修改说明（第二轮）

亲爱的编辑：

我们诚挚感谢主编对稿件提出的中肯意见。针对本轮“标题歧义、卖点不清”的关键问题，我们认真对待并对标题和摘要进行修订，此外，本论文全体作者系统梳理并贯通“标题—摘要—引言—讨论”的逻辑，具体修订如下：

首先，我们将题目改为：价值感知与信息整合的差异：人类与当前大语言模型在爱荷华赌博任务中的对比。我们认为该标题可以（1）将“卖点”前置、直接陈述研究的主要发现和核心讨论内容——人类与当前大语言模型在反映价值感知、信息整合等方面的参数差异。（2）阐述清楚研究的主体（人类 *vs.* 当前大语言模型）和使用的范式（爱荷华赌博任务）。

第二，我们修订摘要的结果部分，使其与题目呼应：“两类决策主体在逆衰减率、价值敏感度和稳定性参数上存在差异：人类在价值信息整合方面占优，而大语言模型在价值感知等方面占优。”作为文章的核心结论，呼应了题目的“价值感知与信息整合的差异”

最后，我们仔细检查了全文逻辑，并梳理“标题—摘要—引言—讨论”的逻辑。**标题：**点明主要结果（价值感知与信息整合的差异）、比较对象（人类与当前大语言模型）和研究范式（爱荷华赌博任务）。**摘要：**交代研究的背景（理解人类与大语言模型的差异以优化人机协同设计）、方法（用于解析决策序列的价值序列探索模型）、结果（模型比较和参数层面的主要发现）和启示（对人机协同设计与决策提升的意义）。**引言：**从人机协同设计欠佳的现实问题与大语言模型的决策表现的学术争议入手，归纳相关理论观点、实证发现并介绍范式选择依据，详细介绍可有效把爱荷华赌博范式的行为表现解析为可估参数的价值序列探索模型，最后，阐明研究的具体设计和基于前述理论的研究假设。**讨论：**聚焦于核心假设和研究发现对结果进行充分探讨，归纳了研究的理论贡献和实践启示。

综上，我们已围绕“标题歧义、卖点不清”的主编意见，完成针对性修订并贯通“标题—摘要—引言—讨论”四部分逻辑，相应修改部分在手稿中标红，谨请再次审阅并惠赐指正。

修改说明（第一轮）

亲爱的编辑和审稿专家：

首先，我们想诚挚感谢编辑和审稿专家对我们的稿件《谁是更好的“赌徒”？人类与大语言模型：在爱荷华赌博范式中的探索与利用》提出的宝贵意见和修改建议。我们根据这些意见和建议，对稿件进行了认真严谨的修改。

根据编辑部审核意见，修订稿中已将 p 值精确到小数点后 3 位，此外，我们已邀请本领域中级及以上职称的两位同行（研究方向分别为认知建模和决策神经科学）进行挑剔性阅读。

根据审稿专家的意见，我们已明确研究核心假设，阐述模型选取依据、完善方法细节、添加候选的策略转换模型、补充基线与 CoT 条件的事后检验差异、人类样本差异、操纵性检验结果、增强讨论深度、归纳理论价值和实践启示。

下面是对每位审稿专家意见的回复，请您审阅：

审稿专家 1#

1. 引言中讲“在此背景下，理解 LLM 系列决策表现与人类的异同，厘清二者的优势，对于优化人机协作系统设计至关重要(Niu et al., 2024)。”那么，本研究的结果对优化人机协作系统设计的启示是什么？对此问题，讨论中并未涉及。

回复：感谢审稿专家提出的重要建议。我们已在讨论中新增“**4.3 理论价值和实践启示**”，从两个角度对启示进行总结：（1）角色分工：认知建模技术启示我们在具体决策情境中基于决策主体的优势进行分工：对高稳定性与精确价值评估要求高的环节交由 LLM；需要整合长期经验并在动态环境中实时调整的环节由人类主导；（2）决策提升：基于模型参数差异，我们提出面向人和 LLM 的双向改进路径。人类可以利用 LLM 驱动的即时反馈训练提高价值敏感度、通过类似于对 LLM 进行 CoT 提示的方式，提升决策表现；LLM 在训练和模型架构设计时可以考虑仿生人类的“躯体标记”或“要义表征”机制，优化注意力机制。具体内容见修订稿 4.3 部分，或参见下文：

“在实践启示方面，第一，指导优势互补型人机协同系统的设计：有必要基于决策主体的优势，进行角色定位与任务分配。例如，在系列决策情境中，由 LLM 处理对估值精度与稳定性要求较高的任务；由人类负责长期、模糊经验的整合，以及在动态情境下的实时策略调整。第二，指出双向提升人与 LLM 的决策能力路径：一方面，人类可以通过基于 LLM 的训练程序，校准对收益/损失信号的感知与评估，也可以通过类似于对 LLM 进行 CoT 提示的方式，利用结构化的出声思维报告来提升决策表现 (Ericsson & Moxley, 2011)；另一方面，LLM 在设计上下文记忆机制时，或可借鉴人类的“躯体标记”或“要义表征”机制来整合信息，利用关键的情绪线索和规律信息优化注意力机制，提升决策质量 (Lee et al., 2024)。上述启示尚需通过具体实验设计，以明确其效果和机制。”

2.“鉴于人类大脑与 LLM 在结构上的根本差异，本研究支持后者观点：在系列决策情境中，LLM 与人类将展现出不同的决策表现。”根本差异具体是什么？作者并未介绍。由于这里并未介绍，对读者理解本文的结果造成困难。如，为什么 LLM 同样拥有理解上下文的能力，比人类更强的记忆能力，但整合历史信息的能力却不如人类。

回复：感谢审稿专家提出的重要建议。我们已在引言中添加段落，阐释现有的关于决策比较研究的结论差异以及人类大脑与 LLM 之间的结构差异。一方面，人类依赖有限的认知资源和情感信号，进行价值信息的整合；而 LLM 则基于其庞大的上下文窗口，进行一种全局性的统计整合；另一方面，我们补充了人类存在认知疲劳的特点，这直接影响了在系列决策任务中的策略稳定性。具体内容见修订稿红色部分，或参见下文：

“本研究从人类大脑与 LLM 结构上的差异出发，支持后者观点。具体而言，人类大脑的认知资源、记忆能力相对有限，但可以依赖于过往经验引发的生理或者情感信号来为选项赋予价值权重，从而高效地整合历史信息以指导未来选择。而基于 Transformer 架构的 LLM 本质上是一台统计机器，展示另一种决策模式，它的优势在于其强大的上下文记忆窗口，可以精确容纳并处理远超人类生物极限的信息量，但自注意力机制仍存在难以区分关键信息与噪声的局限性(Shan et al., 2025)。另一项关键差异体现在决策策略的稳定性上，随着任务的持续，人类会产生认知疲劳，导致决策质量波动、甚至策略切换 (Pessiglione et al., 2025; 胡馨允等, 2023)。而 LLM 则没有生物学意义上的认知负荷或疲劳。上述差异预示着，在需要不断根据历史反馈来调整策略的系列决策任务中，两者将产生不同的决策表现。”

3. 引言部分各段之间的逻辑需要加强，另外，必要的研究假设需要给出。此外，文中部分问题在前言中并未论述，如有关专家及样本的问题，

回复：感谢审稿专家提出的重要建议。根据您的建议，我们对引言部分进行了以下几方面的修订，

(1)关于引言逻辑，我们重塑并梳理了引言部分的结构，以增强其逻辑性和完整性：修订稿中引言形成漏斗状逻辑结构，遵循一条由问题到理论、由理论到实证的清晰叙事路径：从人机协作这一现实问题与学界的争议入手，提出核心问题——人类与 LLM 的系列决策差异；接着，聚焦于本研究的核心假设——人类与 LLM 的系列决策存在差异、特定的实验范式（IGT）与假设的认知建模分析框架（VSE 模型）；最后，落脚于本研究的研究假设、具体设计，为读者提供了进入实证部分的明确路线，并对研究的核心发现建立预期。具体请见修订稿的引言。

(2)关于研究假设，我们在引言的结尾段落，增加了一个编号清晰的研究假设部分。该部分将基于文献回顾和理论框架提出预测，与 VSE 模型的关键参数（如逆衰减率、价值敏感度、稳定性）进行了直接的、明确的关联。
“基于前文的理论框架和 VSE 模型参数设置，本研究提出以下核心假设：（H1）相较于其他候选模型（具体请见补充材料），VSE 能够最佳描述本研究中的人类和 LLM 的 IGT 决策机制。两类决策主体在 VSE 模型的决策参数上存在差异：（H2）人类被试借助“躯体标记”实现更强的历史信息整合能力，体现为具有较高的逆衰减率；（H3）LLM 被试在价值感知、策略一致等方面展现优势，体现为具有较高的价值敏感度和稳定性参数。”

(3)关于专家级样本，我们在引言的结尾段落添加了另设“专家级样本”进行并行分析的参考依据和具体操作：“借鉴现有研究中将 LLM 与普通人群或领域专家进行比较的不同策略 (e.g., Lu, 2025; Luo et al., 2025)，本研究并行执行了两种比较方案：(1)全体人类样本与 LLM 比较：将全部 1763 位人类被试与 LLM 进行对比分析。(2)“专家级”人类样本与 LLM 比较：根据最终净收益指标，从人类被试中筛选出净收益表现最佳的 48 位被试（与 LLM 被试数量匹配），形成“专家级”人类样本，并与 LLM 进行对比分析。”

此外，我们还在补充材料中展示了全体、专家级、全体去除专家级人类样本在六个指标（最终净收益、价值敏感度、逆衰减率、探索学习率、探索增益、稳定性参数）的均值(*M*)、标准差(*SD*)和样本量 (*N*)。表格 S5 还提供了专家级样本和全体去除专家级样本之间的差异，并进行了独立样本 *t* 检验两组差异。有关具体请见修订稿中引言末段、方法 2.4.2、讨论 4.2、补充材料表 S5 部分。

4. 关于 LLM 模型的选择，作者提出“在 LLM 选择上，我们综合考虑了模型的开源、性能与可用性(Ivanova, 2025)，采用了……”性能是如何考虑的，标准是什么？作者在文中多次提及“高性能 LLM”，如不对性能做出解释，读者无法理解何为“高性能 LLM”。此外，作者仅选取了国产的 Deepseek-V3、GLM-4-Flash 和 GLM-Zero-Preview 三类模型，且未包含国际主流 LLM（如 GPT-4/4.1、Claude 3、Llama3 等），模型范围是否能够代表 LLM 整体性能？

回复：感谢审稿专家的宝贵建议。

(1)本次修订时，我们发现“高性能 LLM”在原文中指代的是那些能够顺利理解任务、完成 IGT 实验要求的 LLM 和条件。因此，我们在修订稿中去掉了“高性能”的表述，并清楚阐明排除/纳入策略，以及纳入后续分析的 LLM 条件:

“进一步分析其日志选择序列发现，GLM-Zero-Preview 在约 71%的试次、GLM-4-Flash 在约 67%的试次中呈现出四牌轮流模式（如 A-B-C-D），且几乎每次（GLM-Zero-Preview：94.8%；GLM-4-Flash：96.5%）都更换牌组。这种持续高频探索但缺乏价值利用的选择模式表明，它们在基线条件下未能理解 IGT 的目标。因此，这两组条件的 LLM 被试在后续分析中被排除，最终纳入的 LLM 条件为：Deepseek-V3（基线）、Deepseek-V3_CoT、GLM-4-Flash_CoT、GLM-Zero-Preview_CoT。”

(2)关于模型代表性的问题。首先，本研究的目标并非得到国际范围内的 LLM 的决策表现排名，而是比较 LLM 与人类在 IGT 中的决策机制和参数差异。因此，在实验设计时，我们考虑了性能、开源、可用性（Ivanova, 2025），在中国大陆的合规可访问范围内，选取了：权重开源的 Deepseek-V3，低成本低时延的 GLM-4-Flash（2025-01-16 起智谱平台提供 Flash 免费 API 调用），以及推理导向的 GLM-Zero-Preview。根据 SuperCLUE 年度报告（2025-01-08），我们选择的模型 Deepseek-V3 和 GLM 系列在性能上具有代表性：Deepseek-V3 以总分 68.3 分位于国内第一梯队，特别是在 Hard 维度（如 Agent 任务 75.0 分、深度推理 58.8 分）和文科维度（如语言理解 86.5 分）表现出色，接近 ChatGPT-4o-latest（70.2 分），且在推理速度和性价比方面具有竞争力。GLM 系列以 65+分位居第一梯队，多样化的模型策略也覆盖了决策推理的现实场景。我们在引言和方法中将 SuperCLUE 年度报告（2025-01-08）作为参考依据。

此外，针对每种 LLM，我们同时设置了基线与 CoT 两种提示策略：基线旨在最大程度贴近人类被试流程（仅做选择、无需外化推理）；CoT 则依据既有研究显示“显式推理外化可提升复杂推断任务表现”的结论，在每轮选择前要求模型以固定格式简要外化推理，再给出选择。

最后，在局限的讨论中，我们也强调了在模型代表性上可能的局限和未来的方向：“为克服上述局限，未来研究首先应扩展模型的范围至更全面的 LLM，探索不同模型、不同条件之间的决策表现异同。”

5. LLM 的有效被试数仅为 48，基于这么小的样本，LLM 的有效被试数仅为 48，基于这么小的样本，要对多个认知模型的优劣在群体和个体水平做出区分都是很困难的。当前的被试数量难以有效降低随机误差的影响。

回复：感谢审稿专家的评论。

首先，请允许我们厘清样本研究确定 LLM 样本量的依据和研究目的：本研究 LLM 样本数 48 的设定来自 4 阶拉丁方设计（24 种顺序 ×每顺序 2 个被试 = 48），保证顺序平衡。我们共采集到 3 个模型×2 个条件×48 个被试 = 288 个 LLM 被试。本研究的目的是将人类（*N*=1763）与 LLM 的 IGT 的选择序列置于统一的计算框架下进行解析，研究的核心假设是 VSE 是能够最好的描述两类决策主体 IGT 决策序列的最优模型。

在模型比较时：我们不仅采用了固定效应的模型比较（直接比较不同模型的 BIC、AIC、-2*F 的数值），还结合了贝叶斯随机效应的模型比较方法。随机效应分析将被试间和条件间的异质性纳入模型的评估流程，能够基于个体拟合的情况稳健地反映不同认知模型在群体层面的相对优劣。在随机效应分析中，我们计算了模型频率（该模型能最优拟合多少个体的被试）和超越概率（计算某个特定模型是该样本最优模型的后验概率）。结果显示，无论在“全体人类样本 vs. LLM”还是“专家级人类样本 vs. LLM”的分析方案下，VSE 模型始终获得了最高的模型频率和接近 1 的超越概率，稳健支持 VSE 在群体水平上的拟合优势，优于其他候选模型。

此外，功率分析的结果也证实本研究的实验设计足以检测到 Cohen's $f \geq 0.082$（$\eta 2 \approx 0.006$）的效应量，可以捕捉到微弱的组间差异。

6. 文章在方法部分明确设计了 LLM 的两个实验条件（基线 vs. CoT）。首先，为什么要设置基线和 CoT 条件，在文中并未介绍。其次，在结果部分及后续讨论中，并未见对该变量的主效应进行报告和分析。CoT 是当前大语言模型研究的重要增强手段，其是否显著改善 LLM 的决策表现、是否改变模型参数特征，直接关系到研究结论的完整性。

回复：感谢审稿专家的宝贵意见。

（1）关于基线与 CoT 条件的设置：我们在修订稿中已在引言末尾做出进一步说明。基线条件的设定是为了尽可能贴近人类被试在 IGT 中的实验流程。人类被试在任务中仅需在四副牌中做出选择，而无需外化推理过程，因此我们在基线条件下用指导语限制 LLM 仅以固定格式输出选择结果，而不生成额外的推理内容。CoT 条件的设置考虑了前人研究发现的结果：要求 LLM 在选择之前输出推理步骤，能够显著提升推理能力。研究结果发现，GLM-4-Flash 和 GLM-Zero-Preview 在基线条件下未能理解任务目标，未体现智能决策特征，因此基线条件被后续研究排除。我们已经在引言部分添加了关于实验条件的理由和设置：

“第三，推理是 LLM 实现决策的核心能力 (Evans et al., 1993)，研究表明，当 LLM 参数达到一定规模时，推理能力开始涌现 (Wei, Tay, et al., 2022)。Wei, Wang, et al. (2022)揭示了这种推理通常借助“思维链” (Chain-of-Thought, CoT)——即一系列自然语言中间步骤——来实现或者增强。”

“并为每个模型设计了两种实验条件，即基线条件和 CoT 条件，每个条件下采集了 48 位被试。基线条件旨在尽可能贴近人类被试在 IGT 中的实验流程：人类仅需在四副牌中做出选择，而无需外化推理过程；CoT 条件则要求 LLM 在选择前输出推理步骤。”

（2）关于 CoT 是否显著改善 LLM 的决策表现：我们在补充材料中增加了 Deepseek-V3 模型基线与 CoT 条件的事后比较结果。结果显示，逆衰减率与探索增益在 CoT 条件下显著高于基线条件，而其他参数的差异均未达到显著水平。这些结果表明，CoT 在一定程度上改善了 Deepseek-V3 的历史信息整合与探索动机，但并未将最终收益提升至显著水平。

表 S4 Deepseek-V3 基线与 CoT 条件的事后比较结果

参数	差异 (CoT–Baseline)	*SE*	*t*	*df*	P_{adj}
最终净收益	-381.04	299.41	-1.273	93.94	0.709
价值敏感度	0.05	0.045	1.111	86.52	0.800
逆衰减率	-0.16	0.038	-4.19	93.98	< .001
探索学习率	-0.084	0.053	-1.575	91.05	0.517
探索增益	-0.428	0.14	-3.065	84.47	0.024
稳定性参数	0.367	0.181	2.03	93.45	0.260

7. LLM 在完成 IGT 时包含基线和 CoT 条件，但是人类被试在完成 IGT 时的流程是什么，在文中并未介绍，是更接近基线，还是更接近 CoT 条件？

回复：感谢审稿专家的意见。我们在修订稿中已补充说明人类被试在 IGT 中的实验流程。人类被试在实验中仅需在四副牌中做出选择，而无需报告推理过程。因此，人类实验流程更接近 LLM 的基线条件，而非 CoT 条件。请见修订稿中 2.3 实验流程的表述：

“对于人类被试，实验流程与经典 IGT 一致，被试仅需在四副牌中逐次做出选择，被试无需报告推理过程。对于 LLM 被试，实验设置基线和 CoT 条件：在基线条件下提示 LLM 仅以固定格式输出选择结果，而不生成推理过程；在 CoT 条件下，我们额外要求被试在做出选择之前解释自己的推理过程。”

8. 人类样本逆衰减率更高的原因仅以“躯体标记假说”解释较为单薄。

回复：感谢审稿专家的意见。我们同意仅依赖躯体标记假说来解释人类逆衰减率更高的现象较为单薄。在修订稿中，我们补充了模糊痕迹理论 (Fuzzy Trace Theory, Reyna & Brainerd, 2011)的视角。该理论可以将人类与 LLM 的记忆机制界定为“要义表征”和“字面表征”。在 IGT 任务中，这意味着人类能够将多轮试次的收益损失经验，总结成牌组规律，从而体现出更高的逆衰减率。相比之下，LLM 更依赖逐字式的上下文记忆，在抗记忆衰减方面存在系统局限性。结合情绪加工的躯体标记假说与认知加工的模糊痕迹理论，我们认为人类在历史信息整合上的优势是多机制协同的结果。此外，我们在启示中补充了 LLM 可以借鉴人类的“要义表征”机制，改善信息整合能力，进而提升决策质量(Lee et al., 2024)。

“其次，根据模糊痕迹理论(Fuzzy-Trace Theory; Reyna & Brainerd, 2011)，人类往往能将具体的奖惩经验自动化压缩为“要义表征(Gist Representation)”，这种记忆表征机制具有很强的抗衰减性，可以有效减轻认知负荷，提高信息整合效率。具体到本研究的实验流程，人类被试可能会将每个试次的反馈信息，提炼为牌组的规律。而 LLM 虽然具备上下文记忆能力，但根据模糊痕迹理论的假设，其依赖于纯粹的“字面表征(Verbatim Representation)”机制，是一个典型的逐字记忆系统 (Jiang & Ferraro, 2024)，这在处理需要多轮次信息整合的系列决策任务时，可能整合效率较低。”

“另一方面，LLM 在设计上下文记忆机制时，或可借鉴人类的“躯体标记”或“要义表征”机制来整合信息，利用关键的情绪线索和规律信息优化注意力机制，提升决策质量(Lee et al., 2024)。上述启示尚需通过具体实验设计，以明确其效果和机制。”

9. “专家级”样本的定义、筛选人数与 LLM 样本匹配策略描述较为简略，需进一步明确。文中更多是全体人类样本与 LLM 和专家级样本与 LLM 的并行分析，但是，专家级样本与全体人类样本的差异是什么？并未给出。

回复：我们感谢审稿专家对“专家级”样本的关注。

（1）关于“专家级”样本的描述，我们已在手稿的引言中对“专家级”样本的定义与分析逻辑进行了进一步补充与澄清：

“借鉴现有研究中将 LLM 与普通人群或领域专家进行比较的不同策略 (e.g., Lu, 2025; Luo et al., 2025)，本研究并行执行了两种比较方案：(1)全体人类样本与 LLM 比较：将全部 1763 位人类被试与 LLM 进行对比分析。(2)“专家级”人类样本与 LLM 比较：根据最终净收益指标，从人类被试中筛选出净收益表现最佳的 48 位被试（与 LLM 被试数量匹配），形成“专家级”人类样本，并与 LLM 进行对比分析。”

在讨论中，我们在“4.2 人类和 LLM 的价值利用和探索策略差异：并行分析带来的启示”部分专门强调了并行分析结果的差异带来的启示：

“本研究在进行人类和 LLM 比较时，执行并行分析方案，具体而言，全体人类样本与 LLM 的比较反映了普通人群与 LLM 的差异，而专家级样本与 LLM 的比较则检验了任务表现最佳的人类能否与 LLM 匹敌。两种方案的结果启示我们：人机决策表现的差异结论受到人类样本的任务表现水平的影响，研究者应谨慎设计实验和解释结果。”

（2）关于“专家级样本与全体人类样本的差异”，手稿中的并行分析策略聚焦于人类表现与 LLM 的差异，并不着眼于对比不同水平的人类样本内部的参数差异，在补充材料中，我们添加了对全体、专家级和全体样本去除专家级样本的描述性统计和 t 检验结果。结果表明，专家级样本与全体-专家级样本在多个关键决策指标上存在显著差异，尤其在最终净收益、价值敏感度、逆衰减率和稳定性方面，专家级样本表现出显著的优势。探索学习率在两个样本间未见显著差异，而专家级样本的探索增益表现显著低于全体-专家级样本，提示专家级样本在决策时可能更倾向于精确的评估、稳定的策略、而非广泛探索。具体请见补充材料：

“表格 S5 展示了全体、专家级、全体去除专家级人类样本在六个指标（最终净收益、价值敏感度、逆衰减率、探索学习率、探索增益、稳定性参数）的均值(*M*)、标准差(*SD*)和样本量(*N*)。此外，表格还提供了专家级样本和全体-专家级样本之间的差异，并进行了独立样本 *t* 检验两组差异。结果表明，专家级样本与全体去除专家级人类样本在多个关键决策指标上存在显著差异，尤其在最终净收益、价值敏感度、逆衰减率和稳定性方面，专家级样本表现出显著的优势。探索学习率在两个样本间未见显著差异，而专家级样本的探索增益表现显著低于全体去除专家级人类样本，提示专家级样本在决策时可能更倾向于稳定、精确的评估而非广泛探索。

表 S5 全体、专家级、全体去除专家级样本在各指标上的描述性统计量与独立样本 *t* 检验结果

参数	人类样本	*N*	*M*	*SD*	差异	*t*	*df*	*p*
最终净收益	全体	1763	-1277.15	1034.49				
	专家级	48	1457.71	665.03	2811.41	28.51	1761	<0.001
	全体-专家级	1715	-1353.70	934.17				
价值敏感度	全体	1763	0.28	0.19				
	专家级	48	0.50	0.28	0.22	5.56	1761	<0.001
	全体-专家级	1715	0.28	0.18				
逆衰减率	全体	1763	0.61	0.26				
	专家级	48	0.78	0.20	0.18	6.04	1761	<0.001
	全体-专家级	1715	0.60	0.26				
探索学习率	全体	1763	0.41	0.21				
	专家级	48	0.45	0.27	0.04	0.94	1761	0.351
	全体-专家级	1715	0.41	0.21				
探索增益	全体	1763	1.82	2.41				
	专家级	48	-0.96	2.48	-2.85	-7.88	1761	<0.001
	全体-专家级	1715	1.89	2.36				
稳定性参数	全体	1763	0.67	0.42				
	专家级	48	0.55	0.32	-0.12	-2.63	1761	0.011
	全体-专家级	1715	0.68	0.42				

注释：差异表示 专家级样本均值 - 全体-专家级样本均值。*t* 检验结果比较的是专家级样本和全体-专家级样本之间的差异。“

10. 部分事后检验结果、模型拟合指标放在补充材料中，正文中应至少汇总关键显著性结果。

回复：感谢审稿专家的建议。我们在修订稿中已对关键显著性结果和模型拟合指标进行了汇总。例如，在净收益部分，我们在结果 3.2 部分明确汇总了事后检验结果；在模型拟合指标部分，我们将固定效应、随机效应和预测准确率结果汇总到表 3；在参数比较部分，我们也在结果 3.4 部分阐明了人类与 LLM 之间哪些参数存在显著差异，哪些无显著差异。我们将具体的事后检验统计结果放在了补充材料表 S2 和 S3 中，此外，我们还在补充材料中报告了基线与 CoT 条件间事后检验的结果（表 S4）供感兴趣的读者查阅：

“表 S4 报告了 Deepseek-V3 模型在基线与 CoT 条件下的事后比较结果。本研究的主要目的在于人类与 LLM 的对比，因此这些结果未在正文中呈现。分析显示，逆衰减率与探索增益在 CoT 条件下显著高于基线条件，而其他参数的差异均未达到显著水平。这些结果表明，CoT 在一定程度上改善了 Deepseek-V3 的历史信息整合与探索动机，但并未将最终收益提升至显著水平。”

表 S4 Deepseek-V3 基线与 CoT 条件的事后比较结果

参数	差异 (CoT–Baseline)	*SE*	*t*	*df*	p_{adj}
最终净收益	-381.04	299.41	-1.273	93.94	0.709
价值敏感度	0.05	0.045	1.111	86.52	0.800
逆衰减率	-0.16	0.038	-4.19	93.98	< .001
探索学习率	-0.084	0.053	-1.575	91.05	0.517
探索增益	-0.428	0.14	-3.065	84.47	0.024
稳定性参数	0.367	0.181	2.03	93.45	0.260

注：p_{adj} 为经过 Games-Howell 多重比较校正后的 p 值。

11. 比较模型虽选择的 5 种，但不够全面，作者在讨论中也指出“认知建模的研究支持了人类在进行决策时存在策略切换的现象，因此 VSE 虽能最优的拟合数据，但其设计未能充分捕捉策略切换的动态过程(Feher da Silva et al., 2023; 胡馨允等, 2023)”,既然作者已经认识到人类决策中可能存在 SSO，那么，在模型选择时，应将 SSO 作为补充模型。虽在 100 试次中，策略转换可能表现受限，但将其作为比价模型能够更具说服力。

回复：我们非常感谢审稿专家提出的这一重要意见。正如您所指出的，已有研究表明人类在 IGT 中可能存在策略切换现象 (Feher da Silva et al., 2023; 胡馨允等, 2023)。因此，在本次修订中，我们补充纳入了胡馨允等人提出的 SSO 模型进行拟合与比较。模型拟合的结果如表 3 显示，在固定效应和随机效应分析中，VSE 仍然表现为最优模型，能够在群体水平上最佳地解释人类与 LLM 的决策数据。这一结果提示：可能有部分人类被试遵循策略切换的模式，但 VSE 模型仍是最好描述包含两类决策主体的优胜模型。综上，我们在正文中仍然保持以 VSE 作为主要分析框架的思路，但将 SSO 作为候选模型纳入模型比较的部分。

表 3-候选模型拟合的固定效应、随机效应和预测准确率情况

并行分析方案	模型	固定效应分析			随机效应分析				预测准确率
		BIC	AIC	-2*F	BIC	AIC	F	超越概率	
全体人类样本 & LLM	EV	527348.16	512068.84	517203.24	0.06	0.02	0.11	<.01	0.39
	ORL	489640.02	464174.48	483099.50	0.05	0.10	0.12	<.01	0.49
	PVL	505196.86	484824.43	492190.65	0.01	<.01	0.07	<.01	0.38
	VPP	502895.43	462150.56	475391.92	0.12	0.15	0.02	<.01	0.50
	SSO	524819.06	484074.20	485195.30	0.23	0.10	0.03	<.01	0.42
	VSE	**476542.47**	**451076.94**	**467441.96**	**0.53**	**0.63**	**0.65**	**>.99**	**0.52**
“专家级”人类样本 & LLM	EV	46859.62	44983.9	46499.09	0.01	0.01	0.05	<.01	0.70
	ORL	31142.77	28016.57	29624.56	0.17	0.13	0.17	<.01	0.79
	PVL	31781.97	29281.01	30514.95	0.05	<.01	0.17	<.01	0.77
	VPP	31968.14	26966.22	28578.80	0.26	0.24	0.08	<.01	0.81
	SSO	35691.80	30689.87	30864.75	0.12	0.04	0.01	<.01	0.75
	VSE	**28893.97**	**25767.76**	**27910.75**	**0.42**	**0.56**	**0.53**	**>.99**	**0.82**

12. 文中指出部分 LLM 条件“互信息指数极高，结合其高频探索行为和日志中的选择表现，说明未能理解 IGT 任务目标”。然而，作者并未说明是何种模式的“日志中的选择表现”，导致此处理解困难。

回复：感谢审稿专家的细致建议。我们已在修订稿中具体说明了日志中的选择表现。通过对日志的分析，我们发现 GLM-Zero-Preview 和 GLM-4-Flash 在基线条件下的选择模式具有高度随机性。具体而言，GLM-Zero-Preview 在约 71%的试次中、GLM-4-Flash 在约 67%的试次中呈现出四牌轮流模式（如 A-B-C-D），且其换牌率极高（GLM-Zero-Preview：94.8%；GLM-4-Flash：96.5%），这种持续高频探索但缺乏价值利用的选择模式，说明它们在基线条件下未能理解 IGT 任务的目标。我们据此在修订稿的结果部分明确说明了这一点：

“进一步分析其日志选择序列发现，GLM-Zero-Preview 在约 71%的试次、GLM-4-Flash 在约 67%的试次中呈现出四牌轮流选择模式（如 A-B-C-D），且几乎每次（GLM-Zero-Preview：94.8%；GLM-4-Flash：96.5%）都更换牌组。这种持续高频探索但缺乏价值利用的选择模式表明，它们在基线条件下未能理解 IGT 的目标。因此，这两组条件的 LLM 被试在后续分析中被排除，最终纳入的 LLM 条件为：Deepseek-V3（基线）、Deepseek-V3_CoT、GLM-4-Flash_CoT、GLM-Zero-Preview_CoT。”

13. 讨论部分仅在重复本文发现的结果与已有研究或理论的对这些结果的支持，缺少本研究的理论贡献与实践意义。

回复：非常感谢您提出的宝贵意见。为此，我们在讨论部分新增了“4.3 理论价值和实践启示”一节，旨在总结本实证研究的理论价值和实践意义。具体来说，从三个层面详细论述了本研究的理论价值：第一，在研究范式上，我们强调了认知建模将人机比较从行为表现推向了内在机制。第二，在人类决策理论上，我们阐明了利用 LLM 作为“硅基”对照，为经典人类决策理论提供计算佐证。第三，在 LLM 决策能力界定上，我们基于模型参数，对 LLM 的“理性人”假设进行了精确的量化与解构，并指出了其核心优势与局限。此外，我们也明确提出了研究的实践启示，提出了两点具体的实践方向：（1）指导互补型人机协同系统的设计；（2）为双向提升人与 LLM 的决策能力提供新思路，包括利用 LLM 训练人类和利用 CoT 提示人类决策，以及借鉴人类认知机制优化 LLM 的架构。具体请见讨论 4.3 或如下：

“在理论价值方面，第一，认知建模方法超越了关于人与 LLM“谁更善于决策”的对立，将比较的重点从行为表现优劣转向内在机制差异，这为人机比较、人机对齐提供可借鉴的实现路径。第二，人类与 LLM 模型参数比较的结果，为经典认知理论提供了计算佐证。利用“硅基人类”（LLM）作为人类决策的对照，可以为经典决策理论（如躯体标记假说，模糊痕迹理论）提供精确的计算证据。第三，VSE 模型参数为衡量现阶段 LLM 的决策能力提供了精确的量化证据，支持其为一种有边界的“理性人”：在价值感知和决策一致性等方面优势突出，但在跨试次信息整合方面仍有改进空间。

在实践启示方面，第一，指导优势互补型人机协同系统的设计：有必要基于决策主体的优势，进行角色定位与任务分配。例如，在系列决策情境中，由 LLM 处理对估值精度与稳定性要求较高的任务；由人类负责长期、模糊经验的整合，以及在动态情境下的实时策略调整。第二，指出双向提升人与 LLM 的决策能力路径：一方面，人类可以通过基于 LLM 的训练程序，校准对收益/损失信号的感知与评估，也可以通过类似于对 LLM 进行 CoT 提示的方式，利用结构化的出声思维报告来提升决策表现 (Ericsson & Moxley, 2011)；另一方面，LLM 在设计上下文记忆机制时，或可借鉴人类的“躯体标记”或“要义表征”机制来整合信息，利用关键的情绪线索和规律信息优化注意力机制，提升决策质量 (Lee et al., 2024)。上述启示尚需通过具体实验设计，以明确其效果和机制。”

14. 本研究论文虽写作流畅，但亦存在部分写作问题，但仍需进一步修订文字，如“LLM 实现决策行为主要源于以下三个方面：首先，……其次，……。最终，……。”“最终”一词容易让人产生误解，是对前面内容的总结，而非并列关系；“2.4.1 描述性统计”，这里并非描述性统计，只是在介绍描述性统计指标。本文使用 IGT 任务结合认知建模，比较了大语言模型（LLM）与人类在决策中的差异。整体上，研究问题明确，逻辑清晰，结论较为可靠，且具备一定的理论与实践意义。然而，在正式发表前，仍有若干需要进一步补充和完善的地方。

回复：感谢审稿专家的细致意见。我们已根据建议对相关表述进行了修订：
（1）将“最终”修改为“第三”，以避免歧义，保证三个方面在逻辑上的并列关系
（2）将“2.4.1 描述性统计”修改为“2.4.1 行为指标的计算”，以准确反映该部分的内容。

审稿专家 2#

引言存在的问题：
1. 引言中指出现有人机协同表现不佳，可能源于设计未能有效结合双方潜在优势（Vaccaro et al., 2024），但讨论部分并未充分回应这一问题，尤其是对 LLM 与人类在决策上的各自优势缺乏明确总结。建议在讨论中针对研究结果，明确阐述二者的优势对比。

回复：感谢您的宝贵建议，我们在修订稿中根据具体的研究结果将人类与 LLM 的优势进行了总结，主要体现在讨论 4.2 部分（具体请见修订稿）。此外，我们增加了 4.3 理论价值和实践启示部分，进一步回应“有效结合双方潜在优势”的问题：

“在实践启示方面，第一，指导优势互补型人机协同系统的设计：有必要基于决策主体的优势，进行角色定位与任务分配。例如，在系列决策情境中，由 LLM 处理对估值精度与稳定性要求较高的任务；由人类负责长期、模糊经验的整合，以及在动态情境下的实时策略调整。第二，指出双向提升人与 LLM 的决策能力路径：一方面，人类可以通过基于 LLM 的训练程序，校准对收益/损失信号的感知与评估，也可以通过类似于对 LLM 进行 CoT 提示的方式，利用结构化的出声思维报告来提升决策表现 (Ericsson & Moxley, 2011)；另一方面，LLM 在设计上下文记忆机制时，或可借鉴人类的“躯体标记”或“要义表征”机制来整合信息，利用关键的情绪线索和规律信息优化注意力机制，提升决策质量 (Lee et al., 2024)。上述启示尚需通过具体实验设计，以明确其效果和机制。”

2. 引言中多次引用 Vaccaro et al., 2024，但该文的 meta 分析对象并非仅限于 LLM。建议在引言第三段补充对传统 AI 与 LLM 的差异性说明，并引入相关文献支持。

回复：感谢审稿专家的细致严谨的建议。我们在修订稿中对引言的相关段落进行了重组和补充。具体而言，我们在第二段中明确指出 Vaccaro 等人研究所涵盖的人工智能算法范围更广，包括机器学习与其他深度学习模型，而不仅是 LLM。随后，我们补充阐述了 LLM 相较于传统人工智能模型的区别。紧接着，我们引用了 Filippas 等（2024）将 LLM 比喻为“硅基人类”的表述，梳理了 LLM 实现决策的三个方面：（1）类人化的语言理解与生成能力；（2）强大的上下文记忆功能；（3）推理能力涌现。我们相信这些补充有助于清晰界定 LLM 与传统 AI 的差异，也更好地说明本研究聚焦比较人类与 LLM 系列决策表现的意义。具体请见修订稿引言第二段或如下：

“Vaccaro et al. (2024)的元分析指出，人类与人工智能在决策任务表现的优劣尚无定论。值得注意的是，该研究所涵盖的人工智能算法不局限于 LLM，而是包括了机器学习和其他深度学习模型。LLM 是一种建立在 Transformer 架构之上、通过大规模语料训练，具备语言理解与生成能力的生成式人工智能系统，与传统的人工智能在能力和交互方式上存在根本差异 (Brown et al., 2020; Vaswani et al., 2017)。Filippas et al. (2024)将 LLM 比喻为“硅基人类”，强调其执行复杂决策任务的能力。LLM 实现决策行为主要源于以下三个方面：首先，LLM 拥有与人类高度相似的语言理解和生成能力，这种相似性甚至在神经表征层面也有所体现 (Tuckute et al., 2024)。第二，LLM 具有强大的上下文记忆功能 (Goldstein et al., 2020; Mischler et al., 2024)，为处理需要整合历史信息的系列决策任务奠定了基础。第三，推理是 LLM 实现决策的核心能力 (Evans et al., 1993)，研究表明，当 LLM 参数达到一定规模时，推理能力开始涌现 (Wei, Tay, et al., 2022)。Wei, Wang, et al. (2022)揭示了这种推理通常借助“思维链”(Chain-of-Thought, CoT)——即一系列自然语言中间步骤——来实现或者增强。Wang and Zhou (2024)提供的证据表明，随着模型能力的迭代增强，LLM 已能在无 CoT 的情况下，有效完成推理任务。”

3. 对任务情境的生态效度缺乏反思。IGT 任务高度结构化且封闭，与真实人机协同决策情境存在差异，建议在讨论中分析本研究结果在开放性、多目标、动态任务中的适用性与潜在局限。

回复：感谢审稿专家的宝贵建议，我们在讨论结果时强调了结果可能受到 IGT 高度结构化特点的局限。例如，“值得一提的是，该结论依赖于研究中 IGT 的结构化和重复决策特征，在诸如科研或创造性活动等更为开放的决策领域，探索往往是专家的核心能力(Morgan et al., 2023)。”。

此外，我们在 4.4 局限和展望部分明确指出了该局限，即研究发现依赖于 IGT“特定的、高度结构化的系列决策任务设计”。LLM 展现出的高策略稳定性在面对“更为复杂、开放和动态变化的决策情境时”，可能会表现出“生态脆弱性”。在展望部分，手稿建议未来的研究应当操纵决策任务的“动态性、复杂性、开放性”，以更清晰地界定人类与 LLM 各自的优势边界。具体请见讨论 4.4 部分或如下：

局限：

“最后，本研究的发现依赖于 IGT 特定的、高度结构化的系列决策任务设计。LLM 体现出较高的策略稳定性，能够有效使用探索利用策略。然而，在面对动态变化或更为复杂、开放和的决策情境时，可能体现出“生态脆弱性”(Bender et al., 2021)。例如，当环境规则变化或需要权衡多个目标时，LLM 可能难以有效应对新情况，而人类的探索行为和随机性策略反而更具优势。”

展望：

“最后，未来研究可以考虑操纵决策情境的特点，如决策任务的动态性（变化的奖惩概率）、复杂性（多重目标权衡）和开放性（引入新的任务机制），比较在不同特点的决策情境中人类与 LLM 的决策机制和表现的差异，更加清晰地界定各自的优势边界，为理解和构建不同决策情境下的人机协同系统提供实证基础。”

方法上存在的问题

1. LLM 的实验参数：应补充模型调用时的详细参数（如 temperature、top-p 等），以便评估其对输出稳定性与探索行为的影响。如不同的 temperature 意味着 LLM 有不同的变化。

回复：非常感谢您提出的宝贵意见。我们已在稿件中方法 2.2 部分进行了相应的补充：“在模型调用参数设置上，本研究统一采用了各个模型官方 API 的默认值，旨在反映模型在标准配置下的自然决策行为。具体而言，调用 Deepseek-V3 模型时，其温度默认为 1.0；调用 GLM-4-Flash 和 GLM-Zero-Preview 模型时，其温度默认为 0.7。”

2. LLM 先验知识干扰：尽管设计了牌组匿名与顺序平衡，但 LLM 可能仍能推断出任务类型。建议在实验结束后对 LLM 进行操纵性检验（如询问其识别任务类型的能力），以增强结论的说服力。

回复：感谢这一关键建议。为检验“先验知识干扰”情况，我们新增了两个操纵性检验机制：其一，CoT 事后检查：本研究对手稿中报告的三个 CoT 条件的 LLM 进行了逐被试、逐试次的 CoT 文本分析。其二，自陈操纵性检查：遵循您的建议，我们在返修期间（9 月 10 日）调用了 Deepseek-chat（基线条件，N = 12）和 GLM-4-Flash（CoT 条件，N = 12）作为代表，在实验（与手稿中的流程相同）结束后通过询问其对任务的理解、使用的策略和是否使用了外部知识，实现操纵性检查。具体请见补充材料：

“操纵性检查

为检验“先验知识干扰”情况，我们采取了两个操纵性检验的方法：其一，CoT 事后检查：本研究对手稿中报告的三个 LLM 进行了逐被试、逐试次的 CoT 文本检验。其二，自陈操纵性检查：遵循您的建议，我们调用了 Deepseek-chat（基线条件，N = 12）和 GLM-4-Flash（CoT 条件，N = 12）作为代表，在实验（与手稿中的流程相同）结束后通过询问其对任务的理解、使用的策略和是否使用了任务外的先验知识，实现操纵性检查。

CoT 事后检验：我们首先对 CoT 条件的每个模型（Deepseek-V3、GLM-4-Flash、GLM-Zero-Preview）下全部被试（48 名）与全部试次（100 试次）的 CoT 文本（共 3×48×100=14400 条）做了关键词检索（以“Iowa Gambling Task” “IGT” “Iowa”作为探查词）。接着，我们调用 GLM-4-Flash 对 14400 条 CoT 文本逐条进行独立判别。判别提示语要求模型仅依据单条 CoT 文本回答：是否在该轮推理中推断出具体任务范式（如 Iowa Gambling Task/IGT）。我们要求 GLM-4-Flash 回应仅仅输出“是/否”，并以标准化 JSON 返回结果进行统计。最终，我们请作者列表外的两位心理学专业在读博士生对判别为“是”的 CoT 文本进行人工复核，请他们判断该 CoT 是否涉及爱荷华赌博任务的相关描述，以及是否引用先验知识。

研究结果发现未检索到关键词；LLM 判别结果如下表 S6 所示，在 14400 条 CoT 中，共发现 49 条 CoT 被判别为“是”；人工复核后，两位心理学专业在读博士生均认为这 49 条 CoT 中不存在先验知识干扰的情况。LLM 判别偶有误报，可能由于 CoT 中包含一些基于常识的推理（风险—回报、平衡、探索、假设等）或者基于 LLM“幻觉”做出的假设，误当作范式相关内容或者先验知识。我们也给出了两条 CoT 示例（表 S6）供您和读者审查。

表 S6- CoT 文本独立判别结果

模型	人数	总试次	识别“是”（次数）	占比	存在“是”的被试	人工复核通过
Deepseek-V3	48	4800	0	0.00%	0	/
GLM-4-Flash	48	4800	47	0.98%	7	47
GLM-Zero-Preview	48	4800	2	0.04%	2	2

表 S7-被判别为“是”的两条 CoT 示例

模型	被试	试次	CoT
GLM-4-Flash	Sub3	4	After my third selection, which was Ijp, resulted in a gain of 80 and a loss of 150, my balance is now 2150. Here is my reasoning for the next round selection: 1. Ijp: Although Ijp has a history of gains, the most recent round was a loss, which brings the overall gain from Ijp to 280 (from two previous gains of 100 and 120) minus the loss of 150, resulting in a net loss of 70. Given this, Ijp is no longer the deck with the highest net gain. However, it still has the highest chance of gain at 60%, which makes it a potential long-term strategy. 2. NHj: NHj offers a balanced 50% chance of gain and 50% chance of loss. It doesn't have the highest chance of gain, but it also doesn't carry the same level of risk as hIt or nLO. If I'm looking to minimize the risk of a significant loss, NHj could be a safer bet. 3. hIt: hIt has a 40% chance of gain and 60% chance of loss, which is less favorable than Ijp but also less risky. It has the potential to provide some gains, albeit less frequently. 4. nLO: nLO has the lowest chance of gain (30%) and the highest chance of loss (70%). This deck is the riskiest and should generally be avoided unless there is a compelling reason to choose it. Given that Ijp has not been consistently profitable in the last round, and considering the potential for significant losses, I will shift my strategy to a more balanced

			approach with NHj. This choice allows me to maintain a steady increase in balance without taking on the higher risk associated with hIt or nLO. My selection is 'NHj'.
GLM-Zero	34	1	As I look at the four decks of cards labeled ILl, ysm, KzR, and eQp, I'm trying to determine which one will maximize my balance. Since I have no prior outcomes to guide me, I need to make an initial assumption about the potential gains and losses associated with each deck. Wait, perhaps I should consider the labels themselves. ILl, ysm, KzR, and eQp don't seem to provide any obvious clues about the nature of the cards. They might be arbitrary labels, or they could represent something more meaningful. However, without additional information, I can't make any assumptions based on the labels alone. Maybe I should think about the possible gains and losses associated with each card. Since I have no information about the specific rules for gaining and losing, I'll have to assume that each card has a different probability distribution of gains and losses. Perhaps some cards have higher potential gains but also higher potential losses, while others have lower gains and losses. Wait, but without any data or patterns to analyze, I can't make any informed decisions based on probability. I need to gather some information first. Maybe I should select a card at random and see what happens. That way, I can start to build a understanding of the potential outcomes associated with each deck. However, I have a loan of 2000 play money, and I need to maximize my balance. Selecting a card at random could lead to losses that I can't afford. Maybe I should try to minimize my potential losses while still seeking gains. Wait, but how can I minimize my losses if I have no information about the cards? Perhaps I should spread my selections across the different decks to gather information about each one. That way, I can start to identify which decks have better potential for gains and which ones have higher risks.

However, I only have one selection to make in this round. I can't spread my selections across multiple decks. I need to choose one card and see what happens.

Wait, perhaps I should choose the deck that seems the most balanced. Since I have no information about the cards, I'll assume that each deck has an equal chance of providing gains and losses. Therefore, I'll choose the deck labeled ILl.

My selection is 'ILl'.

自陈操纵性检查：我们在 2025 年 9 月 10 日（返修阶段）调用 Deepseek-chat（基线条件，$N = 12$）和 GLM-4-Flash（CoT 条件，$N = 12$）分别作为基线和 CoT 条件的代表，在实验（与手稿中的流程相同）结束后询问其三个问题并收集 LLM 回应，以判断 LLM 是否利用了先验知识。我们的问题是：“请用一句话说明你认为这个实验在测试什么”、“请用一句话概括你在本任务中的选牌策略”、“在完成任务时，你是否使用了任务外的先验知识？请从两项中选择其一并简要说明理由：Exclusively in-game 或 Used outside knowledge” 陈述结果如下表格所示，我们请作者列表外的两位心理学专业在读博士生根据回答判断是或否调用了先验知识。结果如下表所示，所有的被试均未在自陈操纵性检查中被人工判定为未使用外部知识（否）。

表 S8-自陈操纵性检查结果

模型	被试	测试内容	使用策略	信息调用	人工判断
GLM-4-Flash	1	The goal was to maximize my balance by selecting cards that consistently provided gains with minimal or no losses.	I preferred decks that showed consistent positive outcomes and avoided those with unpredictable or large losses.	Exclusively in-game	否
GLM-4-Flash	2	The goal was to maximize my balance by selecting cards that consistently provided gains with minimal or no losses.	I selected cards that initially showed consistent gains and avoided those with sudden large losses.	Exclusively in-game	否
GLM-4-Flash	3	The goal was to maximize my balance by selecting cards that consistently provided gains with minimal or no losses.	I preferred decks that showed consistent gains and avoided those with unpredictable large losses.	Exclusively in-game: My strategy was based solely on the gains and losses observed during this session.	否

GLM-4-Flash	4	The goal was to maximize my balance by selecting cards that consistently provided gains with minimal or no losses.	I preferred decks that showed consistent gains and infrequent losses based on observed outcomes.	Exclusively in-game: My strategy was based solely on the gains and losses presented during this session without any external information.	否
GLM-4-Flash	5	The goal was to maximize my balance by selecting cards that consistently provided gains without large losses.	I switched decks after experiencing losses to explore more stable options.	Exclusively in-game	否
GLM-4-Flash	6	The goal was to maximize my balance by selecting cards that consistently provided gains without losses.	I selected cards that had shown no losses until a loss occurred, then switched to another deck.	Exclusively in-game	否
GLM-4-Flash	7	The goal was to maximize my balance by selecting cards that consistently provided gains with minimal or no losses.	I preferred decks that showed consistent gains and avoided those with unpredictable or large losses.	Exclusively in-game: My strategy was based solely on the gains and losses observed during this session.	否
GLM-4-Flash	8	The goal was to maximize my balance by selecting cards that consistently provided gains with minimal or no losses.	I preferred decks that showed consistent gains and avoided those with unpredictable or large losses.	Exclusively in-game	否
GLM-4-Flash	9	The goal was to maximize my balance by selecting cards that consistently provided gains with minimal or infrequent losses.	I repeatedly selected zVo due to its consistent gains and low losses.	Exclusively in-game	否
GLM-4-Flash	10	The goal was to maximize my balance by selecting cards that consistently provided gains with minimal or no losses.	I consistently selected bjT after observing its reliable gains and infrequent small losses.	Exclusively in-game: My strategy was based solely on the gains and losses observed during this session.	否

GLM-4-Flash	11	The goal was to maximize my balance by selecting cards that consistently provided gains with minimal or no losses.	I preferred decks that showed consistent positive net outcomes and avoided those with unpredictable large losses.	Exclusively in-game: My strategy was based solely on the gains and losses observed during this session.	否
GLM-4-Flash	12	The goal was to maximize my balance by selecting cards that consistently provided gains with minimal or no losses.	I alternated between Zin and kuE based on recent outcomes to avoid repeated losses.	Exclusively in-game: My strategy was based solely on the gains and losses observed during this session.	否
Deepseek-V3.1	1	The card with the highest potential for gains while minimizing the risk of significant losses.	Select cards based on their historical performance and frequency of selection to balance gains and risks.	Exclusively in-game. The strategy was based solely on the gains and losses presented during the game session, without using any external information.	否
Deepseek-V3.1	2	The goal was to maximize the balance by selecting cards from four decks with different rules for gaining and losing.	I rotated between decks to avoid overreliance and to maintain a diverse strategy, selecting the deck with the highest potential gain based on previous outcomes.	Exclusively in-game	否
Deepseek-V3.1	3	The goal of the game is to maximize the balance by selecting cards from four decks with different rules for gain and loss.	I selected cards based on the historical performance of each deck, aiming to maximize gains and minimize losses.	Exclusively in-game. My strategy was based solely on the gains and losses presented during the game session, without using any external information.	否
Deepseek-V3.1	4	The best strategy was to select the deck with the most consistent positive outcomes.	I continued to select the deck that provided the most consistent gains without losses.	Exclusively in-game. My strategy was based solely on the gains and losses presented during the game session.	否
Deepseek-V3.1	5	The goal of the game was to maximize the balance by selecting cards from four different decks with varying rules for gains and losses.	I alternated between cards based on their historical performance and risk/reward profiles, aiming to balance gains and minimize losses.	Exclusively in-game. My strategy was based solely on the gains and losses presented during the game session, without using any external information.	否

Deepseek-V3.1	6	The card with the highest expected value should be chosen to maximize the balance.	Select cards based on their historical performance and potential for gains.	Exclusively in-game	否
Deepseek-V3.1	7	Maximize the balance by selecting cards with the highest potential for gains and lowest risk of losses.	Select the card that has provided the most consistent gains without losses, while considering the risk of streaking and the potential for significant losses.	Exclusively in-game. The strategy was based solely on the gains and losses presented during the game session, without using any external information.	否
Deepseek-V3.1	8	The card with the highest potential gain and lowest risk of loss was selected to maximize the balance.	Select the card with the highest average gain and lowest risk of loss.	Exclusively in-game	否
Deepseek-V3.1	9	The goal was to maximize the balance by selecting cards with the best gain-to-loss ratio.	I chose cards based on their consistent performance and low risk of loss.	Exclusively in-game	否
Deepseek-V3.1	10	The card with the highest historical gain with no losses should be selected next.	Diversify selections among different cards to manage risk and maximize gains.	Exclusively in-game. The strategy was based on the gains and losses presented during the game session, without using any external information.	否
Deepseek-V3.1	11	The card with the highest average gain over the course of the game was selected.	Specialized in the card that provided the most consistent gains without losses.	Exclusively in-game. The strategy was based solely on the gains and losses presented during the game session, without using any external information.	否
Deepseek-V3.1	12	The goal was to maximize the balance by selecting cards with the best risk-reward profile.	I selected cards based on their historical performance and the risk associated with each.	Exclusively in-game. My strategy was based solely on the gains and losses presented during the game session.	否

综合上述操纵性检查证据，我们认为原始实验设置中牌组匿名化的操作是有效的，“先验知识干扰”对结论的威胁可以忽略不计，LLM 在根据我们的实验要求和反馈进行探索和信息利用。

3. 人类与 LLM 的比较分析策略：既然区分了一般人类与专家级人类，建议额外提供“去除专家级人类被试后的人类样本与 LLM 的比较结果”，以便更清晰地呈现一般人类与 LLM 的差异。

回复：感谢您的建议。为了更清晰地呈现一般人类与 LLM 的决策差异，我们在补充材料中新增了表格 S4。展示了全体、专家级、全体去除专家级人类样本在六个指标（最终净收益、价值敏感度、逆衰减率、探索学习率、探索增益、稳定性参数）的均值(*M*)、标准差(*SD*)和样本量(*N*)。此外，表格还提供了专家级样本和全体去除专家级样本之间的差异，并进行了独立样本 *t* 检验两组差异。结果表明，专家级样本与全体去除专家级人类样本在多个关键决策指标上存在显著差异，尤其在最终净收益、价值敏感度、逆衰减率和稳定性方面，专家级样本表现出显著的优势。探索学习率在两个样本间未见显著差异，而专家级样本的探索增益表现显著低于全体去除专家级人类样本，提示专家级样本在决策时可能更倾向于稳定、精确的评估而非广泛探索。

表 S5 全体、专家级、全体去除专家级样本在各指标上的描述性统计量与独立样本 *t* 检验结果

参数	人类样本	*N*	*M*	*SD*	差异	*t*	*df*	*p*
最终净收益	全体	1763	-1277.15	1034.49				
	专家级	48	1457.71	665.03	2811.41	28.51	1761	<0.001
	全体-专家级	1715	-1353.70	934.17				
价值敏感度	全体	1763	0.28	0.19				
	专家级	48	0.50	0.28	0.22	5.56	1761	<0.001
	全体-专家级	1715	0.28	0.18				
逆衰减率	全体	1763	0.61	0.26				
	专家级	48	0.78	0.20	0.18	6.04	1761	<0.001
	全体-专家级	1715	0.60	0.26				
探索学习率	全体	1763	0.41	0.21				
	专家级	48	0.45	0.27	0.04	0.94	1761	0.351
	全体-专家级	1715	0.41	0.21				
探索增益	全体	1763	1.82	2.41				
	专家级	48	-0.96	2.48	-2.85	-7.88	1761	<0.001
	全体-专家级	1715	1.89	2.36				
稳定性参数	全体	1763	0.67	0.42				
	专家级	48	0.55	0.32	-0.12	-2.63	1761	0.011
	全体-专家级	1715	0.68	0.42				

注释：差异表示专家级样本均值减去全体去除专家级样本均值。*t* 检验结果比较的是专家级样本和全体-专家级样本的差异。

4. 专家级人类样本的选取标准：请进一步详细说明专家级人类样本的选取标准，可以在附录中进行数据展示等。

回复：感谢您的宝贵建议。

我们已经在修订稿中对“专家级”人类样本的筛选标准及人数匹配策略进行了澄清：“借鉴现有研究中将 LLM 与普通人群或领域专家进行比较的不同策略(e.g., Lu, 2025; Luo et al., 2025)，本研究并行执行了两种比较方案：(1)全体人类样本与 LLM 比较：将全部 1763 位人类被试与 LLM 进行对比分析。(2)“专家级”人类样本与 LLM 比较：根据最终净收益指标，从人类被试中筛选出净收益表现最佳的 48 位被试（与 LLM 被试数量匹配），形成“专家级”人类样本，并与 LLM 进行对比分析。”

关于数据展示：为了更直观地展示该样本的特征，我们在补充材料的表格 S4 中（请见上个回复），详细列出了专家级样本在模型参数的数据情况，并与全体去除专家级样本进行了对比。

5. LLM 的选取：请作者进一步阐述 LLM 选取标准，因为在大众普遍认知，或者是目前的一些榜单的得分情况，主流的 LLM 仍然包括 Chatgpt 因此如果作者能够进一步详细的阐释选取标准，能够更加具有说服力。

回复：感谢您的建议。

首先，本研究的目标并非得到国际范围内的 LLM 的决策表现排名，而是比较 LLM 与人类在 IGT 中的决策机制和参数差异。因此，在实验设计时，我们考虑了性能、开源、可用性（Ivanova, 2025），在中国大陆的合规可访问范围内，选取了：权重开源的 Deepseek-V3，低成本低时延的 GLM-4-Flash（2025-01-16 起智谱平台提供 Flash 免费 API 调用），以及推理导向的 GLM-Zero-Preview。根据 SuperCLUE 年度报告（2025-01-08），我们选择的模型 Deepseek-V3 和 GLM 系列在性能上具有代表性。Deepseek-V3 以总分 68.3 分位于第一梯队，特别是在 Hard 维度（如 Agent 任务 75.0 分、深度推理 58.8 分）和文科维度（如语言理解 86.5 分）表现出色，接近 ChatGPT-4o-latest（70.2 分），且在推理速度和性价比方面具有竞争力。GLM 系列以 65+分位居第一梯队，多样化的模型策略也覆盖了决策推理的现实场景。

此外，针对每种 LLM，我们同时设置了基线与 CoT 两种提示策略：基线旨在最大程度贴近人类被试流程（仅做选择、无需外化推理）；CoT 则依据既有研究显示“显式推理外化可提升复杂推断任务表现”的结论，在每轮选择前要求模型以固定格式简要外化推理，再给出选择。

最后，在局限的讨论中，我们也强调了在模型代表性上可能的局限和未来的方向：“为克服上述局限，未来研究首先应扩展模型的范围至更全面的 LLM，探索不同模型、不同条件之间的决策表现异同。”

结果上存在的问题
1. 方差差异解释：图 1 显示人类组的方差明显小于 LLM 组，建议在讨论中加入对此现象的可能原因分析。

回复：感谢审稿专家指出这一重要问题。需要澄清的是，图 1 中累计余额和探索行为的曲线阴影表示的是标准误，其目的是展示均值估计的不确定性，而非组内离散程度。由于人类样本量（N =1763）远大于 LLM 条件（N=48），人类条件下的阴影（标准误）看起来小于 LLM 条件。我们在修订稿中添加了对图 1 的注释："注：阴影区域表示均值的标准误差"

同时，我们理解审稿专家对于关键指标分布情况的关切，在进行关键指标的可视化时，采用小提琴图展示真实分布：如最终净收益（图 2），VSE 模型参数（图 3）。

2. P 值与多重比较校正：由于样本量大，需说明是否进行了多重比较校正及所用方法，以保证结果严谨性（也许可能是我并没有看到，但是为了结果的严谨性我需要进一步核对）。

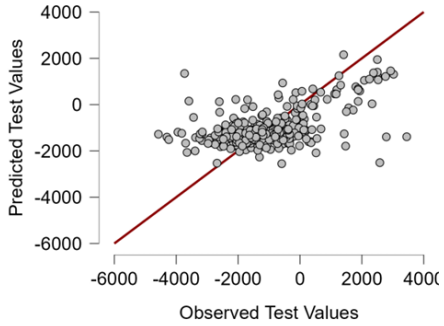
回复：感谢审稿专家的严谨。需要澄清的是，我们确实对研究中涉及的多重比较问题进行了处理。在进行单因素方差分析后，采用了 Games-Howell 方法进行事后检验（如正文 2.4.2 节所述）。Games-Howell 检验是一种专门用于处理多组比较的程序（在表 2，S2, S3, S4 中报告为 p_{adj}）。我们选择该方法也是因为它对各组方差不齐的情况具有稳健性，更适合本研究的数据特征。

我们在修订稿 2.4.4 部分阐述了事后检验的方法："采用单因素方差分析（Games-Howell 事后检验）进行人类与 LLM 的 VSE 模型参数差异检验"替代了原稿的"具体的比较统计方法与 2.3.2 中的描述一致"。此外，我们在表 2，S2, S3, S4 中添加注释："p_{adj} 为经过 Games-Howell 多重比较校正后的 p 值"。

3. 探索性分析：作者具有如此大的数据量，可以进一步的做一些探索性分析，例如 Steyvers et al. (2022)将人类被试和 AI 的结果进行了混合建模，预测了最终的分类结果，如果作者能够在加入一些 LLM 和人类的混合预测结果，如预测最终收益率。是否可以进一步阐释本文所提出的 LLM 与人类各自的优势，二者进行混合或许可以带来更好的结果。

回复：非常感谢您对稿件的细致审阅和提出的宝贵意见。

我们完全同意，这是一个极具研究价值的方向，能够有效探索 VSE 参数对最终收益的影响。基于您的建议，我们进行了一项探索性分析：融合了所有人类和 LLM 被试（包括 1763 名人类与 192 个 LLM 被试）的数据，构建了一个随机森林回归模型，探索哪些 VSE 参数能够预测最终净收益。首先，我们对 VSE 参数特征进行了 Z 标准化处理,并对数据进行了划分：随机抽取 20%数据作为测试集用于最终模型的性能评估，然后在剩余数据中随机抽取 20%数据作为验证集，其余部分则构成训练集。在模型训练阶段，在 1 至 100 棵树的范围内，根据验证集上的袋外误差自动寻找最优树数量以防止过拟合。在构建每一棵决策树时，算法会从训练集中通过自助采样法随机抽取 50%的样本，并在每个节点分裂时自动选择候选特征进行评估。



排序	特征	平均精度下降
1	价值敏感度	675457
2	稳定性	600329
3	探索增益	362342
4	逆衰减率	261266
5	探索学习率	105180

分析结果显示：如图，该模型表现出中等的预测能力，能够解释最终净收益中约 27.9%的变异（$R^2 = 0.279$）。如表，根据平均精度下降（Mean Decrease in Accuracy）指标，对五个 VSE 参数的重要性进行了排序。该结果提示我们价值评估能力（高价值敏感度）与策略一致性（高稳定性）是 IGT 收益的关键因素。尽管这一探索性分析深化了我们对 IGT 决策的理解，但我们从方法和论述逻辑的角度进行了两点考虑：（1）VSE 参数的估计与最终净收益的计算均基于 IGT 的选择序列，可能存在难以排除的共同方法偏差问题。（2）我们手稿的核心叙事逻辑在于利用同一计算模型框架，对比人类与 LLM 决策参数的差异。而这项探索性分析的视角虽然是基于前者的重要延伸，但将其直接并入，可能会分散文章的核心逻辑。因此，我们决定将这一部分作为对您宝贵意见的回应，不在手稿中呈现。再次感谢您的宝贵建议，这无疑为我们未来的研究工作开辟了新的方向。

次要问题 1.图表可读性：部分图表缺乏清晰的图例与条件说明（如颜色对应关系），建议完善图注。

回复：已经完善图例，且确保图 1，图 2，图 3 的颜色与被试群体对应始终保持一致且同模型的不同条件采用相近似的颜色，如 Deepseek 基线是红色，CoT 条件是橙色。

2. 术语解释：部分技术术语（如探索增益、稳定性参数）虽在公式中出现，但初次提及时建议补充简明定义，以方便跨领域读者理解。

回复：感谢宝贵建议，我们对首次出现的参数添加了方便读者理解的描述如：“用来衡量信息保存程度的逆衰减率”、“反映探索值回到基线速度的探索学习率”、“反映探索基线大小的探索增益参数”、“衡量选择随机性的稳定性参数”。

3. 补充材料可获取性：在正文中简要概述提示词设计原则，避免读者在未获取补充材料时缺失关键信息。

回复：感谢宝贵建议，在正文中简要概述提示词设计原则，避免读者在未获取补充材料时缺失关键信息。“提示词的设置原则是尽量模拟人类被试的指导语和实验流程，具体提示词请见补充材料。”

4. 英文摘要优化：部分长句可拆分以提升可读性，并保持 VSE 模型名称与全文一致。

回复：感谢审稿专家的建议。我们已对英文摘要进行可读性优化，并统一术语表述，具体如下：

Who’s the Better Gambler? Comparing Humans and Large Language Models on Exploration–Exploitation in the Iowa Gambling Task

Abstract：Understanding the differences between the sequential decision-making mechanisms of the Large Language Model (LLM) and humans is crucial for optimizing human-AI collaboration. This study employed the Iowa Gambling Task (IGT), combined with Value plus Sequential Exploration (VSE) cognitive modeling, to systematically compare behavioral performance between humans and LLM. We contrasted a large human sample with three representative LLMs

under baseline and Chain-of-Thought (CoT) conditions. We hypothesized that the VSE model could effectively characterize the exploration and exploitation parameters in samples including both humans and LLMs.

The study utilized a clinical version of the IGT with dynamic reward/loss schedules. Data were collected from 1763 undergraduate students via a standard computerized procedure. LLM data were collected via the official APIs of three models (DeepSeek-V3, GLM-4-Flash, GLM-Zero-Preview), simulating 48 participants per model for both baseline and CoT conditions. To prevent LLMs from leveraging pre-existing knowledge about the IGT task acquired during their training, the standard deck labels (A, B, C, D) were replaced with random three-letter strings as anonymized labels. Furthermore, the assignment of the underlying deck characteristics (i.e., their reward and punishment schedules) to the anonymized labels was counterbalanced across participants using a Latin square design. LLM interactions involved system prompts delivering instructions and user prompts requesting choices. They incorporated dialogue history and feedback on gains/losses from the previous trial. An error-correction mechanism ensured valid responses. Cognitive modeling compared the VSE model against four alternative models using Variational Bayesian Analysis for parameter estimation and model fitting with non-informative priors. A parallel analysis approach was adopted. We conducted comparisons between the LLMs and (1) the entire human sample, and (2) the top-performing (based on final net gain) 48 “expert” human participants, matched to the LLM sample size.

Behavioral results indicated that certain LLM conditions (baseline GLM-4-Flash and GLM-Zero-Preview) failed to follow IGT task instructions. They exhibited repetitive non-adaptive patterns and were thus excluded from further analysis. Well-performing LLMs (Deepseek-V3 and GLM-Zero-Preview) achieved significantly higher final net gains compared to the overall human sample, but their performance showed no difference from that of the "expert" human sample. A thoroughgoing investigation involving model comparison analyses robustly identified the VSE model as the superior framework for explaining both humans and LLM choice sequences across two parallel analyses. This conclusion is evidenced by better-fit indices (BIC, AIC, Free Energy) and higher predictive accuracy for choices. Parameter comparisons derived from the VSE model revealed key differences. Regarding value sensitivity (accuracy in perceiving reward/loss utility), LLMs showed greater sensitivity than the general human population, although comparable to that

of expert humans. Conversely, human participants always showed a superior ability to integrate historical value information over time, reflected in higher inverse decay rate parameters compared to LLMs. Differences in exploration-related parameters (learning rate and gain) were also observed, indicating distinct exploration tendencies between the groups and across two parallel analyses. Finally, LLMs consistently demonstrated significantly higher decision stability (consistency) than each human sample.

In conclusion, this research demonstrates the utility of combining the IGT, cognitive modeling, and a multiverse analysis framework to dissect the nuances of sequential decision-making in both humans and LLMs. While high-performing LLMs can match expert human outcomes in the IGT, their underlying cognitive parameters diverge significantly. Key distinctions include the LLMs' heightened value sensitivity (akin to human experts) and greater strategic consistency, contrasted with humans' superior integration of past outcomes, potentially linked to affective systems like somatic markers absent in LLMs. The VSE model provides a unified computational lens to examine these exploration-exploitation dynamics. These findings highlight the unique cognitive profile of LLMs. They offer insights for developing targeted training to enhance decision-making. They also inform the design of more effective human-LLM collaborative systems that capitalize on the distinct strengths of each agent, while acknowledging potential LLM limitations in less structured environments.

Keywords: Iowa Gambling Task, Large Language Model, Value plus Sequential Exploration Model