# 《心理科学进展》审稿意见与作者回应

题目：刻板印象的计算认知机制：社会学习与泛化
作者：王滢洁，张洳源

---

## 第一轮

**审稿人 1 意见：**

本文从计算认知神经科学视角出发，梳理了刻板印象从习得到泛化的认知机制，创新性地整合强化学习与贝叶斯理论两大主流计算框架，并提出了社会认知地图、关系线索整合与大语言模型模拟三个具有启发性的未来研究方向，体现了较强的理论构建能力与学术前瞻性。相较以往多停留于理论描述层面的研究，本文引入计算认知框架，为该领域提供了新的解释路径，在理论整合、视角拓展与前瞻建议方面具有明确贡献。然而，文章在概念界定、理论整合深度、研究逻辑、文献支撑与写作细节上仍存在若干可进一步完善之处，具体总结如下：

**意见 1**：概念界定：

a) 对“刻板印象”这一核心概念的界定在不同理论框架下（如强化学习中的“有偏先验”与贝叶斯理论中的“群体条件概率”）表述较为分散，缺乏贯穿全文的、整合性的操作性定义，一定程度上影响了理论论述的连贯性。

**回应**：感谢审稿专家的指正。修改稿在第 2 节“社会学习”的开头，明确提出刻板印象的核心在于“某群体具有某特质”这一信念结构，并将其分解为“某群体从何而来”和“具有某特质如何习得”两个子问题，分别对应贝叶斯理论与强化学习两大框架。在 2.2 节进一步引入联想表征与命题表征的区分，明确了两大框架在联结习得阶段各自对应的认知基础：强化学习中表现为价值函数 V(群体)（联想表征），贝叶斯框架中表现为条件概率 P(特质 | 群体)（命题表征）。通过这一安排，概念界定与理论分析的展开融为一体，不同框架下的表述均可回溯至同一概念核心。

**意见 2**：概念界定：

b) 研究范围的指向略显模糊。自检报告中提出要解决“习得、表征到泛化全过程”的机制，但正文对“表征”的专门论述不足，且“形成”“更新”“泛化”等阶段在论述中边界不够清晰。建议进一步明确本文的核心是侧重刻画印象的“动态形成与更新过程”，还是“形成后稳定表征的泛化应用机制”。

**回应**：修改稿对研究范围进行了明确界定，将全文聚焦于“社会学习”（形成与更新）和“社会泛化”（应用）两个核心环节。在引言中明确指出：“刻板印象的形成与应用可以分解为两个核心环节：社会学习——个体如何通过经验获取关于群体的知识；以及社会泛化——个体如何将有限经验推广到新情境或未知个体。”相应地，全文结构调整为：第 2 节“社会学习”涵盖群体发现、联结建立和更新/固化三个子过程，第 3 节“社会泛化”涵盖群体识别和联结知识检索两个子过程，各阶段边界更为清晰。

**意见 3**：理论整合

a) 文章将强化学习与贝叶斯理论并列为两大框架，但未深入探讨二者是互补、竞争还是适

用于不同情境，亦未论及其潜在联系（如贝叶斯强化学习）。目前的论述更接近于并列介绍而非有机整合，削弱了理论模型的整体解释力。

**回应**：这是修改稿着力改进的核心问题。主要从三个层面实现了更深入的整合：

（1）通过联想表征与命题表征的区分框架，明确了两大理论各自的适用范围：强化学习侧重于联想表征的形成（群体-价值的自动化联结），贝叶斯理论侧重于命题表征的推断（群体-特质的概率信念），两者不是竞争关系，而是从不同层面解释刻板印象的不同面向（见 2.2 节）。

（2）通过 APE 模型框架，阐明了两种表征（及其背后的计算框架）的交互机制：联想激活可为命题推理提供输入，命题信念也可约束联想反应（见 2.2.3 节）。

（3）明确了两大框架的交汇点：在 2.3.3 节末尾论述了探索-利用困境如何体现贝叶斯先验对强化学习探索策略的引导，形成自我强化循环，并提出贝叶斯强化学习(Bayesian RL)作为未来统一框架的可能方向。

**意见 4**：理论整合

b) 从“习得形式”到“计算机制”再到“泛化路径”之间的逻辑过渡略显生硬，各环节间的因果关系与信息流向不够清晰。例如，条件反射习得的情感偏见如何通过强化学习机制进行价值更新？图 1 虽展示了框架轮廓，但模块间的动态交互关系未能充分体现。

**回应**：回复：修改稿从论述结构与框架图两个层面解决了这一问题：

（1）重组论述结构。新的结构按信息加工流程展开：群体发现（贝叶斯结构学习）→ 联结建立（经验通路/语言通路）→ 联结更新与固化（预测误差/探索-利用）→ 群体识别（知觉/功能线索）→ 联结知识检索与应用。各环节之间以明确的过渡段连接，例如 2.2 节开头“在发现了‘谁是一个群体’之后”、3 节开头“社会学习为刻板印象提供了信息基础，然而刻板印象的认知效力在于泛化”等，使信息流向更加清晰。

具体而言，针对审稿专家提出的“条件反射习得的情感偏见如何通过强化学习机制进行价值更新”这一问题，修改稿在 2.2.1 明确阐述了逻辑衔接：评价性条件反射、工具性学习和观察学习被统一归入“经验通路”，三者的共同特征是通过个体与环境的互动建立联想表征。在计算层面，强化学习为这一过程提供了统一的形式化描述——条件反射对应无模型学习（仅通过刺激-结果配对建立价值联结），工具性学习依赖行动-反馈循环调整价值函数 V(群体) ← V(群体) + αδ，观察学习则通过“价值塑造”机制继承他人的有偏先验。同时，2.2.2 将语言传播归入"语言通路"，区分于经验通路，明确语言通路主要承载命题表征 P(特质|群体)，其概率推断由贝叶斯框架刻画。2.2.3 进一步讨论了两种表征的双向交互机制。这样，习得形式（输入源）、计算机制（双通路的算法描述）和表征类型（联想/命题）三者之间形成了清晰的一一对应，信息流向得以厘清。

（2）重新设计框架图。将原有的框架流程图与各计算环节涉及的脑区进行了整合（见图 1）。新图以三行结构分别对应群体发现、群体-特质联结、联结检索与应用三个核心环节，左侧展示计算过程（包括贝叶斯结构学习、经验通路/联想表征、语言通路/命题表征、知觉线索、功能线索及其整合），右侧以外侧面观和内侧面观标注各环节涉及的关键脑区。通过这种将计算框架与神经基础并置的呈现方式，模块间的信息流向和动态交互关系在图中得到了更为充分的体现。

**意见 5**：理论整合

c) 文章对刻板印象在大脑中具体以何种形式表征（如概率分布、社会图式或认知地图坐标）论述不足。同时，对于价值更新（强化学习）与结构推断（贝叶斯）如何在神经层面协同运作，缺乏深入的神经环路或网络层面的证据与讨论，这与“计算认知神经科学”的定位尚有

距离。

**回应**：修改稿从两个层面加强了表征与神经机制的论述：

（1）表征形式：通过引入联想表征与命题表征的区分，明确了刻板印象以两种形式存储——自动化的价值联结 V(群体)和概率性的条件信念 P(特质 | 群体)，并在 2.2.3 节讨论了两种表征的交互。

（2）神经证据的系统整合：在总结部分（第 4 节）梳理了各计算过程涉及的神经环路——贝叶斯结构学习（rAI、dACC/MCC）、价值编码（vmPFC、纹状体、杏仁核）、特质推断（后扣带回、颞上沟、颞顶联合区）、预测误差更新（前外侧 PFC、STS）、线索整合（dmPFC）。关于审稿专家指出的与“计算认知神经科学”定位的距离，我们经过审慎考虑，认为这一批评是中肯的。本文的核心贡献聚焦于计算-算法层面的框架整合，上述神经证据主要作为各计算过程的支持加以引用，尚未实现计算模型与神经实现的对接。因此，修改稿将全文定位调整为“计算认知”视角（与标题“计算认知机制”一致），在摘要和引言中相应修正了措辞。同时，我们在展望部分明确指出，未来研究可借助 fMRI 结合表征相似性分析等技术，探索计算模型预测与神经表征之间的对应关系，推动从计算认知向计算认知神经科学的进一步发展。

**意见 6**：文献综述

a) 自检报告解释近 5 年文献占比较低，但计算认知神经科学领域发展迅速，2021-2025 年间应有更多相关研究可纳入。建议酌情补充最新实证研究，以增强综述的前沿性。

**回应**：修改稿新增了多篇 2021–2025 年间的重要文献（新补充的文献用红色字体在文献引用中标出），包括 De Houwer et al. (2021), Gawronski & Bodenhausen (2006, 2011), Bao (2024), Kurdi et al. (2023), Hedrich et al. (2024), Rösler et al. (2025), Allidina & Cunningham (2021), Gelpí et al. (2025), Frolichs et al. (2022), Gao et al. (2026), Liu et al. (2025), Babür et al. (2024), Kobayashi et al. (2022), Traast et al. (2025), Mahmoodi & Rushworth (2026) 等。近 5 年文献占比由原来的约 34%提升至约 44%。可在自检报告中看到当前的参考文献比例。

**意见 7**：文献综述

b) 文章主要综述支持计算观点的研究，对相关理论的批判性观点、计算模型自身的局限性（如生态效度、个体差异问题）涉及较少。在“习得形式”等部分，存在引用同一作者系列研究较多、对其他学者工作涵盖不足的情况，建议进一步明确文献筛选标准，并适当增加对替代性解释或争议性观点的评述，以体现综述的全面性与平衡性。

**回应**：修改稿在以下几处加强了批判性讨论：

（1）在 2.2.3 节引入了联想-命题表征争论的文献，指出“联想表征与命题表征的边界比传统双过程理论所假设的更加模糊”。

（2）在总结部分明确讨论了现有模型的局限：生态效度问题（赌博机范式的简化性）、个体差异问题（模型参数变异来源未明）以及两种表征交互机制的研究空白（见第 4 节“当前局限”部分）。

（3）在展望 LLM 部分增加了对其局限性的讨论：“LLM 反映的是语料统计特征而非人类认知过程”等。

**意见 8**：创新性

a) 作者宣称的三大创新点中，“整合框架”与“理性基础”部分在已有综述（如 Amodio, 2021, 2025; Son et al., 2021, 2023）中已有类似论述或基础，本文的实质性推进与差异化体现不够充分。建议更清晰地阐明本文相较于已有综述的独特理论视角、整合方式或解释深化之处。

**回应**：修改稿在自检报告第 5 题和引言中更明确地阐述了本文的差异化贡献：
（1）相较于 Amodio (2021, 2025)主要从强化学习单一视角出发，本文将强化学习与贝叶斯理论进行了有机整合，而非并列介绍。
（2）相较于 Gershman & Cikara (2023)聚焦于结构学习的单一机制，本文构建了从群体发现→联结建立→更新固化→泛化应用的完整闭环框架。
（3）本文引入了联想表征与命题表征的区分框架，这是上述已有综述中未涉及的理论视角，明确了两大计算框架各自的认知基础与适用边界。

**意见 9**：创新性
b) “研究展望”部分提出的方向（如关系线索、LLM 模拟）确有启发性，但论述较为简略，缺乏具体的研究设计构想或理论推导，可进一步拓展其可行性与理论价值。
**回应**：回复：修改稿对三个展望方向均进行了充实：关系线索部分补充了具体的实验设计构想（操纵虚拟社交网络位置与连接模式）及最新神经证据 (Babür et al., 2024)；社会认知地图部分补充了特质空间的神经编码证据 (Gao et al., 2026; Liu et al., 2025)及具体的方法建议；LLM 部分补充了与贝叶斯框架对接的具体路径（利用 FMAT 方法检验 p(群体 | 特质) 的更新规律）以及与神经科学结合的方向。

**意见 10**：写作细节
a) 概念与表述的准确性：部分术语的使用和概念转换不够严谨。例如，第 3.1 节将原文描述的“泛化梯度”概括为“信任迁移效应”，二者在概念内涵上存在差异；同一句中“外貌相似”与“知觉相似度”的表述，在具体语境（面部知觉）下指代相同，但用词不统一易产生歧义。建议全文对关键概念的引用和转述进行复核，确保表述精确。
b) 语句逻辑与严谨性：部分语句存在逻辑跳跃或绝对化表述。例如，第 4.2 节的“而反刻板印象信息，即高 PE，可能促进刻板印象更新的核心动力” 存在语病；5.1 节的“同男同性恋”疑似多打了“同”字。建议进行全面的语言润色，增强论述的逻辑严密性与学术规范性。
c) 格式与参考文献：存在个别格式不一致问题（如标题字体兼容性），建议统一使用常见学术字体。部分参考文献信息不完整（缺少 DOI），请按期刊要求逐一核对并补充完整。
d) 前后文呼应：摘要强调“动态过程”，但正文中对时间维度上的动态更新机制（如信念更新的时间进程）讨论相对不足，可适当加强。
**回应**：
a) 已对全文关键概念进行复核，统一了相关术语的使用。
b) “而反刻板印象信息，即高 PE，可能促进刻板印象更新的核心动力”已修正为“反刻板印象信息产生的高预测误差可能是促进更新的核心动力”（见 2.3.1 节）。“同男同性恋”已修正为“男同性恋”。已反复核对并修正了原稿中的语病，并对全文进行了语言润色。
c) 已统一全文格式，核对并补充了参考文献 DOI 信息。
d) 修改稿在 2.3 节“联结的更新与固化”中加强了动态过程的论述：2.3.1 节详细讨论了预测误差如何驱动更新，2.3.2 节论述了先验偏差如何抵抗更新，2.3.3 节阐述了探索-利用循环如何使信念在反复互动中自我强化，这些内容共同构成了对动态更新机制的论述。本文所表达的“动态”是这种互动循环的过程，并不强调时间维度。

---

**审稿人 2 意见：**

这篇综述从计算视角（Amodio 的强化学习视角与 Gershman 的贝叶斯结构学习视角）梳理了刻板印象的产生过程（习得、表征、泛化），提出刻板印象的习得与应用依赖于社会

学习与泛化，并指出可以利用大语言模型来模拟刻板印象在社会传播中的语义压缩与固化。文章视角新颖，预期能够促进对刻板印象产生与泛化过程中的认知基础的理解。不过，仍有几点建议请作者参考和修改。

**意见 1**：关于刻板印象的语言传播、习得与表征，以及使用大语言模型研究刻板印象的路径，作者需要区分社会认知的联想表征（associative representations）和命题表征（propositional representations）两种视角，并将其融入本文对两种学习过程的讨论，这对于准确理解刻板印象的习得、表征、泛化过程是很重要的。例如，作者文中举的例子“女孩很善良”是一个命题（proposition），而不是群体与属性词之间的简单联想与激活扩散。理论方面，请作者补充参考 De Houwer、Gawronski 等学者关于联想表征和命题表征的文献，特别是需要论述刻板印象的命题表征及其学习与泛化。方法方面，请作者补充参考 Bao 等学者关于 FMAT 这一社会计算方法的研究，特别是其中关于使用 BERT 语言模型计算刻板印象内容的命题表征，涉及到条件概率 p(群体 | 特质) 的直接计算，与本文提到的贝叶斯框架极为相关。

Bao, H. W. S. (2024). The Fill-Mask Association Test (FMAT): Measuring propositions in natural language. *Journal of Personality and Social Psychology, 127*(3), 537–561. https://doi.org/10.1037/pspa0000396

De Houwer, J., Van Dessel, P., & Moran, T. (2021). Attitudes as propositional representations. *Trends in Cognitive Sciences, 25*(10), 870–882. https://doi.org/10.1016/j.tics.2021.07.003

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin, 132*(5), 692–731. https://doi.org/10.1037/0033-2909.132.5.692

Gawronski, B., & Bodenhausen, G. V. (2011). The associative–propositional evaluation model: Theory, evidence, and open questions. *Advances in Experimental Social Psychology, 44*, 59–127. https://doi.org/10.1016/B978-0-12-385522-0.00002-0

**回应**：感谢审稿专家的重要建议。这一区分对本文的理论框架确实至关重要。我们进行了以下修改：

(a) 在第 2 节开头将刻板印象的信念结构分解为两个子问题后，于 2.2 节自然引入联想表征与命题表征的区分，将其作为两条通路的认知基础加以阐述：强化学习侧重联想表征，贝叶斯理论更贴近命题表征；

(b) 将 2.2 节重构为“经验通路与联想表征”（2.2.1）和“语言通路与命题表征”（2.2.2）两个小节；

(c) 新增 2.2.3 节“联想表征与命题表征的交互”，引入 APE 模型和 Kurdi 等人(2023)的元分析证据；

(d) 补充了审稿人推荐的全部文献：De Houwer 等人(2021)关于态度的命题表征、Gawronski 和 Bodenhausen(2006, 2011)的 APE 模型、Bao (2024)的 FMAT 方法——特别是将 FMAT 与贝叶斯框架中的似然估计 p(群体 | 特质)进行了对接。

**意见 2**：涉及刻板印象习得与泛化的认知机制的新近实证研究，建议作者在概括每篇文献之余，花更多笔墨阐述研究证据“如何”能够启发我们对刻板印象认知机制的理解，即帮助读

者从计算认知的视角理解研究发现的意义和启发，而不只是概括研究发现是什么。这一点体现在稿件中，相当于“6 总结与展望”之前的所有章节也需要有更多“蓝色字”（“作者自己独特的思想观点及分析”）。

**回应：**我们在多处增加了从计算认知视角对研究发现进行解读的分析性论述（文中蓝色标注部分）。例如：在 2.1 节末尾从计算角度阐释了亚型化/子群化对干预的启示；在 2.2.1 节总结了联想表征的核心特征及强化学习框架的算法语言贡献；在 2.3.2 节分析了慢变特征偏好对刻板印象顽固性的解释力；在 2.3.3 节讨论了探索-利用困境作为两大框架交汇点的理论意义。

**意见 3：**综述第 2-5 章，虽然有整体的逻辑顺序，但是每个大章节和每个小部分相互之间都比较孤立，没有形成完整系统的理解框架。建议作者提炼出核心观点，特别是不同内容之间的本质联系，并用表或图呈现阐述，以增加本文的实质创新贡献（目前的图 1 只是一种概念罗列和表层关系，读者需要更深层、更能带来启发的理解）。

**回应：**我们以"某群体具有某特质"的操作性定义为核心，将两个子问题与两大计算框架、两种表征类型对应，形成贯穿全文的分析主线。在 2.3.3 节明确讨论了两大框架的交汇点。在第 4 节总结中整合了全文的计算机制与神经证据。图 1 展示了从信息输入到认知表征再到应用的动态循环。

**意见 4：**“4.1 刻板印象对社会学习的影响”一节比较抽象和难懂，而且与“4.2 预测误差对刻板印象更新的作用”的联系比较模糊，让人困惑。

**回应：**在重构后的版本中，原 4.1 和 4.2 节的内容已被重新组织为 2.3 节“联结的更新与固化”，按“预测误差驱动的更新→先验偏差对更新的抵抗→探索-利用困境”的递进逻辑展开，各小节之间的联系更加明确。

**意见 5：**个别用词请斟酌：文章第一段“近乎本能的快速判断”中的“本能”不妥，暗示了刻板印象是天生的，建议改为“自动化的快速判断”。

**回应：**感谢审稿专家指出。已修改为“自动化的快速判断”，避免了暗示刻板印象是天生的这一不当含义。

---

## 第二轮

**审稿人 1 意见：**

作者已经根据我上一轮的意见进行了逐一修改。在内容上我没有其他建议，但是本文有几处行文存在 AI 的感觉，建议作者润色改进。

**回应：**感谢审稿专家的细心提醒。我们已进行了全面的语言润色，重点优化了句式表达与逻辑衔接，并提升了整体学术表达的严谨性。同时，我们还邀请了多位学术同行对修改后的稿件进行审阅，以进一步提升整体表达质量。文中红色字体标注了重点改写的语段，其余细微的文字润色与格式调整未作逐一标记，以保持排版整洁。

**审稿人 2 意见：**

修改稿有很大程度的提升，作者将联想表征与命题表征视角整合进了强化学习和贝叶斯理论，观点具有启发性。以下几点小意见请考虑修改。

**意见 1**：关于“强化学习侧重联想表征，贝叶斯理论更贴近命题表征”这一区分和对应，我认为有一定的合理性和理论启发性，但命题表征本身不一定等同于贝叶斯推理，命题本身是可判断真伪的陈述句，而贝叶斯推理在其中的作用是挖掘条件概率信息，包括 P(特质 | 群体)和 P(群体 | 特质)。需要进一步解释命题表征和贝叶斯推理的联系和区别，两者不宜完全等同。

**回应**：感谢审稿专家的指正，这一区分确实重要。上一稿中“强化学习侧重联想表征，贝叶斯理论更贴近命题表征”的表述容易让读者将命题表征与贝叶斯推理等同起来，但两者更准确地说是内容与加工过程的关系。命题表征界定的是刻板印象的认知内容形式，可被判断真假的陈述性信念（如“该群体具有某特质”）；贝叶斯推理则是对这类命题内容进行概率性加工的一种计算方式，其作用在于挖掘命题中蕴含的条件概率信息，包括 P(特质 | 群体)和 P(群体 | 特质)。命题表征本身并不预设必须通过贝叶斯方式描述，而贝叶斯推理也不仅限于命题表征。

本文将两者相联系，是因为贝叶斯框架可能恰好为命题层面的刻板印象提供了比较好的数学描述，而非主张两者在概念上等同。

修改稿进行了以下调整：

（a）在 2.2.2 节新增语段论述命题表征与贝叶斯推理的联系与区别：“需要强调的是，虽然贝叶斯推断为命题表征提供了可能的计算模型，但两者在理论本质上仍有显著差异： 前者界定的是认知内容的表征形式，后者则是一种描述信息整合的计算机制。”

（b）将“语言通路建立的命题表征在计算层面对应贝叶斯框架中的条件概率”改为“语言通路所形成的命题表征在计算层面可用贝叶斯框架中的条件概率进行刻画”。

（c）摘要中相应修改为“语言通路则通过贝叶斯推理传递概率性的命题表征”。

（d）总结的第一段删除了将“贝叶斯推断”与“命题表征”直接对应的表述。

**意见 2**：摘要中“语言通路则通过贝叶斯概率推断承载命题表征的传递”，如果要与前半句呼应，可以改为“语言通路则通过贝叶斯推理传递概率性的命题表征”（或类似表述）。

**回应**：感谢审稿专家的建议，已将摘要中“语言通路则通过贝叶斯概率推断承载命题表征的传递”修改为“语言通路则通过贝叶斯推理传递概率性的命题表征”，区分了被传递的内容（命题表征）与计算形式（贝叶斯推理）。

**意见 3**：“V(群体) ← V(群体) + α δ”中的α δ代表什么？需要说明。

**回应**：已在 2.2.1 节公式 V(群体) ← V(群体) + αδ 后补充说明：“其中 α 为学习率(learning rate)，控制新信息对已有价值估计的更新幅度；δ 为预测误差(prediction error)，即实际获得的结果与预期之间的差值”。

**意见 4**：2.1 和 2.2 部分多处提到了刻板印象的干预问题，但是干预似乎不属于 2.1 群体的发现与构建、2.2 群体-特质联结的建立，建议调整位置和表述，干预问题可能更符合 2.3 联结的更新与固化。

**回应**：经过考虑，修改稿将 2.1 末尾关于亚型化和子群化的段落移至 2.3.1“预测误差驱动的更新”的末尾。理由是：亚型化和子群化本质上是大脑面对预测误差时的一种结构性应对——不更新联结本身，而是调整群体结构来应对预测误差，因此放在讨论预测误差的小节中逻辑更为通顺。

**意见 5**：Fill-Mask Association Test 的中文翻译是“掩码填空联系测验”，因为 Association 代表 target 和 attribute 的联系（关联），而这个联系除了联想视角的联结，也可以是命题视角

的条件概率，FMAT 侧重于后者，所以不是“联想”。

**回应：**已将“填充掩码联想测验”修正为“掩码填空联系测验”，以准确体现在命题视角下测量概率联系的特性。

**意见 6：**“潜在隐变量(latent variable)”一般翻译为“潜变量”，但不确定在该语境中是否指的是我们常说的潜变量，请斟酌。
**回应：**已将“潜在隐变量”修正为“潜变量”(latent variable)。在本文语境中，该词指功能泛化过程中用于间接匹配的不可观察变量，与统计学中的潜变量概念一致。

**意见 7：**图 1 中，图注写的是命题表征 P(特质 | 群体)，图中是 P(群体 | 特质)，请检查。
**回应：**感谢指出。已将图 1 中命题表征的标注统一修正为 P(特质 | 群体)，与图注及正文保持一致。图中原有的 P(群体 | 特质)标注为误标。

**意见 8：**最后可以加一小段结语。
**回应：**已在第 5 节末尾添加一小段结语，总结计算认知框架对理解刻板印象的意义。

---

## 第三轮

**审稿人 2 意见：**
**意见 1：**作者已修改完善。不错的文章。
有一处“需要强调的是，虽然贝叶斯推断为命题表征提供了可能的计算模型，但两者在理论本质上仍有显著差异： 前者界定的是认知内容的表征形式，后者则是一种描述信息整合的计算机制。”这里的“前者”“后者”略有指向不明，因为前半句的“前者”是贝叶斯，建议把前者、后者直接改成对应的概念。
**回应：**感谢审稿专家的肯定。关于“前者”“后者”指代不明的问题，已将原文修改为“命题表征界定的是认知内容的表征形式，贝叶斯推断则是一种描述信息整合的计算机制”，直接使用对应概念，避免歧义。

---

**编委 1 意见：**建议引用一下 Wang, H., Dong, Y., Liu, N. & Zhu, L. 社会决策信息的抽象与泛化. 心理科学 (2025).
可以发表。

**编委 2 意见：**同意评审老师的意见，建议接收。

**主编意见：**稿件经过多位专家的审阅，作者进行了认真的修改，达到了发表水平，同意发表。