《心理科学进展》审稿意见与作者回应

题目：心理理论的资源适应性推理：有限理性视角下的双系统与递归层级整合
作者：钟永丹，肖风

第一轮

编委意见：

这篇文章尝试提出一个心理理论 TOM 资源有限推理，将心理理论 TOM 的双系统理论和递归心理理论进行结合，来解决当今业内将两者割裂发展的局面。并且作者提出了一系列可以进行证实或证伪的方法，包括测量个体的决策后果、反应时和生理指标等（重点介绍了 IBS 脑间同步）。逻辑严谨，结构完整。但还存在一些问题。

意见 1：双系统理论和递归心理理论不是平行的关系。ToM 双系统是一个有关心理理论 TOM 的经典基础理论；而递归心理理论是目前对于心理理论 TOM 进行研究所总结的一个经典现象、核心特征或表征层级模型，不是一个基础理论。一个是理论，一个是现象。但作者在文章撰写的时候却完全将两者当作平行的进行论述，这一点有待商榷。

回应：感谢编委提出了一个关乎论文核心定位的深刻问题。编委指出在大多数认知心理学与博弈论文献中，双系统被建构为解释性理论，而递归 ToM 更多以“表征层级现象”出现。我们进一步梳理了递归 ToM 相关研究获得以下认识：

在发展心理学中，Perner(1991)在《Understanding the Representational Mind》中提出了系统的表征心理理论(RTM)，以“元表征”为核心解释构念，建立了三层次发展模型，并推导出从一阶(~4 岁)到二阶(~6-7 岁)错误信念推理的阶段性发展预测。在进化心理学中，Corballis(2011)在《The Recursive Mind》中论证递归思维是人类区别于动物界的关键认知架构，ToM 是其三大支柱之一；Dunbar 学派则提出“高阶意向性”(higher-order intentionality)框架，Kinderman 等人(1998)通过 Imposing Memory Task 确立了成人递归上限约 5 阶的可证伪预测，Stiller 与 Dunbar(2007)进一步将其与社会网络规模关联，Oesch 与 Dunbar(2017)揭示了递归心智化对递归句法的认知脚手架作用。上述研究构成了独立于双系统理论的另一理论传统，其关注的是表征深度的认知架构与演化约束，而非加工方式的机制切换。近年来 Apperly(2025)在 Behavioral and Brain Sciences 上对递归心智化即社会能力核心“递归心智化即社会能力核心”的系统性批判，亦从反面印证了这一传统的理论地位，纯粹的现象不需要被理论性地证伪。

因此，本文的定位并非将“基础理论（双系统）”与“纯现象（递归 ToM）”生硬拼接，而是试图在加工机制理论（双系统）与表征深度理论（递归 ToM）之间建立理论接口。这两个理论源于不同学科脉络、使用不同术语体系、关注 ToM 的不同侧面，长期各自平行发展，这正是现有 ToM 研究缺乏统一解释框架的深层学术史根源。

基于上述认识，我们在修改稿中进行了如下调整：

（1）修正核心表述与学术定位：在摘要、引言及正文第二部分中，摒弃了两条平行路径的表述，将其定位为 ToM 研究中两个互补且各自拥有独立理论的分析维度：前者回答如何投入认知资源进行加工，后者回答推理能推至第几层及其演化边界。

（2）补充递归 ToM 学术史溯源：在正文第 2.2 部分新增简要介绍：递归 ToM 的理论传统简介（Perner 表征理论→Corballis 递归思维假说→Dunbar 高阶意向性框架→Apperly 批判性对话），补充 Perner(1991)、Corballis(2011)、Stiller & Dunbar(2007)、Oesch & Dunbar(2017)、

Apperly(2025)等关键文献。

（3）在正文中明确说明，本文无意将递归 ToM 拔高为与双系统同等形式的解释性理论，两个传统关注的是 ToM 不同层面的问题，本文所做的是在它们之间建立理论接口而非消除学科差异。

修改位置 1：引言第 3 段。

修改稿：近年来 ToM 研究聚焦于两个更贴近动态社会互动的分析维度：一是 ToM 双系统，区分快速、低成本的自动心智化与缓慢、高成本的反思性推理，以解释时间压力和认知负荷如何改变加工方式(Apperly & Butterfill, 2009)；二是递归 ToM，植根于表征 ToM 和递归思维假说，关注个体对他人心理状态的多层嵌套表征及其发展、演化与认知边界(Corballis, 2007; Perner, 1991; Perner & Wimmer, 1985; Oesch & Dunbar, 2017; Stiller & Dunbar, 2007)。近年 Apperly(2025)对递归心智化即社会能力核心的批判，凸显了递归 ToM 作为独立理论的地位。然而，双系统主要回答“如何加工”，递归 ToM 主要回答“推理多深”，缺乏解释加工方式与表征深度如何相互制约的统一框架。此外，为什么更高阶递归推理通常需要反思性加工支持？为什么相同递归深度会因互动情境不同而产生不同收益？这些问题提示，ToM 在互动中的表现并非单纯取决于加工方式或能力上限，而更可能是加工机制与表征深度在有限认知资源约束下协同调节的结果。

修改位置 2：第 2.2 节首段。

修改稿：递归心理理论(Recursive Theory of Mind, Recursive ToM)是区别于 ToM 双系统的另一分析维度，主要关注个体能够将信念、意图、欲望等心理状态嵌入他人心理状态之中，并形成多层嵌套表征的能力(Perner & Wimmer, 1985)。该理论植根于发展心理学与进化心理学两条脉络：前者以 Perner(1994)的表征 ToM 为代表，强调儿童通过元表征能力的发展，逐步实现从一阶错误信念推理（约 4 岁）到二阶错误信念推理（约 6-7 岁）的过渡；后者以 Corballis(2007)的递归思维假说和 Dunbar 学派的高阶意向性框架为代表，强调递归心智化在人类社会认知中的演化意义，并指出成人递归推理通常存在约 5 阶的认知上限(Kinderman et al., 1998; Oesch & Dunbar, 2017; Stiller & Dunbar, 2007)。近期 Apperly(2025)对“递归心智化是社会能力核心”的系统批判，进一步说明了递归 ToM 并非单纯现象描述，而是具有独立理论地位的研究传统。

新增参考文献：

Apperly, I. A. (2025). What makes social abilities sophisticated? Not recursive mentalising. *Behavioral and Brain Sciences*, *48*, e164.

Oesch, N., & Dunbar, R. I. (2017). The emergence of recursion in human language: mentalising predicts recursive syntax task performance. *Journal of Neurolinguistics*, *43*, 95-106.

Perner, J. (1994). Understanding the representational mind. *Pedagogische Studiën*, *71*(3), 231-233.

Stiller, J., & Dunbar, R. I. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, *29*(1), 93-104.

意见 2：有关心理理论 TOM 的经典理论、现象完全不止这两个，作者对其他的经典理论、现象（如，依恋理论、执行功能理论等）没有提及。

回应：感谢编委指出本文理论覆盖面不足的问题。原稿确因过度聚焦于双系统加工与递归层级，未能呈现 ToM 研究的更广阔理论版图，容易给读者造成文献覆盖面不够全面的印象。我们系统梳理了 ToM 领域的经典理论及其与本文框架的关系，获得以下认识：

第一，共同注意是与 ToM 最直接相关的发展前兆理论，应优先纳入。Baron-Cohen(1995)在 ToM 的经典模块模型中，明确将共同注意机制(SAM)定位为意向检测(ID)与心理理论模块(ToMM)之间的关键发展过渡：SAM 不是外在于 ToM 的独立理论，而是 ToM 认知架构的内在组成部分。Sodian 与 Kristen-Antonow(2015)的纵向研究进一步表明，12 个月时婴儿的陈

述性指物，即主动分享对客体的注意和兴趣，独立预测 50 个月时的错误信念理解（控制语言和性别后仍显著），而单纯追随他人视线的反应性共同注意则无此预测力。Tomasello(1995, 2018)的共享意向性框架则从比较认知视角论证了共同注意标志着“将他者理解为意向性主体”的认知革命，是后续递归心理推理的演化起点。对本文而言，共同注意可被理解为“最低递归成本的社会认知协同”：婴儿在共同注意中无需嵌套信念，仅通过共享注意指向即可建立与照料者的认知同步，这与本文框架中“在资源允许范围内实现适应性协同”的核心主张高度一致。

第二，执行功能理论与模拟论、模块论分别从领域一般性认知资源和领域特殊性表征机制的角度构成了 ToM 的双重理论基础。执行功能（抑制控制、工作记忆、认知灵活性）与本文的“认知资源约束”核心假设直接关联，解释了为什么 ToM 推理受到资源边界的限制。模拟论与模块论则分别提供了心智化实现的两种互补机制假说。社会脑理论从神经系统层面揭示了上述理论的共享神经基础。

第三，依恋理论虽对理解 ToM 发展的情感基础有意义，但其与心智化认知机制的直接关联弱于上述理论。在本文聚焦的“动态博弈中认知资源如何约束 ToM 推理”这一特定问题上，共同注意作为 ToM 的发展前兆和低成本社会协同的原型比依恋理论具有更直接的理论关联。

修改位置：引言第 2 段

修改稿：现有 ToM 研究已从多个角度揭示心智化的认知基础：理论论、模拟论和模块论分别强调民间心理学知识、自我模拟和专门化表征机制在理解他人心理状态中的作用(Baron-Cohen, 1997; Gordon, 1986; Gallese & Goldman, 1998; Gopnik & Wellman, 1992; Leslie et al., 2004; Scholl & Leslie, 1999)，执行功能取向与社会脑理论表明，ToM 表现不仅依赖心理状态表征本身，也受到工作记忆、抑制控制以及社会认知网络的支持(André & Maintenant, 2024; Balgova et al., 2024; Carlson & Moses, 2001; Devine & Hughes, 2014; Frith & Frith, 2006; Kliemann et al., 2018)。发展研究同样显示，共同注意是 ToM 的重要早期基础，婴儿无需复杂嵌套推理即可通过共享注意实现初步社会协调(Baron-Cohen, 1997; Chuey & Gweon, 2025; Sodian & Kristen-Antonow, 2015)。因此，ToM 既包含领域特殊的心理状态表征，也受到领域一般性认知资源的调节：双系统、模拟论、模块论修正模型、执行功能理论及神经影像证据均表明，反思性心智化和复杂心理状态推理需要执行控制资源的参与(Apperly, 2010; Balgova et al., 2024; Devine & Hughes, 2014; Goldman, 2009; Qureshi et al., 2010)。

新增参考文献：

Apperly, I. A. (2025). What makes social abilities sophisticated? Not recursive mentalising. *Behavioral and Brain Sciences*, *48*, e164.

André, M. L., & Maintenant, C. (2024). About the relationship between executive function, theory of mind, and language abilities in children with autism: a systematic review. *Applied Neuropsychology: Child*, 1-18.

Chuey, A., & Gweon, H. (2025). Theory of minds: early understanding of interacting minds. *Developmental Psychology*, *7*(1), 91-115.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child development*, *72*(4), 1032-1053.

Goldman, A. I. (2009). Précis of simulating minds: the philosophy, psychology, and neuroscience of mindreading. *Philosophical studies*, *144*(3), 431-434.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind & language*, *1*(2), 158-171.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*(12), 493-501.

Oesch, N., & Dunbar, R. I. (2017). The emergence of recursion in human language: mentalising predicts recursive

syntax task performance. *Journal of Neurolinguistics*, *43*, 95-106.

Perner, J. (1994). Understanding the representational mind. *Pedagogische Studiën*, *71*(3), 231-233.

Sodian, B., & Kristen-Antonow, S. (2015). Declarative joint attention as a foundation of theory of mind. *Developmental psychology*, *51*(9), 1190.

Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and 'theory of mind'. *Mind & language*, *14*(1), 131-153.

Stiller, J., & Dunbar, R. I. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, *29*(1), 93-104.

再次感谢编委专业的审查意见，这些修改使得文章的核心概念（理论与现象的区分）更加准确，同时构建了更加全面扎实的综述视野。

……

外审专家 1 意见：

本文基于有限理性框架提出“ToM 资源适应性推理”整合视角，将加工方式选择与递归深度选择统一为认知资源约束下的协同调节过程，并推导出三条可检验的核心预测。总体来说，本文系统梳理了心理理论领域近年来的研究成果，论证思路清晰，为进一步推进该领域的整合提供了有益参考和可操作的理论框架。论文还存在以下问题：

意见 1：“现有 ToM 研究主要沿两条路径展开。一方面，双系统理论区分快速、低成本的自动心智化与缓慢、高成本的反思性推理，解释时间压力和认知负荷如何改变加工方式。另一方面，递归 ToM 强调个体对他人心理状态的多层嵌套表征，刻画推理深度的个体差异和发展轨迹。”此处表述缺少文献支持，要有文献引用。

回应：接受专家的建议。我们已在修改稿引言部分涉及该内容的地方，补充了该领域的代表性的经典与权威文献，以增强论述的学术根基。

修改位置：引言部分（第 3 段）。

修改稿：近年来 ToM 研究聚焦于两个更贴近动态社会互动的分析维度：一是 ToM 双系统，区分快速、低成本的自动心智化与缓慢、高成本的反思性推理，以解释时间压力和认知负荷如何改变加工方式(Apperly & Butterfill, 2009)；二是递归 ToM，植根于表征 ToM 和递归思维假说，关注个体对他人心理状态的多层嵌套表征及其发展、演化与认知边界(Corballis, 2007; Perner, 1994; Perner & Wimmer, 1985; Oesch & Dunbar, 2017; Stiller & Dunbar, 2007)。

新增参考文献：

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states?. *Psychological review*, *116*(4), 953-970.

Corballis, M. C. (2007). The Uniqueness of human recursive thinking: the ability to think about thinking may be the critical attribute that distinguishes us from all other species. *American Scientist*, *95*(3), 240-248.

Oesch, N., & Dunbar, R. I. (2017). The emergence of recursion in human language: mentalising predicts recursive syntax task performance. *Journal of Neurolinguistics*, *43*, 95-106.

Perner, J. (1994). Understanding the representational mind. *Pedagogische Studiën*, *71*(3), 231-233.

Perner, J., & Wimmer, H. (1985). “John thinks that Mary thinks that…” attribution of second-order beliefs by 5-to 10-year-old children. *Journal of experimental child psychology*, *39*(3), 437-471.

Stiller, J., & Dunbar, R. I. (2007). Perspective-taking and memory capacity predict social network size. *Social Networks*, *29*(1), 93-104.

意见 2：“难以同时兼顾二者的高速度与高准确度”的表述存在语法问题，主语缺失。

回应：感谢专家的细致指正。此处确实存在主语缺失的语病，我们已在修改稿中补全了主语

“人类的认知系统”。

修改位置：第 2.1 节“心理理论双系统”第二段。

修改稿：“人类的认知系统受限于有限的计算能力，难以同时兼顾二者的高速度与高准确度”

意见 3：第一次出现“ASD 研究”未写出缩写的全称。“ASD 研究为此分离性证据”的表述也有语法问题，语义难以理解。

回应：感谢专家的建议。我们已补全 ASD 的中英文全称，并重构了该句的语法逻辑，使语义表达更加清晰。

修改位置：第 2.1 节“心理理论双系统：加工方式维度”第二段。

修改后表述：“孤独症谱系障碍(Autism Spectrum Disorder, ASD)的相关研究为这种认知资源的分离提供了临床证据：部分研究发现 ASD 患者的外显心智化（对应系统 2）受损程度显著重于其内隐心智化（对应系统 1），提示二者在执行时依赖不同的认知资源路径（Baron-Cohen et al., 2001; Senju et al., 2009）。”

意见 4：“外显元认知与 ToM 高度相关，表明系统 2 受限于共享的元表征资源(Qureshi et al., 2020)”该研究是否区分了外显元认知与 ToM 的某些任务表现相关，而这些任务表现又是被认为反应系统 2 的，否则不能作为双重分离的证据。

回应：感谢审稿专家指出该表述中的证据链问题。我们重新核对相关文献后发现，原文中“外显元认知与 ToM 高度相关，表明系统 2 受限于共享的元表征资源(Qureshi et al., 2020)”这一表述确有不严谨之处。Qureshi 等人(2020)主要考察的是 ToM 任务中自我—他人视角干扰是否反映共同机制，以及这种干扰与不同抑制控制成分之间的关系；其研究使用 director task 和一级视觉观点采择任务，并未直接区分外显元认知与内隐元认知，也未提供“外显元认知—外显 ToM—系统 2”的双重分离证据。该研究结果更适合用于说明 ToM 任务表现会受到抑制控制等执行功能成分影响，而不宜作为外显元认知与 ToM 共享元表征资源的直接证据。

根据审稿专家建议，我们已删除原文中关于 Qureshi 等人(2020)的相关表述，并相应弱化该段结论。修订后，本文不再将该研究作为外显元认知与 ToM 共享系统 2 资源的直接证据，而是仅在更一般意义上讨论 ToM 任务表现可能受到执行功能、元表征加工和任务负荷等资源因素的影响。

修改位置：第 2.1 节“心理理论双系统：加工方式维度”第二段。

修改稿：ToM 双系统在资源约束下通过权衡速度与准确性实现“满意性”而非最优决策。系统 1 快速生成初步推断，但依赖不完全信息与启发式加工；系统 2 提供精细推理，却消耗执行功能与元表征资源(Carruthers, 2015; Kovács et al., 2014; Schurz et al., 2014; Yousefi et al., 2021; 张慧 等, 2018)。例如，在最后通牒博弈中，系统 1 的快速情绪化拒绝（依赖脑岛、杏仁核等）往往先于系统 2（前扣带皮层等）的缓慢反思调节，特别在时间不足和负荷较高时，人类的认知系统受限于有限的计算能力，难以同时兼顾二者的高速度与高准确度(Rand et al., 2012; 张慧 等, 2018)。孤独症谱系障碍(Autism Spectrum Disorder, ASD)的相关研究为这种认知资源的分离提供了证据：研究发现其外显心智化(对应系统 2)受损程度显著重于内隐心智化(对应系统 1)，提示二者依赖不同的认知资源(Baron-Cohen et al., 2001; Senju et al., 2009)，说明 ToM 推理受资源约束，需要权衡任务需求、时间压力与可用资源以形成满意性判断。

意见 5：“二者随任务复杂度、认知负荷和时间压力动态切换：…… (3)时间压力削弱二级视角计算能力，对一级视角无明显影响(Todd et al., 2019)” 在这部分表述中，除了第一小点外，其他两点的研究证据与观点之间的支撑对应都不够清晰。

回应：感谢专家的批评。我们重新梳理了这三个证据，明确了它们与“动态切换”观点的逻辑映射关系。重点突出了“高认知负荷、时间压力”是如何迫使个体从“系统 2”回落到“系统 1”的。

修改位置：第 2.1 节“心理理论双系统：加工方式维度”第三段。

修改稿：在资源分配上，系统 1 依赖低成本启发式并无意识整合他人视角(Butterfill & Apperly, 2013; Marshall et al., 2018; 李艺, 肖风, 2021)，系统 2 则通过反思加工校正偏差(Jacob, 2019; Nicholson et al., 2019; Van Overwalle & Vandekerckhove, 2013)。二者随任务复杂度、认知负荷和时间压力动态切换：(1)认知控制能力较强的双语者在复杂任务中能更有效平衡内隐与外显加工(Furlanetto et al., 2015; Freundlieb et al., 2018; Michael et al., 2018; Navarro & Rossi, 2022; Oktay-Gür et al., 2018; 王婷 等, 2019)；(2)在认知负荷方面，高工作记忆负荷会消耗系统 2 所需的执行资源，迫使个体回落至系统 1，降低高阶协作表现并减弱反映反思加工的 P3 波振幅(Newman et al., 2024)；(3)在时间压力方面，时间限制阻断了系统 2 的缓慢计算，导致个体在二级视角采择任务中表现显著下降，但并不影响依赖系统 1 直觉加工的一级视角采择(Todd et al., 2019)。神经层面上，内侧前额叶(mPFC)与颞顶联合区(TPJ)构成了支撑两系统的共享核心网络，系统 1 的自动化加工更多依赖该网络与处理基础社会线索的颞上沟(STS)、顶下小叶(IPL)等区域的自发协同；而系统 2 的反思推理因面临更高的计算与表征负荷，必须额外依赖背外侧前额叶(DLPFC)和前扣带回(ACC)等执行控制网络，以支持工作记忆的维持和复杂的抑制控制(Chiavarino et al., 2010; Grosse Wiesmann et al., 2017; Hallajian et al., 2025; Rothmaler & Grosse Wiesmann, 2023; Schüler et al., 2026)。

意见 6：“在人类发展中，ToM 由早期直觉加工逐渐过渡到反思性推理(Wellman, 2011)，体现出适应性优化：例如，在最后通牒博弈中系统 1 快速拒绝不公平分配，而在复杂谈判中系统 2 用于推断他人意图(Apperly, 2012; Pynadath et al., 2013)。”既然观点是 ToM 能力的“发展”和“过渡”，那么冒号后面的实验证据应该是从发展研究中体现，目前引用的文献证据不能支持观点。

回应：感谢专家的建议。感谢专家指出这一逻辑错位。原文冒号后的例子确实是成人成型后的情境适应，不适合作为发展历程的直接证据。我们已将该处的证据替换为经典的儿童发展 ToM 研究证据，以保持前后文逻辑一致。

修改位置：第 2.1 节“心理理论双系统：加工方式维度”第四段。

修改稿：从演化与发展视角看，ToM 双系统具有一定的情境适应性。非人灵长类和鸦科动物研究表明，动物能够根据他者视线、听觉通达、知识状态或旁观经历调整行为，提示低成本社会线索加工与灵活策略调节之间可能存在演化连续性(Hare et al., 2000; Santos et al., 2006; Bugnyar & Heinrich, 2005; Dally et al., 2006; Kano et al., 2019)。在人类发展中，ToM 表现可能经历由早期隐性、自动化的社会认知敏感性向后期显性、反思性心理状态推理的发展(Kloo et al., 2020; Sodian et al., 2020)。例如，儿童早期即可在无需言语报告的任务中表现出对他人信念和目标的敏感性，而在需要显性报告和认知控制的经典错误信念任务中，稳定表现通常要到 4 岁左右才出现(Apperly & Butterfill, 2009; Buttelmann et al., 2009; Wellman et al., 2001)。这表明，早期社会认知可能更多依赖快速、低成本的信念追踪系统，而成熟的错误信念推理则更多依赖反思性加工和执行控制资源(Apperly & Butterfill, 2009; Kloo et al., 2020; Sodian et al., 2020)。

新增参考文献：

Buttelmann, D., Carpenter, M., & Tomasello, M. (2009). Eighteen-month-old infants show false belief understanding in an active helping paradigm. *Cognition*, *112*(2), 337-342.

Kloo, D., Kristen-Antonow, S., & Sodian, B. (2020). Progressing from an implicit to an explicit false belief understanding: A matter of executive control? *International Journal of Behavioral Development*, *44*(2),

107-115.

Sodian, B., Kristen-Antonow, S., & Kloo, D. (2020). How does children's theory of mind become explicit? A review of longitudinal findings. *Child Development Perspectives*, *14*(3), 171-177.

意见 7：“相反，低阶推理则更多依赖启发式和自动化加工，以较低的认知成本快速判断他人意图”，心理理论不止对意图的推理，这一表述不准确。

回应：感谢指正。心理理论确实包含信念、愿望、意图、情绪等。我们已将“意图”修改为更具概括性的“心理状态”。

修改位置：第 6.1 节第 2 段。

修改稿：从该视角看来，高阶递归需要维持多重嵌套的信念，并持续更新对他人的心理预测，这依赖工作记忆、执行控制和元表征等高成本认知资源，通常通过系统 2 的反思性加工来实现。相反，低阶推理则更多依赖启发式和自动化加工，以较低的认知成本快速推断他人的心理状态（包括意图、信念与愿望等），属于系统 1 的典型特征(Apperly & Butterfill, 2009)。但双系统和递归 ToM 并不存在一一对应关系，而是一种基于资源约束的概率性匹配：虽然低阶 ToM 通常以低成本的系统 1 运行，但在遇到新异环境或预期违背时，也可能需要系统 2 的抑制控制与审慎计算介入(Qureshi et al., 2010; Yoshida et al., 2008)；同样，高阶递归虽依赖反思性加工，但在熟练、重复互动的博弈情境中，个体可能通过强化学习不断优化策略表征，使原本高阶的推理经验固化，逐渐脱离系统 2，表现出自动化的系统 1 特征(Yoshida et al., 2008)。总之，系统 1 和系统 2 代表的是认知投入的动态调节，而非固定对应某个递归层级。

意见 8：“IBS 反映的是双方是否建立了相似的资源调节节奏，而不是直接衡量 ToM 水平的高低”。既然作者在文中也指出了这一点，那么 IBS 是否适宜作为揭示心理理论认知机制的理想指标？

回应：感谢专家提出了一个关乎 IBS 在 ToM 研究中方法论定位的根本性问题。专家指出 IBS 反映的是“资源调节节奏”的相似性而非“ToM 水平”本身，进而追问 IBS 是否适宜作为揭示 ToM 认知机制的理想指标。这一追问促使我们重新审视了 IBS 在本文框架中的认识论地位和论证功能。经过深入思考，我们的认识如下：

第一，IBS 在严格意义上不是、也不应被定位为 ToM 认知机制的“理想指标”。IBS 的本质是神经活动的时间序列耦合，其在多种非 ToM 特异性条件（共同注意、共享任务结构、基础动作同步、情绪唤醒）下均可出现(Redcay & Schilbach, 2019; Hamilton, 2021; Azhari et al., 2025)。将 IBS 直接等同于 ToM 推理层级的深层同步确实构成了过度推断。

第二，不是理想指标不等于完全没有信息价值，关键在于在何种条件下、以何种方式使用 IBS。IBS 的核心信息价值不在于直接测量 ToM 水平，而在于：当通过严格的计算建模提取了个体每轮真实的潜状态表征和价值计算轨迹，并将这些计算参数作为回归变量与神经信号对齐后，残留的 IBS 变异中可能包含与推理层级匹配度相关的信号成分(Czeszumski et al., 2020; Kingsbury et al., 2019)。换言之，IBS 在本文框架中的正确地位是：条件性的、模型依赖的辅助观测窗口，只在计算建模剔除混淆变量的前提下才具有辅助性的指示价值，且需要与其他行为指标（如策略收敛速度、反应时耦合）联合使用。

第三，基于以上认识，我们在修改稿中进行了以下调整（详细修改方案见外审专家 2 意见 1）：

（1）全面下调 IBS 权重：在摘要、引言中将 IBS 定位为“可能表现为脑间同步等备选辅助观测指标的增强”，将 IBS 由第 4 节合并至 3.3 节，明确其伴随性而非构成性地位。

（2）补充 IBS 局限性系统讨论：在 3.3 节第 2 段末尾，客观呈现 IBS 的多种混淆来源及其相关性而非因果性相关性而非因果性的认识论局限。

（3）将命题三从 IBS 中心预测重构为多层次互动协同观测命题（详见下文对命题三的

重建说明）：命题三的核心主张是互动双方的资源调节策略在互动过程中趋向协同匹配，其实证验证依赖行为指标（策略收敛、反应时耦合）、计算指标（HMM 潜状态对齐）和神经指标（条件性 IBS）三个层次的汇聚证据，而非单独依赖于 IBS。

（4）修改命题三的神经预测：将原 mPFC-TPJ vs pSTS-rIPL 的静态脑区-层级对应预测，改为与 2.1 部分一致的核心心智化网络与执行控制网络的跨网络动态耦合预测，即预测推理层级匹配度越高，互动中核心网络与执行网络的跨脑功能连接模式越相似（而非特定脑区的 IBS 绝对值增强）。

修改位置 1：3.3 节第 2 段末尾。

修改稿：IBS 水平也会随互动状态发生动态变化。当互动双方策略趋于一致、信息共享程度提高或共同注意增强时，同步程度可能随之提高(Khalvati et al., 2019; Kayhan et al., 2022)。相反，如果双方在推理投入、策略复杂度或信息更新节奏上明显不匹配，互动耦合可能下降。部分特殊群体研究也提示 IBS 与社会互动协同存在关联：ASD 或抑郁个体在互动任务中的 IBS 较低，并伴随策略协调能力下降(Li & Zhu, 2014; 何振宏, 张丹丹, 2018)。需要强调的是，动态互动中的 IBS 主要体现为互动双方神经活动的相关性，可能受到共同注意、联合行动、共享外部刺激和相似任务结构等影响(Czeszumski et al., 2020; Redcay & Schilbach, 2019; Azhari et al., 2025)。因此，本文将 IBS 定位为互动协同的条件性、辅助性观测指标。换言之，IBS 的理论价值不在于单独证明 ToM 机制，而在于为观察互动双方是否形成相似的资源调节节奏提供一个可能的神经过程指标。

修改位置 2：5.2 节核心预测研究的命题三。

修改稿：命题三(互动协同观测命题)：互动双方的资源调节策略会在连续反馈中趋向协同匹配，即互动双方在加工方式切换和递归深度选择逐渐相互协调。这一命题可从三个层面检验：（1）行为层面，若双方处于高推理层级匹配状态，其策略选择、策略切换和反应时变化应表现出更强的时序耦合。一方策略切换在控制任务难度、基础运动同步、选择频率和情绪唤醒后，仍能显著预测另一方在若干回合内的策略调整，则可作为行为证据。（2）计算建模层面，可基于选择、反应时和任务负荷等指标，建模推断个体在低成本启发式反应、不确定性驱动探索、高阶递归更新等潜状态间的转换(Ashwood et al., 2022; Pan et al., 2025)。若双方潜状态序列或转移矩阵更相似，说明其采用了相近的资源投入和更新规则，而不只是外显行为相似。（3）神经层面，在控制共同注意、共享任务结构和行为同步等混淆因素后，将 IBS 作为辅助性过程指标，考察在不同潜在推理状态下功能网络耦合模式的条件性相似(Czeszumski et al., 2020; Hamilton, 2021)。若双方均处于高阶递归，DLPFC 与 mPFC、TPJ 等相关跨脑功能连接应强于不匹配时。但该预测仍需通过动态因果建模、滞后连接分析或其他方向性分析方法检验其状态依赖性和方向性(Bilek et al., 2022)。因此，只有当行为耦合、潜状态对齐与条件性跨脑网络相似性共同出现时，才能支持互动双方资源调节策略趋向协同匹配。

新增参考文献：

Ashwood, Z. C., Roy, N. A., Stone, I. R., International Brain Laboratory, Urai, A. E., Churchland, A. K., ... & Pillow, J. W. (2022). Mice alternate between discrete strategies during perceptual decision-making. *Nature Neuroscience*, *25*(2), 201-212.

Bilek, E., Zeidman, P., Kirsch, P., Tost, H., Meyer-Lindenberg, A., & Friston, K. (2022). Directed coupling in multi-brain networks underlies generalized synchrony during social exchange. *NeuroImage*, *252*, 119038.

Hamilton, A. F. D. C. (2021). Hyperscanning: beyond the hype. *Neuron*, *109*(3), 404-407.

Pan, T. F., Li, J. J., Thompson, B., & GE Collins, A. (2025). Latent variable sequence identification for cognitive models with neural network estimators. *Behavior Research Methods*, *57*(10), 272.

意见 9：心理理论涉及到的脑区相当广泛，包括文章图内呈现的 mPFC、rTPJ、楔前叶、STS、

DLPFC 等，但并没有非常一致的证据表明 STS 对应低阶 ToM 推理，mPFC 和 rTPJ 等对应高阶 ToM 推理。图中对脑区和心理理论不同层级任务的归属与对应是否有明确的文献支持？

回应：非常感谢您严谨且深刻的意见。在原稿件中，我们试图通过图 1 和行文中的脑区划分来直观呈现 ToM 双系统与递归层级的区别。然而，如您所指出的，将 STS 划归为低阶 ToM，将 mPFC 和 rTPJ 划归为高阶 ToM，是对当前认知神经科学证据的过度简化，确实缺乏足够一致的文献支持。

现有的神经成像研究和元分析（如 Schurz et al., 2014; Buergi et al., 2026）表明，内侧前额叶(mPFC)和颞顶联合区(TPJ)构成了 ToM 的“核心心智化网络”，在低阶（内隐/自动化）和高阶（外显/递归）心理状态推理中均发挥着不可或缺的基础表征作用，而非仅对应高阶推理。这两者在神经层面的真正分化机制在于：

在依赖快速启发式的低阶社会互动中，个体主要依赖核心心智化网络（mPFC, TPJ）与处理基础动作意图、视线等社会线索的 STS、IPL 等区域进行底层的自发协同。随着递归深度（如 2 阶及以上）和博弈复杂度的增加，高阶推理因为面临极高的工作记忆负荷和元表征需求，必须在核心网络的基础上额外招募背外侧前额叶（DLPFC）和前扣带回（ACC）等执行控制网络（Zhen & Yu, 2021; Coricelli & Nagel, 2009）。

基于您的建议，我们对稿件进行了修正，摒弃了“特定脑区唯一对应特定层级”的静态视角，采用“核心网络共享与执行网络动态”的调节模型。这修正了神经生物学上的不严谨，也更加契合本文“资源适应性推理”的核心主旨。

修改位置 1：正文“2.1 心理理论双系统：加工方式维度”（关于神经基础层面的描述修正）

修改稿：“神经层面上，内侧前额叶(mPFC)与颞顶联合区(TPJ)构成了支撑两系统的共享核心网络。系统 1 的自动化加工更多依赖该核心网络与处理基础社会线索的颞上沟(STS)、顶下小叶(IPL)等区域的自发协同；而系统 2 的反思推理因面临更高的计算与表征负荷，必须额外依赖背外侧前额叶(DLPFC)和前扣带回(ACC)等执行控制网络，以支持工作记忆的维持和复杂的抑制控制(Chiavarino et al., 2010; Grosse Wiesmann et al., 2017; Hallajian et al., 2025; Rothmaler & Grosse Wiesmann, 2023; Schüler et al., 2026)。”

修改位置 2：正文“5.1 协同机制的解释框架”（增加关于神经机制动态调配的论述）

修改稿：“在神经层面，支持复杂心理推理的资源协同主要体现在核心心智化网络（如 mPFC, TPJ）与领域一般性控制网络（如 DLPFC）的动态耦合上。当任务处于低阶启发式阶段时，基础感知区域与核心网络足以应对；而在高负荷、高递归需求条件下，DLPFC 的额外介入及其与核心网络的协同活动显著增强，以支持多重嵌套意图的维持。当资源受限导致认知过载时，这种前额叶跨网络协同会显著减弱(Meyer & Collier, 2020; Spatola et al., 2022; Yang et al., 2022; Zhang et al., 2024; Zheltyakova et al., 2026)。”

修改位置 3：图 1（ToM 资源适应性推理与脑间同步）重新绘制。

修改方案：已将图 1 的脑间同步图进行了重绘：在原来基础上，将 mPFC 由原来的系统 2（高阶 ToM）调整到协同调节区（共享核心心智化网络），并在图中标记为绿色。ACC 则由原来的协同调节区调整到系统 2（高阶 ToM），并标记为红色。系统 1（低阶 ToM）相关脑区保持不变，标记为蓝色。最后，图中涉及到的所有脑区均可以在文中找到对应参考文献。

意见 10：“命题一(切换阈值命题)：随着任务复杂度、对手不确定性与递归需求提高，个体更可能增加认知资源投入，并表现出更多反思性加工；而在高时间压力、高认知负荷或资源受限条件下，则更倾向于维持低成本推理。”在这一表述中，存在着一种矛盾，任务复杂、对手不确定、递归层级高，是不是也意味着更大的认知负荷及资源不足，那到底是倾向于推高资源投入还是维持低成本呢？

回应：感谢专家指出了这一表述中潜在的逻辑问题。这里确实存在概念上的模糊，我们需要在文中明确区分任务需求与资源限制。我们将原表述重构为：认知资源在任务需求与资源限制的共同作用下，动态调节加工方式与递归表征之间的关系：任务复杂度、不确定性和潜在收益等需求增加时，个体倾向于投入更多认知资源，提高系统 2 的参与概率；而当时间压力、认知负荷或工作记忆限制使计算成本超过可用资源时，个体则倾向于回落至低成本的系统 1。

文中修改位置：第 6.2 节“命题一（切换阈值命题）”。

修改稿：命题一(切换阈值命题)：认知资源在任务需求与内部资源限制的共同作用下，动态调节加工方式与递归表征的关系。任务复杂度、不确定性和潜在收益等任务需求增加时，个体倾向于投入更多认知资源，提高系统 2 的参与概率；而当时间压力、认知负荷或工作记忆限制使计算成本超过内部可用资源时，个体则倾向于回落至低成本的系统 1(Apperly, 2010; Apperly & Butterfill, 2009)。高阶递归并不等同于系统 2，但随着递归层级提升，对工作记忆维持、信念更新和冲突监控的要求增加，系统 2 的参与概率将系统性提高。本文进一步预测，认知资源是通过调节系统 1 与系统 2 的使用概率，影响个体的递归表征深度。可通过双任务范式或时间压力操纵，结合反应时、瞳孔扩张及多阶信念模拟等验证(Hedden & Zhang, 2002; Krockow et al., 2016; Mathôt, 2018; Von Der Osten et al., 2017)。验证核心在于区分能力受损与策略选择：前者表现为各层级准确率的普遍下降，后者则表现为高阶递归表征减少，并伴随从系统 2 向系统 1 的加工方式转移。例如，计算建模可捕捉个体在资源耗竭临界点时，从拟合高阶滑落至低阶，或在隐马尔可夫模型中向低投入隐状态的显著转移(Ashwood et al., 2022; Ho et al., 2022)。

新增参考文献：

Ashwood, Z. C., Roy, N. A., Stone, I. R., International Brain Laboratory, Urai, A. E., Churchland, A. K., ... & Pillow, J. W. (2022). Mice alternate between discrete strategies during perceptual decision-making. *Nature Neuroscience*, *25*(2), 201-212.

Ho, M. K., Saxe, R., & Cushman, F. (2022). Planning with theory of mind. *Trends in Cognitive Sciences*, *26*(11), 959-971.

Mathôt, S. (2018). Pupillometry: Psychology, physiology, and function. *Journal of cognition*, *1*(1), 16.

意见 11：总体来看，本文基于有限理性提出 ToM 资源适应性推理整合视角，其核心是认知资源的分配。心理理论普遍被认为是社会认知中的一种领域特殊性能力，而认知资源分配属于领域一般性能力。本文的基本思路是用领域一般性能力解释领域特殊性能力，这一做法在逻辑和方法论上存在一定问题，需要作者进一步思考和澄清。首先，这属于范畴混淆：领域一般性机制更多反映的是能力表现的支撑条件，而领域特殊性关注的是能力本身的表征内容与模块化属性。用通用资源的差异无法解释心理理论这种专门性社会认知能力的存在与本质。其次，从方法论角度看，这种解释面临还原论风险：基于有限理性视角下 ToM 整合的三个命题，只能说明通用资源对心理理论任务表现的影响，也无法揭示心理理论可能具有的领域特殊表征或加工模块。

回应：感谢审稿专家提出的这一关乎本文理论合法性的意见。审稿人指出，若 ToM 是纯粹领域特殊性能力而本文用领域一般性资源框架加以解释，则存在“范畴混淆”和“还原论风险”。为此，我们对 ToM 的领域属性进行了更系统的文献梳理和理论澄清。

审稿人的论证建立在“ToM 是领域特殊性能力”这一前提之上。然而，来自 ToM 领域内多个经典理论传统的证据表明，ToM 并非纯粹的领域特殊性能力，而是领域特殊性与领域一般性的混合体，这一认识正是本文双系统分析的核心理论基础。具体而言：

第一，双系统理论本身就论证了 ToM 的双重属性。本文第 2.1 节引述的大量证据表明，ToM 的系统 1（自动心智化）主要依赖核心心智化网络（mPFC, TPJ）与基础社会线索区（STS, IPL），表现出较强的领域特殊性；而系统 2（反思性推理）必须额外招募背外侧前额叶（DLPFC）

和前扣带回（ACC），标准的领域一般性执行控制网络，以支持工作记忆维持和冲突监控（Chiavarino et al., 2010; Grosse Wiesmann et al., 2017; Schüler et al., 2026）。工作记忆负荷的增加会选择性损害外显 ToM 而保留内隐 ToM（Newman et al., 2024），以及执行功能在发展中对 ToM 的纵向预测效应（Carlson & Moses, 2001），均表明领域一般性资源是 ToM 高阶运作的“必要条件”而非仅仅是“支撑条件”。如果 ToM 是纯粹的领域特殊性模块，那么 DLPFC 在高阶递归推理中的额外激活和工作记忆负荷的选择性损害将难以解释。

第二，模拟论的内在逻辑同样要求领域一般性资源的介入。模拟论（Goldman, 2006; Gallese & Goldman, 1998）主张个体通过调用自身的认知系统“模拟”他人的心理状态。这一机制本身就意味着：模拟过程必然消耗领域一般性加工资源，即工作记忆维持模拟场景，抑制控制压制自我视角干扰。镜像神经元系统的 fMRI 和 TMS 研究反复证实，即使是最“自动化”的镜像反应，在高认知负荷条件下也需要前额叶执行控制的调制（Gallese, 2007; Keysers & Gazzola, 2007）。如果 ToM 是纯粹的领域特殊性能力，那么“用自身认知资源模拟他人”的过程将无法解释，因为模拟本身即是领域一般性加工。这与双系统理论的预测高度一致：系统 1 的自动模拟（内隐镜像）偏向领域特殊，系统 2 的反思性模拟（外显观点采择）显著依赖领域一般执行资源。

第三，即使在模块论框架内，ToM 也不等同于纯粹“领域特殊性”。审稿人的前提可能源自 20 世纪 90 年代模块论传统（Leslie, 1987; Baron-Cohen, 1995），但即使是最坚定的模块论者也未主张 ToM 的全部运作均由领域特殊性模块完成。Scholl 与 Leslie（1999）区分了 ToM 的核心表征机制（可能为领域特殊）与表现机制（必然涉及领域一般加工资源），这一区分与本文“能推断什么 vs 实际推断到多深”的边界设定完全一致。Baron-Cohen（1995）的四模块模型中的共同注意机制（SAM），其本质正是将领域一般性注意资源与领域特殊性社会信息加工相桥接。因此，即使在模块论框架内，本文用资源约束解释 ToM 的动态表现变异也不构成范畴混淆。

第四，执行功能理论本身就是 ToM 领域内部以领域一般性变量为核心解释工具的理论传统。编委在另一条意见中明确将执行功能理论列为 ToM 的经典理论之一，这一定位表明 ToM 领域内部已存在成熟的、以领域一般性认知资源为核心解释变量的理论路径。Carlson 与 Moses（2001）的纵向研究表明，抑制控制与工作记忆在控制年龄、言语能力和初始 ToM 水平后仍独立预测 ToM 表现；这一发现已被反复验证为发展心理学中的稳健发现（Devine & Hughes, 2014）。如果 ToM 是纯粹领域特殊性能力，那么执行功能（而非 ToM 专门模块的成熟）预测 ToM 表现将构成理论异常，但事实恰恰相反。这一理论传统的存在直接消解了“范畴混淆”的指控：用领域一般性变量解释 ToM 表现变异，不是在两个不相干范畴间建立强行关联，而是在 ToM 文献中早已存在的理论路径上继续推进。

第五，神经影像元分析与社会脑理论提供了多层次的汇聚证据。Balgova 等人（2024）对 344 项 fMRI 研究的元分析表明，ToM 核心网络与语义认知网络存在实质性重叠，专门化心智化回路是嵌套在领域一般性系统之中的。Schurz 等人（2014）的大规模元分析进一步表明，不同 ToM 任务依赖不同的子加工过程和脑区组合，预测力最强的理论是那些包含多重认知子加工过程的理论，挑战了统一领域特殊性假说。Dunbar（1998, 2010）的社会脑假说从演化角度论证了新皮层体积（特别是前额叶这一领域一般性执行控制的核心）与社会群体规模及高阶意向性之间的因果关联。Adolphs（2009）在综述中明确指出社会认知不能仅由专门化模块解释，而是依赖于广泛分布的神经网络，其中许多节点也参与非社会认知。Apperly（2010）在《Mindreaders》中更是明确论证：ToM 的认知需求虽确切，但领域一般性执行加工足以解释 ToM 表现模式。

综合以上来自双系统理论、模拟论、模块论、执行功能理论、社会脑理论和神经影像元分析的汇聚证据，审稿人提出的“范畴混淆”和“还原论风险”建立在 ToM 是纯粹领域特

殊性这一与多重理论传统相矛盾的前提之上。由于 ToM 本身是领域特殊性与领域一般性的混合体，因此：（a）用资源约束框架（领域一般性维度）解释 ToM 在高负荷条件下的动态表现变异（如从高阶回落至低阶、从系统 2 切换至系统 1），就不是用“范畴外”的工具解释“范畴内”的对象，而是对该对象已经内蕴的领域一般性维度系统分析；（b）本文明确界定了理论边界：ToM 的专门性表征内容与模块化属性决定了能推断什么，而领域一般性认知资源决定了实际推断到多深（见引言和 5.1 节），这并非将领域特殊性还原为领域一般性，而是阐释两者在认知资源约束下的协同耦合机制。领域一般性执行功能不是 ToM 的干扰变量而是其正常运作的组成部分（Apperly, 2010），因此在资源约束下考察 ToM 的动态调节，不是还原论而是整合论，即将 ToM 研究中长期分隔的领域特殊性表征机制与领域一般性加工资源在统一框架中考察其协同运作。

修改位置 1：引言第 2 段。强化边界澄清的文献基础

修改稿：现有 ToM 研究已从多个角度揭示心智化的认知基础：理论论、模拟论和模块论分别强调民间心理学知识、自我模拟和专门化表征机制在理解他人心理状态中的作用(Baron-Cohen, 1997; Gordon, 1986; Gallese & Goldman, 1998; Gopnik & Wellman, 1992; Leslie et al., 2004; Scholl & Leslie, 1999)，执行功能取向与社会脑理论表明，ToM 表现不仅依赖心理状态表征本身，也受到工作记忆、抑制控制以及社会认知网络的支持(André et al., 2024; Balgova et al., 2024;Carlson & Moses, 2001; Devine & Hughes, 2014; Frith & Frith, 2006; Kliemann et al., 2018)。发展研究同样显示，共同注意是 ToM 的重要早期基础，婴儿无需复杂嵌套推理即可通过共享注意实现初步社会协调(Baron-Cohen, 1997; Chuey & Gweon, 2025; Sodian & Kristen-Antonow, 2015)。因此，ToM 既包含领域特殊的心理状态表征，也受到领域一般性认知资源的调节：双系统、模拟论、模块论修正模型、执行功能理论及神经影像证据均表明，反思性心智化和复杂心理状态推理需要执行控制资源的参与(Apperly, 2010; Balgova et al., 2024; Devine & Hughes, 2014; Goldman, 2009; Qureshi et al., 2010)。

修改位置 2：第 5.1 节第 1 段末尾，在整合视角提出后增加领域特殊性与领域一般性交互的实质论述

修改稿：ToM 双系统理论和递归 ToM 理论分别侧重于加工方式和推理深度，但各自都存在局限：前者很难解释为什么高阶推理特别依赖反思性加工，后者则难以说明推理深度为什么会随情境动态变化。在真实社会互动中，ToM 并非简单的系统切换或层级提升，更可能是个体在认知资源有限的情况下，对推理复杂度和加工投入联合调节的动态过程。因此，本文提出“ToM 资源适应性推理”(Resource-Adaptive ToM Reasoning)整合视角，将 ToM 视为一种资源分配过程：在动态博弈中，个体并不是孤立地决定采用哪种加工方式或达到哪一层递归深度，而是持续评估增加认知投入能否带来足够的策略收益，并据此灵活调整自己的推理方式和深度。该视角将资源约束引入 ToM 是为了阐明表现差异背后的调节机制。系统 1 与系统 2 的切换、低阶与高阶递归的选择，实质上是领域特殊性模块（如支持信念表征的心智化核心网络）在领域一般性资源（如支持工作记忆的前额叶执行控制网络）支持下的动态耦合结果。这种协同调节不涉及 ToM 内在表征模块的改写，但能解释为何在面临特定任务负荷与时间压力时，个体潜在的高阶社会认知能力会被压抑或转化为低成本的直觉表现。

外审专家 2 意见：

本文针对心理理论双系统加工与递归心理理论长期分立、缺少统一解释框架的学术痛点，基于有限理性的认知资源约束，创新性提出 ToM 资源适应性推理整合模型，符合本刊问题导向型综述要求，避免了教科书式平铺罗列。文章文献覆盖面广，梳理了发展、博弈决策、社会认知神经等多领域证据，提炼出三条可证伪的具体研究命题，从行为、计算建模、临床

干预及人机智能多方面给出后续研究路径，理论创新突出，整体学术质量较好。
文章主要存在三处需要修改完善的问题：

意见 1：IBS 相关论证不合理：现有研究表明动态博弈中脑间同步（IBS）只是互动伴随的相关性指标，同步现象可在合作、竞争、共同注意等多种条件下出现，无法直接用来判定双方推理层级匹配，尚不具备因果推论效力。作者未客观呈现该领域研究争议，造成 IBS 在全文定位摇摆，时而作为外在观测指标，时而被视作内在协同机制，第三条预测论证根基不足。建议降低 IBS 在核心框架中的权重，改为备选辅助观测指标，补充 IBS 研究现存局限与剔除混淆变量的实验控制思路。

回应：非常感谢专家对 IBS 论证结构的诊断。专家指出 IBS 在全文定位摇摆，导致“第三条预测论证根基不足”。专家的诊断可分解为三个层次的问题，我们对每个层次进行了针对性修正：

问题一：IBS 定位摇摆（观测指标↔协同机制）。根源在于：原文在构建命题三时，将 IBS 同时赋予了两种不兼容的角色：既用它来观测协同，又用它来构成协同（机制角色），导致在不同段落中读出不同的 IBS 定位。

修正方案：（1）4.3 节标题从“脑间同步作为协同机制的证据”改为“3.3 脑间同步作为互动协同的辅助观测指标”；（2）摘要和引言中将 IBS 表述为“可能表现为脑间同步等备选辅助观测指标的增强”；（3）5.1 节中明确 IBS 可以作为这种协同调节的过程性指标之一，但必须与控制性分析联合使用。

问题二：第三条预测论证根基不足。根源在于：当 IBS 的定位被纠正后，命题三的论证结构出现了问题，如果 IBS 不构成协同机制而只是辅助指标，那么命题三究竟在预测什么？

修正方案：将命题三从以 IBS 为中心的单一指标预测重构为资源协同匹配的多层面可观测预测体系。本次修改进一步在命题三的验证思路中明确指出：IBS 的证据价值是高度条件性的，只有在计算建模剔除混淆变量后才有辅助价值，且不能单独作为命题三成立的决定性证据。

修改位置：6.2 核心预测研究 命题三

修改稿：命题三(互动协同观测命题)：互动双方的资源调节策略会在连续反馈中趋向协同匹配，即互动双方在加工方式切换和递归深度选择逐渐相互协调。这一命题可从三个层面检验：（1）行为层面，若双方处于高推理层级匹配状态，其策略选择、策略切换和反应时变化应表现出更强的时序耦合。一方策略切换在控制任务难度、基础运动同步、选择频率和情绪唤醒后，仍能显著预测另一方在若干回合内的策略调整，则可作为行为证据。（2）计算建模层面，可基于选择、反应时和任务负荷等指标，建模推断个体在低成本启发式反应、不确定性驱动探索、高阶递归更新等潜状态间的转换(Ashwood et al., 2022; Li et al., 2023)。若双方潜状态序列或转移矩阵更相似，说明其采用了相近的资源投入和更新规则，而不只是外显行为相似。（3）神经层面，在控制共同注意、共享任务结构和行为同步等混淆因素后，将 IBS 作为辅助性过程指标，考察在不同潜在推理状态下功能网络耦合模式的条件性相似(Czeszumski et al., 2020; Hamilton, 2021)。若双方均处于高阶递归，DLPFC 与 mPFC、TPJ 等相关跨脑功能连接应强于不匹配时。但该预测仍需通过动态因果建模、滞后连接分析或其他方向性分析方法检验其状态依赖性和方向性(Bilek et al., 2022)。因此，只有当行为耦合、潜状态对齐与条件性跨脑网络相似性共同出现时，才能支持互动双方资源调节策略趋向协同匹配。

意见 2：ToM 测量范式章节逻辑脱节：测量范式演变部分内容偏孤立，没有紧扣“传统范式缺陷催生资源适应性推理模型”这条主线展开，未能阐明过往静态测量难以同步操控加工方

式与递归深度，因此需要本文整合框架予以完善，段落与前后文关联性偏弱。建议重构该小节，围绕范式短板引出本模型的必要性，紧密服务全文核心立论。

回应：感谢专家的建议。原稿在第 3 节的描述确实大多停留在范式演变的罗列，资源约束的分析几乎只在表 1 中呈现。我们在修订稿中对第 3 节和第 4 节进行了逻辑重构，将其核心主线：把“静态范式到动态博弈”“加工方式切换与推理深度有限性”“IBS 作为辅助观测指标”放在同一章，逻辑修改为：为什么传统范式不够→为什么动态博弈是关键场域→为什么需要有限理性整合。

修改位置 1：删除第 4 节所有内容，并修改第三节所有标题为：3 心理理论测量范式改为：3 动态博弈中的心理理论；3.1 ToM 测量范式的演进改为：3.1 从静态范式到动态博弈范式：ToM 测量范式的资源化转向；3.2 测量范式对整合框架的启示改为：3.2 动态博弈中的加工方式切换与推理深度有限性；并新增 3.3 脑间同步作为互动协同的辅助观测指标。

修改位置 2：第 3 节所有内容。

修改稿：3 动态博弈中的心理理论

前文指出，ToM 双系统与递归 ToM 分别回答“如何加工”和“推理多深”的问题，缺乏同时解释加工投入、递归深度和互动结果的统一接口。要弥合这一断裂，仅依赖传统静态 ToM 任务并不充分。传统任务有助于界定个体是否具备基本心理状态理解能力，难以呈现真实互动中持续反馈、策略调整、时间压力、认知负荷和收益权衡共同作用。而动态博弈不仅是 ToM 研究的应用场景，更是检验 ToM 的重要范式：能同时呈现个体如何调用加工系统、选择递归深度，并在互动中形成策略协调或失配(详见**错误!未找到引用源。**)。

3.1 从静态范式到动态博弈范式：ToM 测量范式的资源化转向

传统静态离线测验主要包括错误信念任务、观点采择任务、提示任务、奇异故事任务、失礼识别测试和眼神读心任务等，评估个体对他人信念、意图、情绪和社会语境的理解水平，但任务结构相对固定，互动反馈有限，更多呈现个体在相对稳定条件下的能力上限。其资源约束体现在：第一，基础信念理解虽在自动化条件下资源需求较低，但在需要外显报告时，个体需要抑制自身视角或现实知识的干扰，并调用执行控制资源(Rakoczy, 2022; Carlson & Moses, 2001)。第二，高阶意图与语言推理要求个体维持非字面语义、整合上下文并进行执行控制，对语言能力、加工速度和工作记忆有更高要求(Osterhaus & Bosacki, 2022; Montgomery et al., 2024)。第三，复杂社会情境任务同时追踪多个角色的信念、情绪和社会规范，涉及冲突信息协调与复杂执行控制；二阶错误信念理解与执行能力之间的元分析证据也说明，高阶 ToM 表现并非单纯反映表征能力，而是与一般认知资源存在关联(Peloquin et al., 2023)。因此，静态范式的资源约束通常被任务结构所弱化，难以充分呈现真实互动中资源投入与推理深度的动态关系。

动态博弈将 ToM 置于更接近真实互动的情境中，凸显推理深度是信念、计算能力与认知资源共同作用的结果。在选美博弈、斯塔克伯格博弈和同时行动博弈中，个体根据对手行为实时调整递归推理策略，但受工作记忆和计算能力限制。例如在 level-k 相关策略互动任务中，推理深度通常集中于有限层级，外显 level-k 行为也不一定等同于真实 level-k 思维(Jin, 2021)。在信任博弈、囚徒困境和猎鹿博弈中，个体还需权衡合作意图推断、公平偏好与收益计算，认知负荷或时间压力以进行合作维持、信任表达和协调选择(Mieth et al., 2021; Zhao et al., 2024)。在独裁者博弈和最后通牒博弈中，公平判断涉及直觉情绪反应与理性收益计算的冲突，眼动证据也支持在最后通牒博弈中，公平偏好与自利计算分别对应自动化加工和控制性加工的相对贡献(Wei et al., 2022)。在石头剪刀布博弈和其他策略对抗任务中，个体还需要根据反馈结果、对手策略不确定性和预测误差实时调整推理投入，避免过度递归造成认知资源耗竭。由此，动态博弈不仅测量个体“能推理到多深”，还揭示其在有限资源条件下“是否值得继续推理”以及“应以何种方式推理”。

综上，ToM 测量范式经历了从静态、非互动情境到动态互动博弈情境的演进。早期经典范式主要考察个体是否具备理解他人信念、意图和情绪的能力，强调有无心理状态表征及其层级结构；动态博弈范式则关注个体在时间压力、认知负荷、不确定性和收益权衡中，如何调节加工方式、分配认知资源并选择适当的推理深度(Camerer et al., 2004; Coricelli & Nagel, 2009; Yoshida et al., 2008)。关于 ToM 测量效度的讨论也表明，经典 ToM 任务并不总是纯粹测量心理状态表征本身，其表现还可能受到语言理解、执行功能、语境整合和任务要求等因素影响(Montgomery et al., 2024; Quesque & Rossetti, 2020)。因此，测量范式的演进反映了 ToM 研究问题的转向：从考察个体是否具备 ToM，转向考察个体在资源约束下如何调用 ToM。

3.2 动态博弈中的加工方式切换与推理深度有限性

动态博弈揭示了 ToM 加工方式会随情境约束发生系统性变化：在时间压力大或认知负荷高时，个体更倾向于采用快速、低成本的直觉式加工（系统 1）；而在需要持续预测对方意图、协调合作或应对复杂策略变化时，则更容易转向反思性推理（系统 2）。例如，在最后通牒博弈中，时间有限时个体更可能做出基于情绪的快速拒绝，而有足够时间思考时则更可能权衡收益并调整决策(Rand et al., 2012; 张慧 等, 2018)；在多轮信任博弈或合作任务中，若互动连续且对方策略不断变化，个体通常会逐步投入更多认知资源，更精细地表征和预测他人意图(Thompson et al., 2021; Zhang et al., 2021)。这种加工方式的切换受到环境结构和预测误差的共同调节：在稳定、低不确定性情境中，个体倾向于维持启发式策略；当预测误差增加或对手策略复杂度提高时，反思性加工的边际收益上升，个体更可能启动系统 2 以提高预测精度(Coricelli & Nagel, 2009; Enevoldsen & Waade, 2020)，并根据对手行为反馈持续调整其加工方式与策略选择(Freire et al., 2023; Li et al., 2014; Osawa & Imai, 2013)。因此，动态博弈中的加工方式切换本质上反映了个体对认知资源消耗与策略收益的权衡。

个体在互动中不会无限递归推理，表现出稳定的有限层级结构。在选美博弈、斯塔克伯格博弈和同时行动博弈等任务中，个体需根据他人可能的选择递归推理，但受工作记忆和计算能力限制，推理深度通常集中于有限层级，大多数参与者的推理深度集中于 1-2 阶，仅在高激励或高复杂度条件下才出现更高阶推理(Camerer et al., 2004; Nagel, 1995; Wilson et al., 2023)，这表明个体在社会互动中通常追求可操作范围内的适度推理，而非理论上的最大递归深度。推理深度随任务复杂度、对手不确定性和认知负荷动态变化：当环境更复杂或对手行为更难预测时，个体可能适度提高推理层级以增强对他人行为的预测(De Weerd et al., 2015; Hu et al., 2017; Krill & Platek, 2012; Markiewicz et al., 2024; 李晶, 刘希平, 2009)，但这种提升并非线性递增：随着递归层级不断加深，个体需要维持更多嵌套信念并持续更新预测，认知资源消耗会快速上升，当推理成本超过收益时，过度递归反而会降低决策效率和整体策略表现(Anufriev et al., 2024; Ridinger & McBride, 2025)；在高不确定性或高认知负荷情境下，个体更可能退回较低阶推理，或采用简单启发式策略(Lieder & Griffiths, 2020)。这说明，互动中的有效 ToM 并非越深越好，而是取决于推理深度、加工成本和情境收益的适配程度。

3.3 脑间同步作为互动协同的辅助观测指标

如果说动态博弈为考察 ToM 的资源调节提供了行为范式，那么超扫描研究则为观察互动双方的协同过程提供了神经层面的辅助线索。与传统单脑研究不同，IBS 关注互动双方神经活动在时间上的耦合，可作为双方在注意、预测、策略协调和心理状态推断过程中是否形成协同节奏的辅助性观测指标。在合作、竞争和沟通等情境中，mPFC、TPJ 和 STS 等区域常出现与社会认知相关的同步现象(Azhari et al., 2025; Tang et al., 2016)。当任务需要理解他人意图和信念时，mPFC 与 TPJ 的跨脑耦合可能增强，并与策略协调和互动成功相关(Goodie et al., 2012; Lei et al., 2025)。在加工社会线索和预测他人行为的情境中，STS 和顶叶区域也可能出现同步活动，同时右角回(rAG)与欺骗情境中的心理表征效率相关(Liu et al., 2017;

Wang et al., 2020)。此外，在多轮互动或复杂决策任务中，前额叶控制网络与心智化网络的协同加强，可能反映个体在持续互动中整合信息、更新策略和调节认知资源的需求(De Dreu et al., 2020; Jiang et al., 2015; Zhang et al., 2023)。

IBS 水平也会随互动状态发生动态变化。当互动双方策略趋于一致、信息共享程度提高或共同注意增强时，同步程度可能随之提高(Khalvati et al., 2019; Kayhan et al., 2022)。相反，如果双方在推理投入、策略复杂度或信息更新节奏上明显不匹配，互动耦合可能下降。部分特殊群体研究也提示 IBS 与社会互动协同存在关联：ASD 或抑郁个体在互动任务中的 IBS 较低，并伴随策略协调能力下降(Li & Zhu, 2014; 何振宏, 张丹丹, 2018)。需要强调的是，动态互动中的 IBS 主要体现为互动双方神经活动的相关性，可能受到共同注意、联合行动、共享外部刺激和相似任务结构等影响(Czeszumski et al., 2020; Redcay & Schilbach, 2019; Azhari et al., 2025)。因此，本文将 IBS 定位为互动协同的条件性、辅助性观测指标。换言之，IBS 的理论价值不在于单独证明 ToM 机制，而在于为观察互动双方是否形成相似的资源调节节奏提供一个可能的神经过程指标。

综上，动态博弈为 ToM 的研究提供了以下三方面启示：第一，ToM 测量已经从静态能力评估转向资源约束下的动态调用；第二，动态博弈揭示了加工方式切换和递归深度有限性，说明个体会权衡认知成本与策略收益以选择适度推理；第三，IBS 可作为互动双方资源调节是否趋于协同的辅助观测指标，但不能被视为因果机制。现有研究仍缺乏一个能够同时整合加工方式、递归层级和互动协同的理论框架。有限理性正是在这一意义上成为关键接口：它既能够解释推理深度为何存在边界，也能够说明个体为何会在不同情境中调整加工投入，并为后续提出 ToM 资源适应性推理整合视角奠定基础。

新增参考文献：

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child development*, *72*(4), 1032-1053.

Jin, Y. (2021). Does level-k behavior imply level-k thinking?. *Experimental Economics*, *24*(1), 330-353.

Montgomery, D. E., Tompkins, V., & Feng, X. (2024). The theory of mind construct in adulthood: perspective taking in relation to language and executive function. *Frontiers in Psychology*, *15*, 1435685.

Mieth, L., Buchner, A., & Bell, R. (2021). Cognitive load decreases cooperation and moral punishment in a Prisoner's Dilemma game with punishment option. *Scientific Reports*, *11*(1), 24500.

Osterhaus, C., & Bosacki, S. L. (2022). Looking for the lighthouse: A systematic review of advanced theory-of-mind tests beyond preschool. *Developmental Review*, *64*, 101021.

Peloquin, C. M., McDermott, C. H., & Moses, L. J. (2023). Individual differences in second-order false-belief understanding and executive abilities: a meta-analytic review of evidence from school-age children and adults. *Cognitive Development*, *68*, 101367.

Quesque, F., & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on psychological science*, *15*(2), 384-396.

Zhao, Y., Li, Z., Jin, S., & Zhang, X. (2024). The impact of cognitive load on cooperation and antisocial punishment: Insights from a public goods game experiment. *Behavioral Sciences*, *14*(8), 638.

意见 3：部分行文细节欠严谨：个别段落行文容易让人误解双系统与递归深度是固定一一对应的关系，需要修改表述，明确二者为资源约束下的概率性匹配，不存在硬性绑定。

回应：感谢专家的严谨把关。将双系统与递归深度绑定确实会违背有限理性框架下的动态性特征。我们在文中进行了明确的界定和概念澄清，强调二者的关系是基于认知成本与预测收益权衡下的概率性匹配。

修改位置：5.1 协同机制的解释框架的第 2 段

修改稿：从该视角看来，高阶递归需要维持多重嵌套的信念，并持续更新对他人的心理

预测，这依赖工作记忆、执行控制和元表征等高成本认知资源，通常通过系统 2 的反思性加工来实现。相反，低阶推理则更多依赖启发式和自动化加工，以较低的认知成本快速推断他人的心理状态（包括意图、信念与愿望等），属于系统 1 的典型特征(Apperly & Butterfill, 2009)。但双系统和递归 ToM 并不存在一一对应关系，而是一种基于资源约束的概率性匹配：虽然低阶 ToM 通常以低成本的系统 1 运行，但在遇到新异环境或预期违背时，也可能需要系统 2 的抑制控制与审慎计算介入(Qureshi et al., 2010; Yoshida et al., 2008)；同样，高阶递归虽依赖反思性加工，但在熟练、重复互动的博弈情境中，个体可能通过强化学习不断优化策略表征，使原本高阶的推理经验固化，逐渐脱离系统 2，表现出自动化的系统 1 特征(Yoshida et al., 2008)。总之，系统 1 和系统 2 代表的是认知投入的动态调节，而非固定对应某个递归层级。

意见 4：建议全文论述进一步向资源适应性推理核心观点收拢，各章节均服务于模型的提出与论证，逻辑会更加紧凑完整。

回应：接受专家的宏观建议。为了使文章更具整体性，我们对全文的过渡段落和各章节的小结进行了全面梳理。

修改位置 1：第 2 节（心理理论研究的两个分析维度）：删减了部分冗长的经典 ToM 理论演进背景，集中于“双系统为何需要递归深度作为表征客体”以及“递归深度为何受限于双系统的资源分配”，直接为下文的断裂与整合铺垫。

修改位置 2：第 3 节（ToM 在动态博弈中的应用及脑间同步）：在每一小节的结尾增加了收拢性语句。例如，在 3.3 节探讨脑间同步时，明确指出：动态互动中的 IBS 主要体现为互动双方神经活动的相关性，可能受到共同注意、联合行动、共享外部刺激和相似任务结构等影响(Czeszumski et al., 2020; Redcay & Schilbach, 2019; Azhari et al., 2025)。因此，本文将 IBS 定位为互动协同的条件性、辅助性观测指标。换言之，IBS 的理论价值不在于单独证明 ToM 机制，而在于为观察互动双方是否形成相似的资源调节节奏提供一个可能的神经过程指标。

最后，再次感谢各位专家的悉心指导！我们已严格按照意见修改论文，如仍有需完善之处，敬请进一步指正。期待您的回复。

第二轮

外审专家 1 意见：

本文作者对上一轮审稿意见逐条进行了认真、细致的回应。针对我提出的核心关切——即文章中存在的“用领域一般性能力解释领域特殊性能力可能导致的还原论倾向”以及“范畴混淆风险”，作者也进行了有效的澄清与修正。修订后的文章在理论阐述上有了显著提升，对心理理论的资源适应性推理机制进行了富有创见的阐释，为相关领域的研究提供了新的理论视野。

外审专家 2 意见：作者认真回应了本人审稿意见，并虚心接受相关建议，对文章做了认真修改，解决了我的问题，特别是我也看到作者对编委及其他审稿人的回应也很认真专业，同意接受发表。

编委 1 复审意见：注意格式和写作规范完善。同意审稿人意见，接受发表。

编委 2 复审意见：本轮修改质量达标，作者对 17 条审稿意见（编委 2 条+外审专家 1 共 11 条+外审专家 2 共 4 条）全部进行了回应，核心学术问题已得到解决。同意发表。

主编终审意见：

论文选题有意义，但可读性有待进一步加强，一些表述有待推敲和澄清。

需要进一步澄清的内容举例：

意见 1：“这两个理论源于不同学科脉络、使用不同术语体系、关注 ToM 的不同侧面”，请明确指出各自的学科脉络区别、术语体系的差异、以及关注了 ToM 的哪些不同侧面。

回应：感谢主编的宝贵建议。我们已分别在 2.1 心理理论双系统：加工方式维度和 2.2 递归心理理论：递归深度维度部分补充了相关内容，明确指出了两者各自的学科脉络区别、术语体系的差异、以及关注了 ToM 的哪些不同侧面。

修改位置 1：2.1 节第 1 段。

修改稿：ToM 双系统主要形成于认知心理学、发展心理学与心灵哲学关于信念追踪机制的讨论，旨在解释个体为何既能快速、低成本地追踪他人的信念状态，又能进行缓慢但灵活的显性心理状态推理(Apperly & Butterfill, 2009; Butterfill & Apperly, 2013)。该理论通常使用“系统 1 与系统 2”、“高效系统与灵活系统”、“自动与控制加工”以及“内隐与外显信念追踪”等术语，描述心理状态信息加工的速度、自动性、灵活性与资源消耗(Apperly & Butterfill, 2009; Low & Perner, 2012; Low et al., 2016)。具体而言，ToM 由两个功能互补的系统构成：系统 1（即自动心智化）依赖启发式、快速、低成本加工他人的注意、视角与情绪线索；系统 2 则是较慢且成本较高的反思性推理，依赖工作记忆与执行功能建构心理状态表征(Apperly, 2010)。系统 2 在任务策略不透明、互动重复或需要预测复杂意图时提高推理质量，而系统 1 在时间紧迫、附加认知负荷高时以低成本维持效率(Apperly & Butterfill, 2009)。因此，ToM 双系统作为解释 ToM 运作的基础理论，为理解认知资源的分配提供了底层机制。

新增参考文献：

Low, J., Apperly, I. A., Butterfill, S. A., & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: A primer on the two-systems account. *Child Development Perspectives*, *10*(3), 184-189.

Low, J., & Perner, J. (2012). Implicit and explicit theory of mind: state of the art. *British Journal of Developmental Psychology*, *30*(1), 1-13.

修改位置 2：2.2 节第 1 段。

修改稿：递归 ToM 关注个体将信念、意图、欲望等心理状态嵌套于他人心理状态之中，形成多层表征的能力(Perner & Wimmer, 1985)。近期，Apperly(2025)对递归心智化即社会能力核心的批判，凸显了递归 ToM 作为独立理论的地位。递归 ToM 源于发展心理学与进化心理学两条脉络：前者以 Perner(1991)的表征 ToM 为代表，强调儿童通过元表征能力的发展，逐步从一阶错误信念推理（约 4 岁）到二阶错误信念推理（约 6-7 岁）的过渡，关注儿童理解嵌套信念的年龄进程；后者以 Corballis(2007)的递归思维假说和 Dunbar 学派的高阶意向性框架为代表，强调递归 ToM 在人类社会认知中的演化意义，并指出成人通常存在约 5 阶的认知上限(Kinderman et al., 1998; Oesch & Dunbar, 2017; Stiller & Dunbar, 2007)。相应地，递归 ToM 研究通常使用“一阶与二阶错误信念”、“高阶意向性”、“嵌套表征”、“递归层级”、“递归深度”等术语，主要描述心理状态嵌套的结构与阶数(Kinderman et al., 1998; Perner & Wimmer, 1985; Stiller & Dunbar, 2007)。因此，递归 ToM 侧重表征中嵌套了多少层主体及其心理状态，回答“推理多深”的问题。

意见 2：文中提到二者“长期各自平行发展”，又言本文“打破了以往将双系统与递归层级硬性绑定的误区”，不知何意。

回应：感谢主编指出这一表述上的矛盾。原表述确实容易引起误解。我们的本意是：以往文献中这两个理论是长期平行、相对独立发展的；而在少部分尝试结合两者的讨论中，容易陷入“低阶必定等同于系统 1、高阶必定等同于系统 2”的简单对应误区。我们已在“总结与

展望”（第 6 节第 1 段）中修改了表述，澄清为：“为缓解以往 ToM 双系统与递归 ToM 研究之间的解释断裂提供了理论接口，并澄清了将特定加工方式与特定递归层级简单对应的潜在误区(例如误认为系统 1 必定等同于 1-2 阶推理)。”。

修改位置：6 总结与期望

修改稿：本文在有限理性框架下提出 ToM 资源适应性推理整合视角，为联系加工方式与递归深度两个维度提供了理论接口。该视角将加工方式、递归深度和互动协同纳入同一逻辑主线，并将 IBS 视为互动协同的辅助观测指标：ToM 并非静态的能力特质，而是个体在认知资源约束与动态情境下，对加工方式与递归深度的概率性匹配与协同调节的适应过程，为缓解以往 ToM 双系统与递归 ToM 研究的解释断裂提供了理论接口，并澄清了将特定加工方式与特定递归层级简单对应的潜在误区(例如误认为系统 1 必定等同于 1-2 阶推理)。与此同时，本文提出的预测命题为后续从行为耦合、计算建模和跨脑神经活动等层面检验互动协同提供了路径。

意见 3：将“脑间同步(IBS)作为该过程的辅助观测指标”，但前面论述都是个体内的认知资源、加工方式与推理深度，而 IBS 是个体间的指标。且它和“资源适应性推理”有何必然关联？

回应：感谢主编的意见。我们已在引言第 4 段和 5.1 节尾段补充阐明了个体内指标与个体间指标的必然关联：“需要强调的是，ToM 资源适应性推理的核心机制是个体内的资源约束调节。在连续反馈中，个体调节结果通过策略与反应时序等外显化，成为对方更新资源调节策略的信号；双方逐渐匹配时，可能因共同加工与策略协调而增强跨脑耦合(Lei et al., 2025)。因此，IBS 仅作为资源调节策略是否趋向匹配的辅助指标，而非机制本身或因果证据，须在控制共同注意、共享任务结构与行为同步等混淆因素后，结合行为时序与计算建模证据解释(Azhari et al., 2025; Czeszumski et al., 2020; Hamilton, 2021)。

修改位置 1：引言第 4 段

修改稿：基于此，本文从有限理性出发提出 ToM 资源适应性推理整合视角，聚焦 ToM 双系统与递归 ToM 两个分析维度，将加工方式与递归深度整合为认知资源约束下的协同调节。在动态博弈中，互动双方会根据彼此的选择与反馈持续调整策略，这一协调过程可能表现为双方行为的时序依赖，并伴随神经活动的跨脑同步(Kayhan et al., 2022; Wang et al., 2020; Zhang et al., 2020)。因此，本文认为可将脑间同步(Inter-Brain Synchrony, IBS)作为该过程的辅助观测指标进行考察。本文首先梳理 ToM 双系统、递归 ToM 并解释其断裂，其次讨论动态博弈范式与 IBS 对 ToM 研究的启示，最后提出 ToM 资源适应性推理视角及三类可检验命题。

修改位置 2：5.1 协同机制的解释框架的尾段

修改稿：此外，在互动层面，IBS 可作为加工方式与递归深度协同调节的辅助观测指标。当双方在推理复杂度与资源投入方式上较接近时，更易出现稳定耦合和较高 IBS；反之则下降(Réveillé et al., 2024; Kayhan et al., 2022)。需要强调的是，ToM 资源适应性推理的核心机制虽然是个体内的资源约束调节，但是在连续反馈中，个体调节结果通过策略与反应时序等外显化，成为更新资源调节策略的信号；双方逐渐匹配时，可能因共同加工与策略协调而增强跨脑耦合(Lei et al., 2025)。因此，IBS 仅作为资源调节策略是否趋向匹配的辅助指标，而非因果证据，须在控制共同注意、共享任务结构与行为同步等混淆因素后，结合行为时序与计算建模解释(Azhari et al., 2025; Czeszumski et al., 2020; Ma & Tan, 2023; Hamilton, 2021)。

意见 4：“任务复杂度、时间压力和认知负荷共同影响可用认知资源状态”，认知负荷通常是由任务复杂度和时间压力共同导致，而不是一个独立的外部变量。三者并列不妥。

回应：感谢主编指出这一表述可能造成的概念混淆。我们在原文中的表述确实不够精细，容

易引发概念上的重叠感。在日常语境中，任务复杂度与时间压力都可能增强个体的“负荷感”，但从认知负荷理论和 ToM 实验研究来看，三者是概念上可区分、实验中可分别操纵且可能相互作用的任务约束。首先，任务复杂度取决于任务本身需要加工和整合的信息数量及其交互程度。例如，相较于一阶信念推理，二阶或多阶递归推理需要维持更多嵌套的心理状态表征，因而具有更高的计算需求。它与认知负荷理论中的内在负荷密切相关，但二者并不完全等同，因为内在负荷还取决于信息要素之间的交互程度以及个体已有的知识经验(Sweller et al., 2011)。其次，本文所称的“认知负荷”特指通过次级任务施加的附加负荷。例如，研究者要求被试在完成 ToM 或视觉观点采择任务的同时记忆无意义数字或进行数字奇偶判断，以额外占用工作记忆和执行控制资源(Maehara & Saito, 2013; Qureshi et al., 2010)。最后，时间压力是对任务完成时限的外部操纵，如比较限时反应与自定步调条件。它主要压缩信息提取、整合和检验的时间窗口，并不直接等同于对工作记忆存储容量的占用；相关研究也表明，时间压力会限制对他人心理状态的审慎思考，使判断更容易受到自我中心信息的影响(Todd & Tamir, 2024)。因此，任务复杂度、外加认知负荷与时间压力分别从计算需求、可用认知资源和加工时间三个方面约束 ToM 加工。三者均可能降低灵活、反思性 ToM 的参与程度，使个体更多依赖快速且低成本的加工方式，但其作用路径并不相同(Apperly & Butterfill, 2009; Apperly et al., 2009)。由此，三者在理论机制和实验操作上具有可区分性，但其影响可能相互叠加或交互。

为避免引起概念混淆，我们在修改稿中对这三个概念进行了更严谨的词汇界定，将“认知负荷”明确修改为“外加认知负荷”，同时补充了它们各自代表的限制属性，以更清晰地体现其独立的并列关系。

修改位置 1：引言第 3 段

修改稿：一是 ToM 双系统区分快速、低成本的自动心智化与缓慢、高成本的反思性推理，并考察任务复杂度（内在认知需求）、时间压力（加工时限约束）以及外加认知负荷（如双任务造成的资源占用）等如何共同限制可用的认知资源状态，进而改变个体的 ToM 加工方式(Apperly & Butterfill, 2009; Qureshi et al., 2010; Todd et al., 2019)。一是 ToM 双系统区分快速、低成本的自动心智化与缓慢、高成本的反思性推理，并考察任务复杂度（内在认知需求）、时间压力（加工时限约束）以及外加认知负荷（如双任务造成的资源占用）等如何共同限制可用的认知资源状态，进而改变个体的 ToM 加工方式(Apperly & Butterfill, 2009; Qureshi et al., 2010; Todd et al., 2019)。

修改位置 2：5.1 协同机制的解释框架第 3 段

修改稿：ToM 的理性边界受到个体资源、任务约束与环境结构的共同影响。个体资源包括工作记忆、执行控制能力及疲劳状态等，它们共同影响个体能够维持的加工投入与递归深度(Baddeley, 2003; Devine & Hughes, 2014; O’Grady et al., 2015; Van der Linden et al., 2003)；任务约束包括时间压力、外加认知负荷和任务复杂度，可能限制控制加工与高阶递归(Newman et al., 2024; Stanley, 2026; Todd et al., 2019; 宗清瑶 等, 2024)；环境结构则包括对手策略复杂度、可预测性、反馈质量和激励水平。

相关参考文献：

Apperly, I. A., & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, *116*(4), 953-970.

Apperly, I. A., Samson, D., & Humphreys, G. W. (2009). Studies of adults can inform accounts of theory of mind development. *Developmental Psychology*, *45*(1), 190-201.

Maehara, Y., & Saito, S. (2013). Cognitive load on working memory both encourages and discourages reasoning bias regarding the mental states of others. *Australian Journal of Psychology*, *65*(3), 152-159.

Qureshi, A. W., Apperly, I. A., & Samson, D. (2010). Executive function is necessary for perspective selection, not

Level-1 visual perspective calculation: Evidence from a dual-task study of adults. *Cognition*, *117*(2), 230-236.

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer.

Todd, A. R., & Tamir, D. I. (2024). Factors that amplify and attenuate egocentric mentalizing. *Nature Reviews Psychology, 3*(3), 164-180.

意见 5：P2 已出现递归心理理论，P5 才给出对应的英文。以上仅为几处举例，建议通篇修改后请他人做挑剔性阅读。

回应：感谢您的细致审阅。我们已将英文全称及缩写“(Recursive Theory of Mind, Recursive ToM)”提前至正文引言中第一次出现“递归心理理论”的位置，且后文统一写为递归 ToM。

通篇修改说明：

除上述逐条修改外，我们还进行了挑剔性校读并完成了以下整体调整：一是统一将 ToM 双系统与递归 ToM 表述为“加工方式”和“递归深度”两个分析维度，避免层级混同；二是将 IBS 严格限定为控制混淆变量后的辅助观测指标，行为耦合与计算建模作为互动协同的证据；三是核查并统一术语首次定义、章节标题、图注、标点及中英文格式。

再次感谢主编及各位专家的悉心指导。上述意见显著提升了本文的概念清晰度、论证严谨性和可读性。恳请继续批评指正。

第三轮

主编意见：作者做了认真回应和修改，同意发表。