心理科学进展 2027, Vol. 35, No. 2, 524–540 https://doi.org/10.3724/SP.J.1042.2027.0524
Advances in Psychological Science https://cstr.cn/32111.14.2027.0524

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*

钟永丹 肖 风

(贵州师范大学心理学院, 贵州省脑功能与脑疾病防治全省重点实验室, 贵阳 550025)

摘 要 心理理论(ToM)的加工方式与递归深度分别回答社会推理“如何加工”与“推理多深”, 但二者如何在认知资源约束下共同变化, 尚缺乏统一解释。本文基于有限理性提出“ToM 资源适应性推理”整合视角, 将加工方式与递归深度视为认知资源约束下相互关联、概率性匹配的两个维度, 认为个体会依据任务需求、可用资源、认知成本和预期收益调节加工投入与递归层级。该视角推导出三条可检验预测：(1)系统 2 参与程度与可维持的递归深度共同受任务需求、可用资源和预期收益约束, 二者的作用路径可通过实验操纵与模型比较检验; (2)在递归计算具有成本且对手复杂度有限的情境中, 递归深度与决策收益呈倒 U 形关系, 其最优点随环境结构变化; (3)在反馈连续且策略具有一定可预测性的互动中, 互动双方的资源调节策略会在连续反馈中趋向协同匹配。该视角为区分能力边界、策略选择及互动协同提供了可检验框架。

关键词 有限理性, 动态博弈, 心理理论双系统, 递归心理理论, 互动协同

分类号 B849: C91

1 引言

传统博弈论通常以完全理性人为前提, 假定个体拥有充分信息和无限计算能力, 能计算并选择最优策略。然而, 大量研究表明人类在博弈中的行为表现出有限理性(Bounded Rationality), 即个体受到信息、时间与认知资源等限制, 通常依赖启发式策略而非最优策略(Simon, 1955)。当决策环境由同样具有思维和意图的人类构成时, 这种有限理性便进入社会理性的讨论范畴(Gigerenzer, 2000)。社会理性是生态理性的一种特定形式, 强调在社会互动中, 理性并非孤立地追求个人收益最大化, 而是通过利用社会环境的结构和他人行为的规律性进行适应性决策(Gigerenzer & Selten, 2002)。这就要求个体不仅要权衡收益, 更要持续理解、预测并基于他人的信念、愿望及意图调整自身策略, 以实现互动协调与关系稳定。这种对他人心理状态的推理与追踪能力, 正是心理理论(Theory of Mind, ToM)的核心, 即个体理解他人信念、愿望及意图, 并据此预测其行为的能力, 也因此成为支持社会理性决策的关键认知基础(Baron-Cohen et al., 1985; Frith & Singer, 2008)。

ToM 相关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, 1995; 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 & Adolphs, 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, 1995; Chuey & Gweon, 2025;

收稿日期：2026-05-16
* 国家自然科学基金(32460208); 贵州省基础研究计划面上项目(黔科合基础 MS (2026)474)。
通信作者：肖风, E-mail: xiaofeng19850328@gmail.com

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)。

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, 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)等如何限制可用的认知资源状态, 改变个体的 ToM 加工方式(Apperly & Butterfill, 2009; Qureshi et al., 2010; Todd et al., 2019)。二是递归 ToM (Recursive Theory of Mind, Recursive 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; Oesch & Dunbar, 2017; Perner, 1991; Perner & Wimmer, 1985; Stiller & Dunbar, 2007)。加工方式与递归深度分别揭示了 ToM 的资源依赖性与递归边界, 但长期相对独立, 尚未充分说明两者在动态互动中的关联。由此产生两个关键问题：认知资源如何共同约束反思性加工的参与程度与可维持的递归深度？不同递归深度的策略收益为何随互动情境而变化？这些问题提示, ToM 表现可能并非由加工方式或递归深度单独决定, 而可能是两者在资源约束下的动态调节。

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, 该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)。因此, 本文将脑间同步(Inter-Brain Synchrony, IBS)作为该协同调节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预测。

2　心理理论研究的两个分析维度

2.1　心理理论双系统：加工方式维度

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 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 则是较慢且认知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、外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 的加工方式, 为理解认知资源的分配提供了加工层面的解释。

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, 2017; Kovács et al., 2010; Schurz et al., 2014; Yousefi et al., 2020; 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(INS)、杏仁核(AMY)等)往往先于系统 2(前扣带皮层(ACC)等)的缓慢反思调节, 特别在认知负荷较高时, 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(张慧 等, 2018)。孤独症谱系障碍(Autism Spectrum Disorder, ASD)研究发现自发性与显性 ToM 表现可能呈现不同受损模式, 但结果受任务要求、语言能力和执行功能等因素影响(Baron-Cohen et al., 2001; Senju et al., 2009), 说明 ToM 推理受任务需求与时间压力等约束进行满意性判断。

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(Nicholson et al., 2019; Van Overwalle & Vandekerckhove, 2013)。二者受语言经验、外加认知负荷和时间压力影响：(1)认知控制能力较强的双语经验者在复杂任务中能更有效平衡内隐与外显加工(Michael et al., 2018; Navarro & Rossi, 2022; Oktay-Gür et al., 2018; 王婷 等, 2019); (2)高负荷任务会消耗系统 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(Newman et al., 2024); (3)时间限制阻断了系统 2 的缓慢计算, 导致个体的二级视角采择任务表现显著下降, 但不影响依赖系统 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 的反思推理则额外依赖背外侧前额叶(DLPFC)和 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(Grosse Wiesmann et al., 2017; Hallajian et al., 2026; Rothmaler & Grosse Wiesmann, 2023; Schüler et al., 2026)。

从演化与发展视角看, ToM 双系统具有情境适应性。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(Bugnyar & Heinrich, 2005; Dally et al., 2006; Hare et al., 2000; Kano et al., 2019; Santos et al., 2006)。在人类发展中, ToM 表现可能经历由内隐、自动化的社会认知敏感性向外显、反思性心理状态推理的转变(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, 而在需要外显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。

ToM 是否对应两套相对独立的表征与加工机制仍存在争议。有研究指出, ToM 的自动化与控制性加工更可能源于单一心智化网络与通用认知资源(如执行控制、元表征)的动态交互, 而不是两个完全独立系统(Carruthers, 2017; Conway et al., 2017; Nicholson et al., 2019)。即便区分内隐与外显加工, 二者激活的核心脑区也大多重叠(Balgova et al., 2024; Jacob, 2019)。因此, ToM 更可能是在统一资源约束下, 权衡“算得快”与“算得准”的适应性调节。这可能意味着仅凭加工方式无法充分解释推理表现, 还需考虑个体实际构建了多深的心理状态表征。由此, 引出下文对递归 ToM 的考察。

2.2 递归心理理论：递归深度维度

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)。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。

递归 ToM 有理性边界。社会互动中的多层心智表征不能无限嵌套, 只能在认知资源允许的范围内展开(Krupenye et al., 2016; 隋光远, 刘聪, 2014)。一阶 ToM 为“我认为你相信 X”, 二阶为“我认为你认为我相信 X”, 三阶为“我认为你认为我认为你相信 X”, 依此类推(Kinderman et al., 1998)。个体在任务中可观察到较高阶表现, 但高阶表现高度依赖任务形式、语言能力、记忆负担与激励条件(O'Grady et al., 2015)。此外, 听障大学生自身语言条件会影响递归深度, 高工作记忆

容量个体的递归加工可达到 2~3 阶, 而工作记忆低容量个体仅能维持在 1 阶(宗清瑶 等, 2024)。因此, 递归深度并非无限扩展, 存在明确的递归上限。

递归 ToM 的认知成本与收益权衡。随着嵌套层级增加, 个体需要更多注意、记忆与执行控制资源, 以维持复杂心智表征和跟踪他人信念的变化(Hampton et al., 2008)。随着任务复杂度的提高, 心智化网络如 mPFC、TPJ、楔前叶(Precuneus)、后侧 STS 等区域广泛参与心理状态推断, 而 DLPFC 等控制区域可能额外参与需要维持、更新或控制复杂社会信息的任务(Frith & Frith, 2006; Kliemann & Adolphs, 2018; Schurz et al., 2014)。在博弈中, 较节省资源的低阶推理快速生成关于他人意图的判断, 而需要更多认知资源的高阶推理则维持嵌套信念间的逻辑一致性并预测他人策略变化(Zhen & Yu, 2021)。因此, 递归深度可能受到认知成本与预期收益的共同调节, 个体未必总是采用其能力允许的最高层级。

递归 ToM 的适应性, 一方面体现在物种进化: 类人猿在某些非言语预期任务中表现出对错误信念的敏感性, 黑猩猩和灌丛鸦等非人动物能够根据竞争者的视线、知觉经验或知识状态调整行为, 但通常难以通过二级视角采择任务, 表明其可能依赖高效但受限、缺乏灵活性的心理状态追踪系统(Apperly & Butterfill, 2009; Krupenye et al., 2016)。另一方面, 递归 ToM 表现随年龄发展: 儿童通常在学前期逐渐通过一阶错误信念任务, 随后发展出二阶错误信念理解(Gopnik & Astington, 1988; Perner & Wimmer, 1985; Wellman et al., 2001)。成人虽能完成更高阶的心理状态嵌套任务, 但表现通常随嵌套层级增加而下降, 其可维持的递归深度还受到任务形式、语言能力和工作记忆负荷等因素影响(Kinderman et al., 1998; Stiller & Dunbar, 2007)。同时发展阶段影响个体的策略使用, 如儿童更多依赖低阶 ToM 和直觉判断, 并随年龄增长 ToM 会更有益于策略性决策(Castelli et al., 2014)。由此可见, 递归深度的选择受到认知能力和资源条件的限制, 可能需权衡认知成本与策略收益。

2.3　两个分析维度的解释断裂与整合需求

ToM 双系统与递归 ToM 分别从加工机制和递归深度两个维度解释社会推理。前者关注个体在资源约束下如何切换系统 1 与系统 2, 回答“以何种方式推理”的问题; 后者关注个体如何构造多层嵌套的心理表征, 回答“推理能达到何种深度”的问题。然而, 二者长期相对独立, 缺乏清晰的理论接口: 双系统研究较少说明加工方式切换对应的是何种层级的社会表征, 递归 ToM 研究也较少解释个体为何会在不同资源状态下调整推理层级。由于两类研究在范式、指标和解释语言上不一致, 加工方式与递归深度的关系往往只能被间接推断, 未形成关于何种情境会诱发何种推理组合的统一说明。

现有框架不能完全填补这一空白。资源限制说虽指出高负荷会降低认知表现, 但往往停留在资源少则表现差的宽泛描述上, 难以解释系统切换和递归层级之间的关系(Kahneman, 1973)。认知层级模型、level-k 模型等模型能估计个体的递归深度分布, 却常把层级当作相对静态的行为参数, 较少讨论不同层级通过何种加工方式实现, 更难说明加工方式如何嵌入层级选择(Camerer et al., 2004; Goodie et al., 2012)。因此, 真正需要的不是在双系统加工方式与递归深度之间简单搭桥, 而是一个能同时回答三个问题的统一框架: 其一, 为什么特定情境会触发某种加工方式; 其二, 为什么这种加工方式会支持某种递归深度; 其三, 加工方式与递归深度如何表现为策略收益和互动耦合差异。有限理性为回答这三个问题提供了框架。

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。传统任务虽有助于界定个体是否具备基本心理状态理解能力, 但难以呈现真实互动中持续反馈、策略调整、认知负荷和收益权衡共同作用。而动态博弈提供了同时操纵资源约束、对手复杂度和反馈结构的实验环境, 使研究者能够考察 ToM 何时被调用及其递归深度的变化(见表 1)。

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、

表 1 静态与动态 ToM 测量范式及其资源约束

范式类别	核心维度	代表性范式	考察内容	资源约束	关键参考文献
静态离线测验	基础信念理解	错误信念任务、观点采择任务	理解他人信念与现实或自身信念不同；从他人空间视角判断物体。	涉及对他人视角的抑制性控制，任务结构单一。自动化条件下需求极低，外显报告时依赖抑制控制。	李艳，徐胜, 2024; Perner & Wimmer 1985; Samson et al., 2010; Tian et al., 2021;
	高阶意图与语言推理	提示任务、奇异故事任务	理解间接请求、讽刺、谎言及话语背后的真实意图。	维持多重非字面语义，并利用执行功能进行上下文整合。语言依赖性强，对加工速度有要求。	Corcoran et al., 1995; Happé, 1994
	复杂社会情境协调	失礼识别测试、眼神读心任务	识别无意冒犯行为；从微弱情绪线索中辨别心理状态。	同时追踪多个角色的信念与情感，涉及复杂执行控制协调冲突信息。	Baron-Cohen et al., 2001; 苏彦捷，马天舒, 2014
动态互动博弈	信念建构与策略推理	选美博弈、斯塔克伯格博弈、同时行动博弈	测量递归深度；评估个体对他人平均行为的预测精度。	受个体工作记忆限制，推理深度有上限，随任务结构和激励等变化。	Nagel, 1995; Von Stackelberg, 2010
	合作意图与信任推理	信任博弈、囚徒困境、猎鹿博弈	考察在不确定环境下对他人的信任预测、合作动机及共同信念的协调能力。	个体需权衡收益与认知成本，认知负荷可能改变合作、信任和反馈整合。	Cheng et al., 2022; Skyrms, 2004; Thompson et al., 2021
	公平感知与拒绝预期	独裁者博弈、最后通牒博弈	评估对公平规范的感知及对他人拒绝行为的预判。	面临不公平分配时，系统 1(直觉愤怒)常与系统 2(收益计算)冲突，时间压力会强化系统 1。	Wei et al., 2022; 张慧等, 2018
	竞争推理与策略对抗	石头剪刀布博弈、棋盘类博弈	追踪在策略对抗中的实时信念更新及对对手策略复杂度的动态适应。	受预测误差和对手不确定性驱动，个体实时调整推理投入，防止过度递归导致认知枯竭。	Wang et al., 2014; Weimer et al., 2022

意图、情绪和社会语境的理解水平，但任务结构相对固定，反馈有限，更多呈现个体在特定任务要求下的表现。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(Carlson & Moses, 2001; Rakoczy, 2022)。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要求更高(Montgomery et al., 2024; Osterhaus & Bosacki, 2022)。第三，复杂社会情境任务需要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，而是与认知资源相关(Peloquin et al., 2023)。但静态范式的资源约束通常被任务结构弱化，难以区分心理状态表征能力与资源限制对表现的影响。

动态博弈将ToM置于更接近真实互动的情境中，凸显 ToM 的选择是信念、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(Jin, 2021)。在信任博弈、囚徒困境和猎鹿博弈中，个体需权衡合作意图推断、公平偏好、策略收益与认知负荷，以维持合作、表达信任和协调选择(Mieth et al., 2021; Zhao et al., 2024)。在独裁者博弈和最后通牒博弈中，个体的公平决策通常需要权衡情绪性公平反应与自利性收益计算。最后通牒博弈的眼动研究表明，公平偏好与收益计算可能分别更多依赖自动化加工和控制性加工(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，避免过度递归造成资源需求超过可用容

量。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、互动博弈情境的拓展。静态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、不确定性和收益权衡中, 如何分配认知资源以选择加工方式和调节递归深度(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, 测量范式的拓展反映了 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递归深度有限性

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, 时间限制可能增强直觉性公平反应对决策的影响, 使个体更倾向于拒绝不公平提议; 随着反思加工增加, 个体则更可能权衡拒绝的成本并接受不公平提议(Calvillo & Burgeno, 2015; Neo et al., 2013; Wei et al., 2022); 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)。加工方式的切换受到环境结构和预测误差共同调节: 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), 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, 动态博弈中加工方式的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中, 个体需递归推理他人可能的选择, 但受工作记忆和计算能力限制, 递归深度通常集中于有限层级, 如在选美博弈中, 大多数参与者的递归深度集中于 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), 表明个体在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。递归深度随任务复杂度、对手不确定性和工作记忆等因素动态变化: 当环境更复杂或对手行为更难预测时, 个体可能适度提高递归层级以增强对他人行为的预测(李晶, 刘希平, 2009; De Weerd et al., 2015; Markiewicz et al., 2024)。但这种提升并非线性递增: 随着递归层级不断加深, 个体的认知资源消耗上升以维持更多嵌套信念和持续更新预测, 但当推理成本超过收益时, 过度递归反而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 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

动态博弈为考察 ToM 的资源调节提供了行为范式, 超扫描则为观察互动协同提供神经层面的辅助线索。合作、竞争与交流可能伴随 mPFC、TPJ 及前额叶等社会认知相关区域的 IBS, 但这种同步也可能源于共同刺激、注意和行为同步, 不能直接证明双方采用了相同的加工方式或递归层级(Czeszumski et al., 2022; Hamilton, 2021; Tang et al., 2016)。因此, 只有在行为时序和计算模型已识别潜在推理状态, 并控制共同输入等混淆因素后, IBS 才能作为双方信息整合、相互预测与策略协调的辅助神经指标。

IBS 水平随互动状态动态变化。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(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 症状较重或抑郁倾向较高的个

体在互动任务中表现出较弱的行为协调，右侧 TPJ 的 IBS 也相应降低(Kruppa et al., 2021; 张丹丹 等, 2020)。需要强调的是，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(Azhari et al., 2025; Czeszumski et al., 2020; Redcay & Schilbach, 2019)。因此，本文将 IBS 定位为互动协同的辅助性观测指标，为观察互动双方是否形成相似的资源调节策略提供可能指标。

综上，动态博弈为 ToM 的研究提供了以下三方面启示：第一, ToM 测量正在从静态能力评估转向资源约束下的动态调用；第二，相较于静态测量范式，动态博弈范式更能揭示加工方式切换和递归深度有限性；第三, IBS 可作为互动双方资源调节策略是否趋于协同的辅助观测指标，但不能视为因果机制。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视角。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，也能说明个体为何会在不同情境中调整加工投入，为后续提出 ToM 资源适应性推理整合视角奠定基础。

4 有限理性作为整合框架

将有限理性引入 ToM 研究，有助于解释个体如何依据认知资源与互动需求调整加工方式和推理层级，为解释为何在复杂互动中采用“足够好”推理提供框架，也便于将递归 ToM 建模与检验。有限理性修正了传统博弈论的完全理性假设，即人类在现实决策中受到信息不完整、认知能力有限以及时间和注意资源不足等约束，通常难以穷尽所有可能的计算，更倾向于选择“足够好”方案(Simon, 1976)。而社会理性强调个体的理性决策嵌入人际关系与互动结构之中，旨在提升个体收益的同时促进与他人的协调、维持社会关系的稳定及群体层面的适应(Gigerenzer & Selten, 2002)。在此基础上，有限理性强调理性的边界可测量性、资源分配机制与情境适应性，共同构成一个系统化、可操作的有限理性框架(刘永芳, 2022)。

有限理性体现为可量化的认知边界、资源分配机制与情境适应性。工作记忆容量、信息处理能力及可承受的计算复杂度共同限定个体能够维持的推理层级，博弈者的策略推理通常也集中于较低层级(Camerer et al., 2004)。由于认知资源有限，个体难以同时处理大量信息和穷尽所有策略，因而会权衡预期收益与计算成本，选择“足够好”的启发式方案。由此产生的近似推理或认知偏差并非必然意味着非理性，而可能是有限资源下提高决策效率的适应性结果(Lieder & Griffiths, 2020)。此外，个体还会根据任务复杂度、环境不确定性和既往经验调整认知投入：在熟悉、稳定的情境中更依赖低成本策略，在潜在收益提高且资源允许时则增加审慎加工(Todd & Gigerenzer, 2012)。因此，有限理性并非单纯的能力不足，而是个体根据资源限制与情境需求动态调节推理投入和策略复杂度的过程。

有限理性为连接 ToM 的加工方式与递归深度提供了理论接口。ToM 双系统与递归 ToM 虽处于不同分析维度，但均受可用资源、认知成本与预期收益影响。由此，个体在社会互动中采用何种加工方式、维持何种递归深度，并非彼此独立或固定对应，而可能是根据任务需求与环境结构形成的适应性组合。基于这一逻辑，下文将进一步提出 ToM 资源适应性推理视角，阐明加工方式与递归深度在资源约束下如何相互关联，并据此推导可检验预测。

5 有限理性视角下的 ToM 整合：推理方式与递归深度的协同调节

5.1 协同机制的解释框架

ToM 双系统与递归 ToM 分别解释“如何加工”和“推理多深”，但未说明二者如何随资源与情境共同变化(Apperly & Butterfill, 2009; Camerer et al., 2004)。为此，本文提出“ToM 资源适应性推理”(Resource-Adaptive ToM Reasoning)整合视角(见图 1)，将加工方式与递归深度视为相互关联但非一一对应的两个维度。该视角认为，个体会依据可用资源、认知成本和预期收益，动态调整加工投入与递归层级，这种动态调整反映了心理状态表征与一般执行控制资源的共同作用，可解释资源不足时高阶推理表现受限并转向低成本策略的现象。相较于既有框架，本视角的新增解释对象不是“个体是否投入更多资源”本身，而是资源状态如何分别影响加工方式与递归深度、二者如何形成概率性组合，以及这种组合如何在连续互动中形成跨个体的状态依赖。

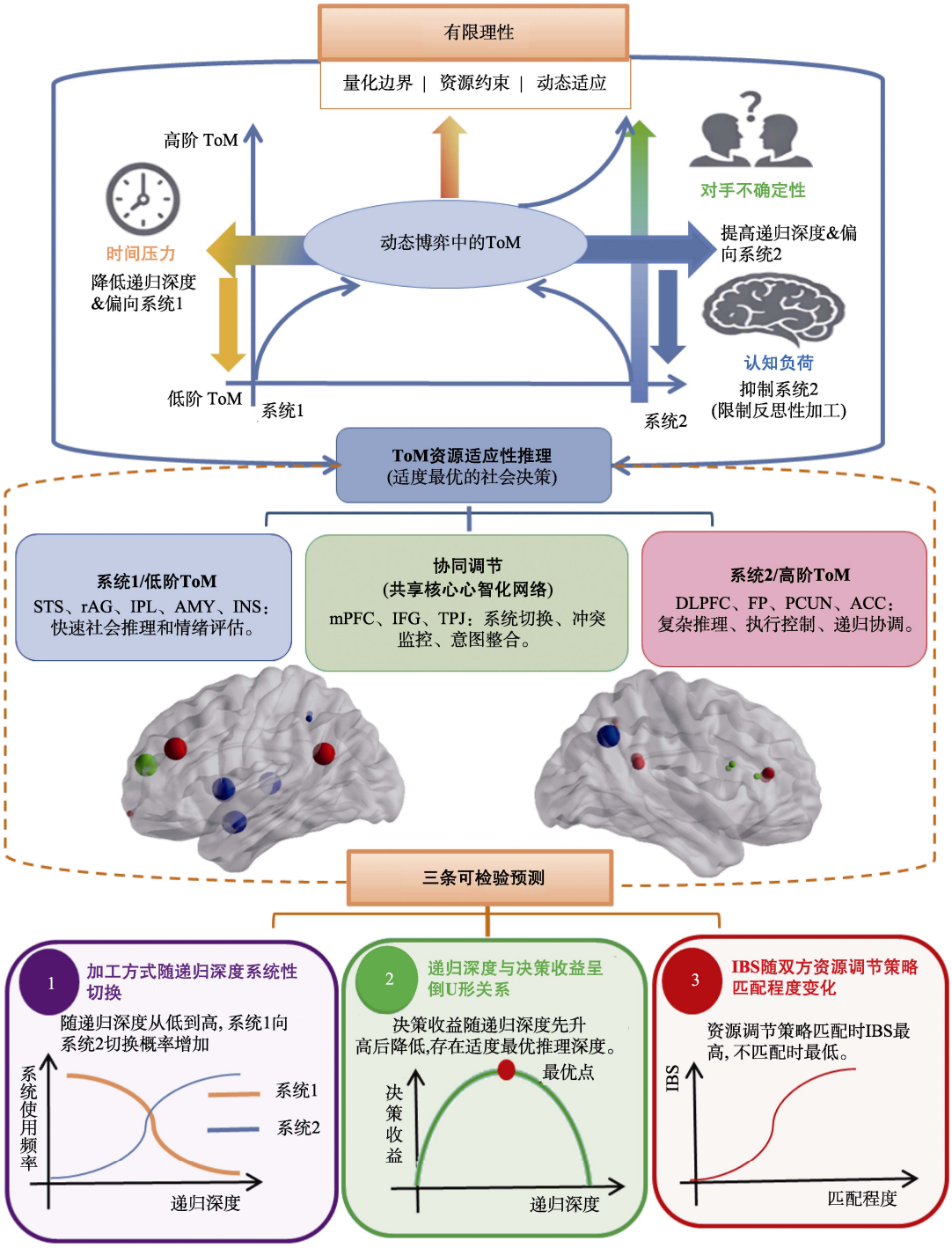


图 1　ToM 资源适应性推理与可检验预测

注：TPJ: 颞顶联合区; mPFC:内侧前额叶皮层; DLPFC: 背外侧前额叶皮层; IFG: 额下回; STS: 颞上沟; IPL:顶下叶; rAG: 右角回; ACC: 前扣带回皮层; AMY: 杏仁核; INS: 岛叶; FP: 额极; PCUN: 楔前叶。脑网络可视化图采用 BrainNet Viewer 软件绘制(Xia et al., 2013)。

从该视角看来，随着递归层级提高，维持和更新嵌套心理状态通常需要更多工作记忆与执行控制资源，因此更可能伴随控制性、反思性加工投入增加；较低阶推理则更容易借助熟悉线索和启发式策略完成(Apperly & Butterfill, 2009)。但这种对应具有概率性和情境依赖性，而非固定映射：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., 2020; 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选择，使原本高阶递归经验固化，表现出自动化的系统 1 特征(Yoshida et al., 2008)。总之，加工方式的切换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。

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。在稳定且可预测的重复互动中，低成本启发式和较低阶推理通常足以支持有效预测(Apperly & Butterfill, 2009; Halder et al., 2015)；当对手策略更复杂、竞争或激励增强、预测误差持续扩大时，提高控制加工投入与递归深度的潜在收益可能增加(Colman, 2003; Hedden & Zhang, 2002; Zhen & Yu, 2021)。因此，个体可能根据自身资源、任务要求与环境结构，动态调节加工投入和递归深度。

这一视角获多层面研究证据的间接支持。在发展层面，儿童由低阶直觉推理逐步发展出依赖反思加工的高阶递归，反映认知资源积累对递归深度的约束(Perner & Wimmer, 1985; Valle et al., 2015; Wellman et al., 2001)。在临床层面，部分 ASD 个体在内隐心理状态加工中的表现可能相对保留，而外显及二阶错误信念推理更易受到语言、工作记忆和递归表征限制，且具有明显的任务依赖性(Lu et al., 2024; Polyanskaya et al., 2022)。在行为层面，递归深度随激励和对手策略复杂度增加，却在资源受限时下降，提示其存在适应性上限(Anufriev et al., 2024; Camerer et al., 2004; Ridinger & McBride, 2025)。在神经层面，复杂心理推理依赖心智化网络与控制网络的动态耦合：高负荷、高递归任务会增强 DLPFC 与核心网络的协同，而认知过载则可能削弱前额叶协同(Meyer & Collier, 2020; Yang et al., 2022; Zhang et al., 2024; Zheltyakova et al., 2026)。这些证据提示，加工方式与递归深度可能随认知资源共同变化，但具体作用路径仍需检验。

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(Kayhan et al., 2022; Réveillé et al., 2024)。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; Hamilton, 2021; Ma & Tan, 2023)。

5.2 可检验预测

ToM 资源适应性推理将双系统加工方式与递归深度统一为资源约束下的协同调节，即个体在动态博弈中的加工方式、递归层级以及互动协调状态随资源状态与环境需求动态调整。在此基础上，本文提出三条可检验预测。

命题一(资源阈值命题)：系统 2 参与程度与递归深度由任务需求、可用资源和预期收益共同决定。当资源充足且继续推理有助于提高收益时，任务复杂度增加将促进系统 2 投入和递归加深；当计算需求超过可用资源，或推理的边际收益低于认知成本时，二者将停止上升或出现回落(Apperly, 2010; Apperly & Butterfill, 2009; Lieder & Griffiths, 2020)。高阶递归并不等同于系统 2，但随递归层级提升系统 2 参与概率系统性提高。本文预测，认知资源同时约束系统 2 参与程度与

可维持的递归深度, 二者之间的作用路径及方向可通过实验操纵与模型比较加以检验(Hedden & Zhang, 2002; Krockow et al., 2016; von der Osten et al., 2017)。未来可比较三种模型：模型 A 认为资源通过改变加工方式的相对参与间接影响递归深度(Apperly & Butterfill, 2009; Qureshi et al., 2010); 模型 B 认为资源分别约束加工方式与递归深度, 二者相关但无必然中介(Lieder & Griffiths, 2020); 模型 C 认为递归需求升高会增加系统 2 投入(Hedden & Zhang, 2002)。通过分别操纵认知负荷与对手策略复杂度, 检验各模型的解释力。若操纵可用资源只影响一般反应速度而不改变递归模型参数; 或递归需求增加并未提高控制加工指标, 则相应模型不获支持。

命题二(收益与深度非线性命题)：递归深度与决策收益之间并非单调正相关, 而是倒 U 形关系, 即适度递归在大多数博弈情境下收益最优, 而过度递归因认知耗竭将导致效率下降。该命题可通过操纵对手策略复杂度, 结合外显信念报告、层级化对手模型与选择数据估计递归深度。操作化检验采用分层线性模型, 考察递归深度与收益的二次项系数是否显著为负。在计算建模层面, 可通过分层贝叶斯建模与动态信念更新方法, 参数化刻画个体如何根据环境不确定性调节 ToM (Buergi et al., 2026), 为从固定递归层级转向动态调节推理投入的建模思路提供方法学支持。若递归深度与收益呈单调正相关(即二次项系数不显著或为正), 或在所有博弈情境中均无倒 U 形拐点, 则该命题被否证。

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, 即互动双方的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, 若双方的递归深度相近, 其策略选择、策略切换和反应时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, 借鉴离散潜状态模型识别个体随时间变化的策略状态(Ashwood et al., 2022)。若双方潜状态序列或转移矩阵更相似, 可作为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的证据之一(Pan et al., 2025)。(3)神经层面, 可在控制共同注意、共享任务结构和行为同步等因素后, 将 IBS 作为不同潜在推理状态下功能网络耦合模式的状态依赖的辅助观测指标(Czeszumski et al., 2020; Hamilton, 2021)。该预测可借鉴动态因果建模、滞后连接分析或其他方向性连接方法检验其状态依赖性和方向性(Bilek et al., 2022)。因此, 行为耦合与计算建模构成支持互动双方资源调节策略趋向协同匹配的主要证据, 而在控制共同输入后 IBS 可作为辅助指标。

综上, 三条命题共同构成了 ToM 资源适应性推理的检验：资源阈值命题揭示资源状态如何影响加工方式与递归深度, 收益与深度非线性命题说明高阶推理并非总是最优, 互动协同观测命题将个体的资源调节扩展到互动的协同匹配。由此, 后续研究可通过行为操纵、计算建模与超扫描等联合验证, 考察个体如何动态平衡认知成本、递归深度与收益。

6　总结与展望

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。

本文提出的整合视角仍存在局限：首先, 加工方式与递归深度的协同机制及具体影响仍需通过更精细的实验设计验证。其次, IBS 在 ToM 推理中的功能定位有待进一步探究。最后, 本文所提出的视角在操作化测量与计算建模层面仍有待发展。未来可从以下几个方面对 ToM 资源适应性推理进行拓展和检验：

推理调节阈值的量化检验。ToM 资源适应性推理预测个体在不同情境下对加工方式与递归深

度的调整可能存在可量化的调节阈值, 该阈值由预期策略收益与认知成本的相对权衡决定。当环境复杂度或对手策略不确定性超过一定水平时, 个体可能增加推理复杂度与加工投入; 反之, 则趋向于采用低成本策略。未来研究可系统操纵对手策略复杂度, 如在选美博弈中逐步提高对手行为的层级, 结合反应时、眼动及超扫描技术, 识别行为与神经层面的变化拐点检验该阈值(Camerer et al., 2004; Coricelli & Nagel, 2009)。同时, 可结合认知负荷操纵考察该阈值是否随资源变化动态调整(Newman et al., 2024; Todd et al., 2019)。

递归复杂度的适应性收益函数。该视角认为, 递归深度与决策收益之间可能呈现非线性关系。在高噪声或低可预测性的环境中, 过高的推理复杂度反而易导致过拟合和策略退化, 而适度降低复杂度却可能提升决策表现(Anufriev et al., 2024; Ridinger & McBride, 2025)。此外, 额极皮层对适应性学习信号和层级推理信息的整合, 提示大脑在动态环境中能够自动调节推理策略的复杂度, 以更好地适应收益变化(Feng et al., 2026)。未来研究可以通过操纵环境可预测性和对手行为结构, 来检验递归复杂度与收益的函数关系, 同时结合计算建模来刻画适度最优推理深度如何随环境变化而动态调整(Friston et al., 2017; Yoshida et al., 2008)。

互动协同机制的严谨检验。该视角关注的核心并非 IBS 本身, 而是互动双方是否在连续反馈中形成推理策略、资源投入和更新节奏的协同匹配。未来研究可在多轮互动任务中结合行为时序、策略转换和计算建模, 刻画双方推理层级与潜在资源调节状态的匹配程度; 在此基础上, IBS 可作为辅助指标, 用于考察协同匹配是否伴随跨脑神经活动的条件性耦合。为避免将共享刺激、共同任务结构或单纯行为同步误解为推理协同, 研究需设置严格控制条件, 并在排除关键混淆后谨慎解释 IBS 指标(Czeszumski et al., 2020)。

面向临床与干预的应用拓展。ToM 资源适应性推理为社交障碍群体的干预和人工智能(AI)的发展提供理论视角。未来可检验部分 ASD 个体的社会认知困难是否不仅涉及心理状态表征差异, 也涉及根据情境需求调节加工投入(如僵化使用低阶启发式)和推理复杂度的困难(如对简单情境过度分析), 以探索在传统 ToM 训练之外增加情境识别和推理深度调节训练的可能性。此外, 还可探索将 ToM 资源适应性推理应用于 AI, 在 AI 的 ToM 模型中引入计算成本和深度调节参数, 检验资源受限模型是否比无成本的无限递归模型更能预测人类行为(Baker et al., 2017; Li et al., 2025)。

综上, ToM 资源适应性推理为系统理解人类在复杂社会互动中的行为提供了整合性解释框架, 未来研究正逐步从是否具备 ToM 转向如何在资源约束下动态调节 ToM, 并结合计算模型与跨脑神经指标实现多层级整合(Buergi et al., 2026; Friston et al., 2017), 为后续研究社会互动中的资源调节、递归深度和策略协调提供可检验的理论路径。

参考文献

李晶, 刘希平. (2009). 对手推理水平的转变对心理理论推理水平发展趋势的影响. *第十二届全国心理学学术大会论文摘要集*. 济南.

李艳, 徐胜. (2024). 孤独症谱系障碍个体视觉观点采择研究进展. *中国特殊教育*, (8), 44−51.

李艺, 肖风. (2021). 自动观点采择: 内隐心智化与潜心智化的争议. *心理科学进展*, *29*(10), 1887−1900.

刘永芳. (2022). 有限理性的本质辨析与价值之争. *心理学报*, *54*(11), 1293−1309.

苏彦捷, 马天舒. (2014). 孤独症个体的公平决策: 心理理论和自我/他人情绪识别的作用. *心理科学*, *37*(4), 985−991.

隋光远, 刘聪. (2014). 策略性博弈中的心理理论推理研究评述. *洛阳师范学院学报*, *33*(3), 121−124.

王婷, 王丹, 张积家. (2019). 双语与双言影响心理理论内隐系统与外显系统的分离. *心理与行为研究*, *17*(4), 442−451.

张丹丹, 王驹, 赵君, 陈淑美, 黄琰淋, 高秋凤. (2020). 抑郁倾向对合作的影响: 双人同步近红外脑成像研究. *心理学报*, *52*(5), 609−622.

张慧, 马红宇, 徐富明, 刘燕君, 史燕伟. (2018). 最后通牒博弈中的公平偏好：基于双系统理论的视角. *心理科学进展, 26*(2), 319−330.

宗清瑶, 宫慧娜, 邓柳. (2024). 不同语言条件下工作记忆对听障大学生心理理论机制的影响. *中国特殊教育*, (11), 39−48.

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.

Anufriev, M., Duffy, J., & Panchenko, V. (2024). Individual evolutionary learning in repeated beauty contest games. *Journal of Economic Behavior & Organization*, *218*, 550−567.

Apperly, I. A. (2010). *Mindreaders: The cognitive basis of*

"theory of mind". Psychology Press.

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

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.

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.

Azhari, A., Rai, A., & Chua, Y. V. (2025). A systematic review of inter-brain synchrony and psychological conditions: Stress, anxiety, depression, autism and other disorders. *Brain Sciences*, *15*(10), 1113.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*(10), 829−839.

Baker, C. L., Jara-Ettinger, J., Saxe, R., & Tenenbaum, J. B. (2017). Rational quantitative attribution of beliefs, desires and percepts in human mentalizing. *Nature Human Behaviour*, *1*(4), 0064.

Balgova, E., Diveica, V., Jackson, R. L., & Binney, R. J. (2024). Overlapping neural correlates underpin theory of mind and semantic cognition: Evidence from a meta-analysis of 344 functional neuroimaging studies. *Neuropsychologia*, *200*, 108904.

Baron-Cohen, S. (1995). *Mindblindness: An essay on autism and theory of mind*. MIT Press.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37−46.

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" test revised version: A study with normal adults, and adults with Asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry*, *42*(2), 241−251.

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.

Buergi, N., Aydogan, G., Konovalov, A., & Ruff, C. C. (2026). A neural signature of adaptive mentalization. Nature Neuroscience, *29*, 934−944.

Bugnyar, T., & Heinrich, B. (2005). Ravens, Corvus corax, differentiate between knowledgeable and ignorant competitors. *Proceedings of the Royal Society B: Biological Sciences*, *272*(1573), 1641−1646.

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.

Butterfill, S. A., & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language*, *28*(5), 606−637.

Calvillo, D. P., & Burgeno, J. N. (2015). Cognitive reflection predicts the acceptance of unfair ultimatum game offers. *Judgment and Decision Making*, *10*(4), 332−341.

Camerer, C. F., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. *The Quarterly Journal of Economics*, *119*(3), 861−898.

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

Carruthers, P. (2017). Mindreading in adults: Evaluating two-systems views. *Synthese*, *194*(3), 673−688.

Castelli, I., Massaro, D., Bicchieri, C., Chavez, A., & Marchetti, A. (2014). Fairness norms and theory of mind in an ultimatum game: Judgments, offers, and decisions in school-aged children. *PLOS ONE*, *9*(8), e105024.

Cheng, X., Zhu, Y., Hu, Y., Zhou, X., Pan, Y., & Hu, Y. (2022). Integration of social status and trust through interpersonal brain synchronization. *NeuroImage*, *246*, 118777.

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

Colman, A. M. (2003). Depth of strategic reasoning in games. *Trends in Cognitive Sciences*, *7*(1), 2−4.

Conway, J. R., Lee, D., Ojaghi, M., Catmur, C., & Bird, G. (2017). Submentalizing or mentalizing in a Level 1 perspective-taking task: A cloak and goggles test. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 454−465.

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.

Corcoran, R., Mercer, G., & Frith, C. D. (1995). Schizophrenia, symptomatology and social inference: Investigating "theory of mind" in people with schizophrenia. *Schizophrenia Research*, *17*(1), 5−13.

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(23), 9163−9168.

Czeszumski, A., Eustergerling, S., Lang, A., Menrath, D., Gerstenberger, M., Schuberth, S., ... König, P. (2020). Hyperscanning: A valid method to study neural inter-brain underpinnings of social interaction. *Frontiers in Human Neuroscience*, *14*, Article 39.

Czeszumski, A., Liang, S. H. Y., Dikker, S., König, P., Lee, C. P., Koole, S. L., & Kelsen, B. (2022). Cooperative behavior evokes interbrain synchrony in the prefrontal and temporoparietal cortex: A systematic review and meta-analysis of fNIRS hyperscanning studies. *eNeuro*, *9*(2), ENEURO.0268−21.2022.

Dally, J. M., Emery, N. J., & Clayton, N. S. (2006).

Food-caching western scrub-jays keep track of who was watching when. *Science*, *312*(5780), 1662−1665.

Devine, R. T., & Hughes, C. (2014). Relations between false belief understanding and executive function in early childhood: A meta-analysis. *Child Development*, *85*(5), 1777−1794.

De Weerd, H., Verbrugge, R., & Verheij, B. (2015). Higher-order theory of mind in the tacit communication game. *Biologically Inspired Cognitive Architectures*, *11*, 10−21.

Feng, J., Jin, J., Zhao, S., Derrington, E., Qin, X., Fu, S., Shen, Q., & Dreher, J.-C. (2026). Neural correlates of interactions between adaptive learning and hierarchical reasoning in repeated strategic games. Games and Economic Behavior, 158, 167−190.

Freire, I. T., Arsiwalla, X. D., Puigbò, J.-Y., & Verschure, P. (2023). Modeling theory of mind in dyadic games using adaptive feedback control. *Information*, *14*(8), 441.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017). Active inference: A process theory. *Neural Computation*, *29*(1), 1−49.

Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron*, *50*(4), 531−534.

Frith, C. D., & Singer, T. (2008). The role of social cognition in decision making. Philosophical Transactions of the Royal Society B: Biological Sciences, 363(1511), 3875−3886.

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

Gigerenzer, G. (2000). Adaptive thinking: Rationality in the real world. Oxford University Press.

Gigerenzer, G., & Selten, R. (Eds.). (2002). *Bounded rationality: The adaptive toolbox*. MIT Press.

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

Goodie, A. S., Doshi, P., & Young, D. L. (2012). Levels of theory-of-mind reasoning in competitive games. *Journal of Behavioral Decision Making*, *25*(1), 95−108.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change and its relation to the understanding of false belief and the appearance-reality distinction. *Child Development*, *59*(1), 26−37.

Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, *7*, 145−171.

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

Grosse Wiesmann, C., Friederici, A. D., Singer, T., & Steinbeis, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental Science*, *20*(5), e12445.

Halder, T., Sharma, K., Ghosh, S., & Verbrugge, R. (2015). How do adults reason about their opponent? Typologies of players in a turn-taking game. *Proceedings of the Annual Meeting of the Cognitive Science Society, 37*, 854−859.

Hallajian, A.-H., Dehghani-Arani, F., Sima, S., Heydari, A., Sharifi, K., Rahmati, Y., … Salehinejad, M. A. (2026). Theta-burst stimulation of the right temporoparietal junction and implicit theory of mind in autism. Neuromodulation: Technology at the Neural Interface, 29(2), 306−315.

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

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. *Proceedings of the National Academy of Sciences*, *105*(18), 6741−6746.

Happé, F. G. (1994). An advanced test of theory of mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*(2), 129−154.

Hare, B., Call, J., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771−785.

Hedden, T., & Zhang, J. (2002). What do you think I think you think? Strategic reasoning in matrix games. *Cognition*, *85*(1), 1−36.

Jacob, P. (2019). Challenging the two-systems model of mindreading. In A. Avramides & M. Parrott (Eds.), Knowing Other Minds (pp. 79−106). Oxford University Press.

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

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kano, F., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2019). Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proceedings of the National Academy of Sciences*, *116*(42), 20904−20909.

Kayhan, E., Nguyen, T., Matthes, D., Langeloh, M., Michel, C., Jiang, J., & Hoehl, S. (2022). Interpersonal neural synchrony when predicting others' actions during a game of rock-paper-scissors. *Scientific Reports*, *12*(1), 12967.

Kliemann, D., & Adolphs, R. (2018). The social neuroscience of mentalizing: Challenges and recommendations. *Current Opinion in Psychology*, *24*, 1−6.

Kinderman, P., Dunbar, R., & Bentall, R. P. (1998). Theory-of-mind deficits and causal attributions. *British Journal of Psychology*, *89*(2), 191−204.

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.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others' beliefs in human infants and adults. *Science*, *330*(6012), 1830−1834.

Krockow, E. M., Colman, A. M., & Pulford, B. D. (2016). Exploring cooperation and competition in the Centipede game through verbal protocol analysis. *European Journal*

of Social Psychology, *46*(6), 746−761.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110−114.

Kruppa, J. A., Reindl, V., Gerloff, C., Oberwelland Weiss, E., Prinz, J., Herpertz-Dahlmann, B., ... Schulte-Rüther, M. (2021). Brain and motor synchrony in children and adolescents with ASD—A fNIRS hyperscanning study. *Social Cognitive and Affective Neuroscience*, *16*(1-2), 103−116.

Lei, Y., Liu, S., Guo, X., Zuo, B., & Wen, F. (2025). Neural synchronization and its impact on intergroup attitudes in dynamic interactions. *Communications Biology*, *8*(1), 312.

Leslie, A. M., Friedman, O., & German, T. P. (2004). Core mechanisms in 'theory of mind'. *Trends in Cognitive Sciences*, *8*(12), 528−533.

Li, J., Liu, X. P., & Tang, W. H. (2014). Flexibility of theory of mind in a matrix game when the opponent's level changes. *The American Journal of Psychology*, *127*(1), 75−85.

Li, X., Ding, Y., Jiang, Y., Zhao, Y., Xie, R., Xu, S., … Xu, B. (2025). *DPMT: Dual process multi-scale theory of mind framework for real-time human-AI collaboration*. arXiv. https://doi.org/10.48550/arXiv.2507.14088

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1.

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.

Lu, H., Niu, J., Wang, J., Liu, M., & Xu, M. (2024). Characterization of implicit and explicit mind-reading in children with autism based on eye movements. *Frontiers in Psychiatry*, *15*, 1449995.

Marshall, J., Gollwitzer, A., & Santos, L. R. (2018). Does altercentric interference rely on mentalizing? Results from two level-1 perspective-taking tasks. *PLOS ONE*, *13*(3), e0194101.

Ma, Y., & Tan, H. (2023). Representation of intergroup conflict in the human brain. *Neuron*, *111*(10), 1692−1696.

Markiewicz, R., Rahman, F., Apperly, I., Mazaheri, A., & Segaert, K. (2024). It is not all about you: Communicative cooperation is determined by your partner's theory of mind abilities as well as your own. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(5), 833−844.

Meyer, M. L., & Collier, E. (2020). Theory of minds: Managing mental state inferences in working memory is associated with the dorsomedial subsystem of the default network and social integration. *Social Cognitive and Affective Neuroscience*, *15*(1), 63−73.

Michael, J., Wolf, T., Letesson, C., Butterfill, S., Skewes, J., & Hohwy, J. (2018). Seeing it both ways: Using a double-cuing task to investigate the role of spatial cuing in Level-1 visual perspective-taking. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(5), 693−702.

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.

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.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review*, *85*(5), 1313−1326.

Navarro, E., DeLuca, V., & Rossi, E. (2022). It takes a village: Using network science to identify the effect of individual differences in bilingual experience for theory of mind. *Brain Sciences*, *12*(4), 487.

Neo, W. S., Yu, M., Weber, R. A., & Gonzalez, C. (2013). The effects of time delay in reciprocity games. *Journal of Economic Psychology*, *34*, 20−35.

Newman, L. A., Cao, M., Täuber, S., & van Vugt, M. (2024). Working memory load impairs tacit coordination but not inter-brain EEG synchronization. *Social Cognitive and Affective Neuroscience*, *19*(1), nsae017.

Nicholson, T., Williams, D. M., Grainger, C., Lind, S. E., & Carruthers, P. (2019). Relationships between implicit and explicit uncertainty monitoring and mindreading: Evidence from autism spectrum disorder. *Consciousness and Cognition*, *70*, 11−24.

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.

O'Grady, C., Kliesch, C., Smith, K., & Scott-Phillips, T. C. (2015). The ease and extent of recursive mindreading, across implicit and explicit tasks. *Evolution and Human Behavior*, *36*(4), 313−322. https://doi.org/10.1016/j.evolhumbehav.2015.01.004

Oktay-Gür, N., Schulz, A., & Rakoczy, H. (2018). Children exhibit different performance patterns in explicit and implicit theory of mind tasks. *Cognition*, *173*, 60−74.

Osawa, H., & Imai, M. (2013). Evolution of mutual trust protocol in human-based multi-agent simulation. In P. Liò, O. Miglino, G. Nicosia, S. Nolfi, & M. Pavone (Eds.), Advances in artificial life, ECAL 2013: Proceedings of the twelfth European conference on the synthesis and simulation of living systems (pp. 692−697). Cambridge, MA: MIT Press.

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.

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.

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.

Perner, J. (1991). *Understanding the representational mind*. MIT Press.

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.

Polyanskaya, I., Eigsti, I. M., Brauner, T., & Blackburn, P. (2022). Second-order false beliefs and linguistic recursion in autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *52*(9), 3991−4006.

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

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.

Qureshi, A. W., Monk, R. L., Samson, D., & Apperly, I. A. (2020). Does interference between self and other perspectives in theory of mind tasks reflect a common underlying process? Evidence from individual differences in theory of mind and inhibitory control. *Psychonomic Bulletin & Review*, *27*(1), 178−190.

Rakoczy, H. (2022). Foundations of theory of mind and its development in early childhood. *Nature Reviews Psychology*, *1*(4), 223−235.

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495−505.

Réveillé, C., Vergotte, G., Perrey, S., & Bosselut, G. (2024). Using interbrain synchrony to study teamwork: A systematic review and meta-analysis. *Neuroscience & Biobehavioral Reviews*, *159*, 105593.

Ridinger, G., & McBride, M. (2025). Theory-of-mind ability and cooperation. *Journal of Economic Behavior & Organization*, *230*, 106895.

Rothmaler, K., & Grosse Wiesmann, C. (2023). Evidence against implicit belief processing in a blindfold task. *PLOS ONE*, *18*(11), e0294136.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. Journal of Experimental Psychology: Human Perception and Performance, 36(5), 1255−1266.

Santos, L. R., Nissen, A. G., & Ferrugia, J. A. (2006). Rhesus monkeys, Macaca mulatta, know what others can and cannot hear. *Animal Behaviour*, *71*(5), 1175−1181.

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

Schüler, C., Berger, P., & Grosse Wiesmann, C. (2026). A dorsal versus ventral network for understanding others in the developing brain. *The Journal of Neuroscience*, *46*(17), e1419242026.

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience & Biobehavioral Reviews*, *42*(1), 9−34.

Senju, A., Southgate, V., White, S., & Frith, U. (2009). Mindblind eyes: An absence of spontaneous theory of mind in Asperger syndrome. Proceedings of the National Academy of Sciences, 106(9), 3353−3357.

Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, *69*(1), 99−118.

Simon, H. A. (1976). From substantive to procedural rationality. In T. J. Kastelein, S. K. Kuipers, W. A. Nijenhuis, & G. R. Wagenaar (Eds.), 25 years of economic theory: Retrospect and prospect (pp. 65−86). Springer.

Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge University Press.

Sodian, B., & Kristen-Antonow, S. (2015). Declarative joint attention as a foundation of theory of mind. *Developmental Psychology*, *51*(9), 1190−1200.

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.

Stanley, G. N. (2026). Strategic reasoning under pressure: Testing heuristics in higher-order theory of mind. *Cognition*, *266*, 106331.

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

Tang, H., Mai, X., Wang, S., Zhu, C., Krueger, F., & Liu, C. (2016). Interpersonal brain synchronization in the right temporo-parietal junction during face-to-face economic exchange. *Social Cognitive and Affective Neuroscience*, *11*(1), 23−32.

Thompson, K., Nahmias, E., Fani, N., Kvaran, T., Turner, J., & Tone, E. (2021). The Prisoner's Dilemma paradigm provides a neurobiological framework for the social decision cascade. *PLOS ONE*, *16*(3), e0248006.

Tian, M., Luo, T., Ding, J., Wang, X., & Cheung, H. (2021). Spatial ability and theory of mind: A mediating role of visual perspective taking. *Child Development*, *92*(4), 1590−1604.

Todd, A. R., Simpson, A. J., & Cameron, C. D. (2019). Time pressure disrupts level-2, but not level-1, visual perspective calculation: A process-dissociation analysis. *Cognition*, *189*, 41−54.

Todd, P. M., & Gigerenzer, G. (2012). Ecological rationality: Intelligence in the world. Oxford University Press.

Valle, A., Massaro, D., Castelli, I., & Marchetti, A. (2015). Theory of mind development in adolescence and early adulthood: The growing complexity of recursive thinking ability. *Europe's Journal of Psychology*, *11*(1), 112−124.

Van der Linden, D., Frese, M., & Meijman, T. F. (2003). Mental fatigue and the control of cognitive processes: Effects on perseveration and planning. *Acta Psychologica*, *113*(1), 45−65.

Van Overwalle, F., & Vandekerckhove, M. (2013). Implicit and explicit social mentalizing: Dual processes driven by a shared neural network. *Frontiers in Human Neuroscience*, *7*, 560.

von der Osten, F. B., Kirley, M., & Miller, T. (2017, August). The minds of many: Opponent modeling in a stochastic game. In C. Sierra (Ed.), Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (pp. 3845−3851). Melbourne, Australia: International Joint Conferences on Artificial Intelligence.

Von Stackelberg, H. (2010). *Market structure and equilibrium*. Springer Science & Business Media.

Wang, Z., Wang, Y., Zhou, X., & Yu, R. (2020). Interpersonal brain synchronization under bluffing in strategic games. *Social Cognitive and Affective Neuroscience*, *15*(12), 1315−1324.

Wang, Z., Xu, B., & Zhou, H. J. (2014). Social cycling and conditional responses in the Rock-Paper-Scissors game. *Scientific Reports*, *4*(1), 5830.

Wei, Z. H., Li, Q. Y., Liang, C. J., & Liu, H. Z. (2022). Cognitive process underlying ultimatum game: An eye-tracking study from a dual-system perspective. *Frontiers in Psychology*, *13*, 937366.

Weimer, A. A., Cortez, N., & Razo, N. (2022). Does chess-playing relate to theory of mind? An examination of the interrelations among theory of mind, perspective-taking, and empathic concern in chess-players (¿ El juego de ajedrez tiene relación con la teoría de la mente? Un análisis de las interrelaciones entre la teoría de la mente, la toma de perspectiva, y el interés empático de los jugadores de ajedrez). *Studies in Psychology*, *43*(2), 389−413.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655−684.

Wilson, R., Hruby, A., Perez-Zapata, D., van der Kleij, S. W., & Apperly, I. A. (2023). Is recursive “mindreading” Really an exception to limitations on recursive thinking? *Journal of Experimental Psychology: General*, *152*(5), 1454−1468.

Xia, M., Wang, J., & He, Y. (2013). BrainNet Viewer: A network visualization tool for human brain connectomics. PLOS ONE, 8(7), e68910.

Yang, R., Ma, Y., Pan, B. B., Bhatt, M. A., Lohrenz, T., Gu, H. G., ... Luo, Q. (2022). Dynamic neural reconfiguration for distinct strategies during competitive social interactions. *NeuroImage*, *263*, 119585.

Yoshida, W., Dolan, R. J., & Friston, K. J. (2008). Game theory of mind. *PLOS Computational Biology*, *4*(12), e1000254.

Yousefi, Z., Heinke, D., Apperly, I., & Siebers, P. O. (2020, October). An agent-based model to understand a simple theory of mind: Belief representation systematic approach (BRSA). In *Conference of the Computational Social Science Society of the Americas* (pp. 355−380). Cham: Springer International Publishing.

Zhang, M., Jia, H., & Wang, G. (2021). Interbrain synchrony of team collaborative decision-making: An fNIRS hyperscanning study. *Frontiers in Human Neuroscience*, *15*, 702959.

Zhang, Y., Ye, W., Yin, J., Wu, Q., Huang, Y., Hao, N., ... Cai, D. (2024). Exploring the role of mutual prediction in inter-brain synchronization during competitive interactions: An fNIRS hyperscanning investigation. *Cerebral Cortex*, *34*(1), bhad483.

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.

Zheltyakova, M., Kireev, M., Knyazeva, I., Myznikov, A., Kiselev, V., Didur, M., ... Korotkov, A. (2026). Task-dependent increases and decreases of BOLD signal in theory of mind brain regions during strategic social interaction. *Frontiers in Neural Circuits*, *20*, 1741762.

Zhen, S., & Yu, R. (2021). Neural correlates of recursive thinking during interpersonal strategic interactions. *Human Brain Mapping*, *42*(7), 2128−2146.

Resource-adaptive theory-of-mind reasoning: Integrating dual-process mechanisms and recursive hierarchies from the perspective of bounded rationality

ZHONG Yongdan, XIAO Feng

(*School of Psychology, Key Laboratory of Brain Function and Brain Disease Prevention and Treatment of Guizhou Province, Guizhou Normal University, Guiyang* 550025, *China*)

Abstract: The processing mode and recursive depth of theory of mind (ToM) address, respectively, how social reasoning is carried out and how far it extends. However, a unified account of how these two dimensions jointly vary under cognitive resource constraints remains lacking. Drawing on bounded rationality, this article proposes an integrative perspective of "resource-adaptive ToM reasoning," which conceptualizes processing mode and recursive depth as two interrelated and probabilistically coupled dimensions under cognitive resource constraints. It posits that individuals adjust their processing investment and level of recursive reasoning according to task demands, available resources, cognitive costs, and expected benefits. This perspective yields three testable predictions: (1) the degree of System 2 engagement and the sustainable depth of recursive reasoning are jointly constrained by task demands, available resources, and expected benefits, and the pathways linking these factors can be examined through experimental manipulations and model comparisons; (2) when recursive computation is costly and opponents are limited in complexity, recursive depth has an inverted-U-shaped relationship with decision benefits, with the optimal depth varying across environmental structures; and (3) in interactions characterized by continuous feedback and some degree of strategic predictability, the interaction partners' resource-regulation strategies will tend toward coordinated alignment over time. This perspective provides a testable framework for distinguishing between capacity limits, strategic choices, and interpersonal coordination.

Keywords: bounded rationality, dynamic games, dual-process accounts of theory of mind, recursive theory of mind, interaction coordination