接受）在概念上与“反馈接受度”较为契合。因此，我们将这一单题目指标作为反馈接受度的测量，对其在不同反馈来源和互动阶段下的动态演变模式进行了探索性分析（详见附录 C）。

分析结果显示：（1）在多轮次负性反馈条件下（实验 1），被试对机器人队友的反馈接受度在第一次负性反馈时显著高于人类队友，但随后逐渐下降至与人类组无异，这一模式与我们在正文中所提到的“个体在初期对来源于机器人的负性反馈表现出较低的反应敏感性”和“累积效应”的解释方向一致。而被试对人类队友的反馈接受度在第一次负性反馈时就已经下降到较低值，这也支持了我们在正文中所提到的“即时响应”和“高敏感性”特征；（2）在正性反馈条件下（实验 2），反馈接受度整体保持稳定，且不受反馈来源的显著影响，这与正性反馈下行为信任和未来合作意愿的稳定模式相呼应。

当然，我们也充分认识到这一补充分析存在局限性：其一，反馈接受度仅以单题目测量，

信度和效度有待进一步验证；其二，样本量有限，无法开展正式的中介分析；其三，缺少其他关键机制变量（如感知能动性、归因等），使得我们仍无法构建完整的机制链条。因此，我们在附录 C 中明确标注此为“探索性分析”，并在讨论部分增加了相应的局限性说明和未来研究建议。

综上，虽然我们按您的建议所做的补充在线实验的结果不够理想，但通过对现有数据的深度挖掘，仍然获得了与理论解释方向一致的初步证据。我们诚恳希望这一补充分析能够在一定程度上增强解释链条的初步可信度，为未来更深入的研究提供基础。

再次感谢您的耐心指导和宝贵建议，我们从中受益良多！

第三轮

审稿人 1 意见：

作者已回答了我的所有问题，没有进一步修改意见

审稿人 2 意见：

作者对第二轮审稿意见进行了认真、细致的回复和修改。针对统计模型，作者在附录中补充了混合效应模型的详细参数设置、断点选择依据及可视化结果，显著提升了结果的可信度。针对理论推论，作者修正了关于“非对称性”效应的表述，客观承认了三重交互效应的统计局限，结论更加严谨。在数据处理细节（如公式修正、多重比较校正）和效应量报告方面，均已按要求完善。

编委意见：

感谢投稿学报和在稿件修改过程中付出的努力。文章几经修改，整体质量有了明显提升，基本可以达到发表的水平。但在这之前，有几点小建议供你们参考：（1）文章汇报的研究并不是很多很复杂，建议对篇幅进行精简，提升整体可读性；（2）不论是中文还是英文的摘要目前都过长，建议删减和进一步凝练；（3）图 2 包括一些英文单词，建议换成中文。

回应：非常感谢编辑提出的细致建议。我们根据您的建议对全文内容进行了进一步的修改。具体说明如下：

（1）我们已对稿件篇幅进行了精简，重点删减了前言中的宏观背景与案例性描述，并对讨论部分中与前文重复的表述进行了合并，同时对方法部分做了必要的概括，以提升行文紧凑性与整体可读性。

（2）我们已对中文摘要与英文摘要进行了显著删减与凝练，使摘要长度与内容结构更符合期刊要求。

（3）我们替换了图 2，并检查了全文图片，确保都是中文版本。

再次感谢学报编辑的宝贵建议！

第四轮

主编意见：

稿件围绕机器人与人类来源的多轮正负反馈对信任与合作动态演变的影响展开，选题契合人机协作与智能时代心理学研究方向，具有较好的现实意义和一定理论贡献。综合目前稿件和三轮审稿回复来看，稿件不存在影响发表的重大理论、设计或统计根本缺陷；作者对前

期专家意见总体回应较为充分，已在概念区分、行为信任指标说明、混合效应模型、稳健性检验、多重比较校正、效应量报告、机制未测局限和结论收束等方面作出实质性修改。以下几点，仍需考虑：

意见 1：请进一步约束“正负反馈非对称性”的表述。整合分析中三重交互在行为信任上未达显著，在未来合作意愿上仅达边缘显著，因此摘要、5.2 节和结论部分应统一表述为“负性反馈情境呈现更清晰的来源差异趋势”或“提示可能存在非对称性”，避免写成已经稳健证明的效应。

回应：感谢主编对我们前期工作的肯定以及提出的宝贵建议。我们根据您的建议对全文中有关“正负反馈非对称性”的表述进行了进一步约束与统一。具体说明如下：

我们已全面检查全文表述，并修改了摘要、4.4 节、5.2 节和结论部分中关于“正负反馈非对称性”的相关表述，避免将其表述为已经得到稳健证明的效应。

在摘要部分和 4.4 节部分，我们将相关表述统一调整为“负性反馈情境呈现更清晰的来源差异趋势”以及“提示正负反馈对信任与合作的影响可能存在非对称性”，以更加准确地反映整合分析结果。

在 5.2 节中，我们将标题和正文中的“非对称性效应”等较强表述修改为“潜在非对称性”“趋势性结果”等更为审慎的表达，并补充说明整合分析中三重交互效应在行为信任上未达显著、在未来合作意愿上仅达边缘显著，因此当前结果尚不足以支持稳健非对称效应的结论。

在结论部分，我们进一步统一了相关表述，将研究发现限定为“负性反馈情境中反馈来源差异趋势更为清晰”，并表述为“提示可能存在非对称性，仍需未来研究进一步检验”。

意见 2：请约束 5.3 节实践启示。当前证据尚不足以直接支持“组织应安排机器人传递敏感反馈”或“机器人与人类管理者进行明确反馈分工”等具体管理建议。请将相关内容改为方向性启示，并明确具体应用仍需真实组织场景、高生态效度任务和更多类型机器人的验证。

回应：感谢主编提出的意见和建议。根据您的建议，我们已对第 5.3 节“理论贡献与实践启示”中的相关表述进行了进一步约束和修改。具体说明如下：

一方面，将原文中关于“机器人传递敏感反馈”及“机器人与人类管理者进行明确反馈分工”等具体建议调整为方向性实践启示，强调未来智能组织在探索机器人参与反馈沟通时，应关注不同反馈来源可能带来的差异化心理影响，并重视持续互动过程中的信任变化；

另一方面，明确补充说明机器人在组织反馈系统中的潜在应用仍需在真实组织场景、高生态效度任务以及更多类型机器人的条件下进一步验证。

意见 3：请明确区分机制解释与机制检验。关于归因、意图性感知和情绪反应的解释应写成可能机制或理论解释路径，不应使读者误认为本研究已经完成中介机制检验。请在 5.1 节和 5.4 节同步调整，使表述与证据强度一致。

回应：非常感谢主编对文章表述的精准指正。此前的表述确实存在一定的误解可能性，我们已对第 5.1 节和第 5.4 节进行了同步修改。具体说明如下：

（1）在第 5.1 节中，我们对归因、意图性感知和情绪反应相关表述进行了调整，将其明确写作基于既有理论和研究提出的可能解释路径，而非本研究已经检验的中介机制。

（2）我们在相关段落中进一步弱化了因果性和机制性表述，改用“可能”“或许可以解释”“有助于理解”“理论路径”等表述，使讨论内容与本研究证据强度保持一致。

（3）在第 5.4 节中，我们同步补充了研究局限，明确说明本研究尚未直接测量或检验归因、意图性感知和情绪反应等机制变量，未来研究可进一步加以检验。

意见 4：请核对 4.3.5 节统计报告。该节应统一使用“整体调整幅度”而非“整体提升程度”；同时请复核行为信任结果中的 t 值、p 值、Cohen’s d 及检验方向。若采用双侧检验，请据实修正显著性判断和相应讨论。

回应：非常感谢主编的细致审阅。我们已对第 4.3.5 节的统计报告及相关讨论进行了核对和修改，并再次检查了全文统计分析的数据。具体说明如下：

一方面，我们已将第 4.3.5 节中相关表述统一修改为“整体调整幅度”，避免与“整体提升程度”等方向性表述混用，并进一步明确该分析以 T4 与 T0 差值的绝对值（|T4−T0|）来进行表征。

另一方面，我们重新核对了行为信任结果中的 *t* 值、*p* 值、Cohen’s d 及检验方向，并按照双侧独立样本 *t* 检验据实修正统计报告。同时，我们也对实验小结和讨论部分中涉及正负反馈非对称性的表述进行了约束，使其更加准确、审慎。

意见 5：请修正自检报告与正文之间的数据处理不一致。自检报告第 7 项不能写“没有剔除数据”，应说明实验 1 和实验 2 均有少量试次因分母为 0 被剔除，并报告比例、原因及对结果的影响。请同时核对正文、附录和回复信中相关表述，确保完全一致。

回应：非常感谢主编提出的宝贵建议。我们已全面核对并统一修改自检报告第 7 项、正文、附录及回复信中有关数据处理的表述。具体而言，我们明确说明预实验有 8 名被试因中途退出或所有题目均选择同一选项而被排除；实验 1 和实验 2 均未剔除被试，但分别有 27 个试次（1.93%）和 28 个试次（2.00%）因队友建议与被试初始答案相同、导致行为信任计算公式分母为 0 而无法计算，故从相关分析中剔除。同时，我们补充说明了该处理仅涉及行为信任相关试次，不影响未来合作意愿等其他指标。

意见 6：请统一弱化原创性宣称。摘要、5.3 节、结论和自检报告“研究亮点”中如仍有“首次揭示”、“突破传统框架”等表述，请改为更稳健的学术表述，如“从动态视角考察”、“进一步揭示”、“提供经验证据”等。

回应：感谢主编对文章表述的细致审阅。根据您的建议，我们已对全文中涉及原创性宣称的相关表述进行了系统检查，并对措辞较强或可能超出研究证据边界的表述进行了统一弱化和调整。具体修改如下：

我们重点修改了摘要、第 5.3 节“理论贡献与实践启示”、第 5.4 节研究局限与未来方向、第 6 节结论以及自检报告“研究亮点”中的相关内容，将原文中“首次揭示”、“突破传统框架”、“填补空白”、“超越传统静态‘快照’研究局限”等较强原创性表述，调整为“从动态视角考察”、“进一步揭示”、“提供经验证据”“补充研究视角”、“提供参考”等更为稳健的学术表达。与此同时，我们也对相关段落的上下文进行了同步润色，使研究贡献的表述更加审慎、连贯，并与本研究的实验设计和数据结果保持一致。

再次感谢您的耐心指导和宝贵建议！