《心理科学进展》审稿意见与作者回应

题目：反应风格建模的深化：项目反应树的框架演进及应用
作者：董洁源 刘红云

第一轮

审稿人 1 意见：

本文聚焦项目反应树（IRTree）模型在反应风格（RS）建模中的框架演进，选题具有较强的方法学前沿性，也契合当前对问卷作答过程和反应风格控制精细化建模的研究趋势，论文具有一定的理论和实践价值。但论文还存在不少的问题，需要进一步修改。

主要问题：

意见 1：引言的问题意识不够突出。就作者所写的引言来说，作者似乎是以反应风格（RS）的建模为应用背景，对 IRTree 模型的发展脉络进行梳理，并顺带给出具体的应用示例。这更像是在介绍 IRTree 模型的发展史，而其问题意识则还没有充分建立。

回应：衷心感谢专家的宝贵意见。针对专家指出的引言部分问题意识不够突出的问题，我们对引言部分进行了重点修改。在概述项目反应树（IRTree）模型的应用现状后，突出本文要解决的核心问题，即 IRTree 模型变体日益丰富但不同模型的理论假设与适用场景尚缺乏系统梳理的问题。具体改动内容为在简述现状后增加了如下一段论述（第 2 页第 2 段）：

“然而，面对日益丰富的 IRTree 模型变体，一个基本问题尚未得到系统回答：不同模型各自隐含了怎样的理论假设，又各自适用于怎样的分析场景？已有的关于 IRTree 模型的教程或综述，或仅关注某一类拓展(Plieninger, 2021; Xue & Chen, 2023)，或在综述各种 RS 建模方法时概括性地论及 IRTree 模型(Bolt & Meng, 2025; 张缨斌, 王烨晖, 2019)；尤其在国内，相关综述由于时间较早，仅介绍了 IRTree 发展初期的基础模型(张缨斌, 王烨晖, 2019)。这一现状对研究者构成一定障碍：方法学研究者因缺乏对 IRTree 演进脉络的系统梳理，难以进一步改进和拓展模型；应用研究者因缺乏清晰的理论假设说明和实践指引，面对众多变体时容易无所适从。

为解决上述问题，本文系统梳理 IRTree 模型在 RS 建模领域的框架演进。……”

意见 2：文章明确了要做什么，但对于为什么要做，文章并没有提供强有力的依据。目前文章所写的引言对综述必要性的论证更多的侧重于 RS 研究领域的热度和 IRTree 建模方法的先进性，但对于为什么需要专门聚焦“IRTree 模型的框架演进”进行综述的学术理由，作者阐述的并不充分。

回应：衷心感谢专家的宝贵意见。我们进一步强化了综述必要性的论述。修改后的引言在提出核心问题之后，补充了如下内容（第 2 页第 3 段），以明确说明为何要专门聚焦“IRTree 模型的框架演进”：

“为解决上述问题，本文系统梳理 IRTree 模型在 RS 建模领域的框架演进。之所以聚焦于此，主要有如下考虑。第一，围绕 IRTree 开展的方法学研究持续积累，推动了 RS 研究从单纯的偏差控制转向兼顾过程理解，这一研究取向的转变值得阶段性回顾。第二，IRTree 模型的一系列拓展也并非零散的技术更新，而是围绕作答过程表征这一核心命题的建模精细

化，这构成了 RS 研究深化的另一重要层次。此外，当前国内对 IRTree 的应用仍十分有限，症结之一即在于缺乏兼顾理论深度与实践操作的中文指引。”

结合意见 1，我们在修改稿中进一步明确指出当前要解决的问题，即缺乏对 IRTree 模型变体的理论假设与适用场景的系统梳理。在此基础上，上述补充进一步论证了本综述的学术必要性：首先是 IRTree 的演进与 RS 研究进展紧密关联（推动研究取向转变与建模精细化），其次是能回应国内应用受限的现实需求。再次感谢专家的指导。

意见 3：对本文贡献的阐述不够清晰，没有明确说明相对于已有综述，这篇文章到底弥补了哪些内容，重点覆盖到了什么程度。

回应：非常感谢专家的意见。针对“对本文贡献的阐述不够清晰”的问题，我们已在引言中明确已有综述的不足（或仅关注某一类拓展，或仅概括性提及，国内仅介绍了发展初期的基础模型），并将引言末尾的贡献表述修改为（第 2 页最后一段）：

“为此，本文首次系统梳理 IRTree 从基础模型到前沿变体的演进历程，结合已有研究成果总结 IRTree 主要变体模型的统计结构、理论假设与适用边界，并提供递进式分析流程与实证演示，以服务研究者的理论理解与实践应用。”

这一表述更加清晰地说明了相对于已有综述，本文所弥补的内容（从局部或概括性介绍转向系统性框架梳理并增加实践指引），以及覆盖的程度（演进脉络、统计结构、理论假设、适用边界与分析演示）。

意见 4：在介绍 IRTree 模型时，作者尝试将其置于 MPT 和 GLMM 两个更广的建模框架下加以理解（1.2 章节），这一思路本身具有一定理论意义。但是，就目前的这种写法来说，这部分内容并没有显著提升读者对 IRTree 核心机制的理解，反而在一定程度上增加了理解负担。第一，IRTree 模型本身的基本思想、节点含义以及它和传统 IRT 模型的关键差异都还没有充分阐明的情况下，就进一步引入两个上位框架，容易使读者在还没有把握核心对象时就要承受额外的概念负荷；第二，作者并没有清楚地说明 MPT 与 GLMM 各自究竟帮助理解了 IRTree 模型的哪一方面，前者更接近作答过程的树状表示，后者更接近统计表达与估计实现，但文中未对这种差异进行明确区分。

回应：非常感谢专家的宝贵意见，也感谢专家对“将 IRTree 置于 MPT 和 GLMM 框架下理解”这一思路理论意义的认可。针对专家的意见，我们对 1.2 节做了以下重点修改。

首先，我们将原 1.3 节开头对 IRTree 模型的介绍合并至 1.2 节的开头，扩充了对 IRTree 基本思想、节点含义及其与传统 IRT 模型关键差异的阐述。修改后的 1.2 节第 1 段表述为：

“与传统的多级计分 IRT 模型如等级反应模型（graded response model. GRM）或广义分部评分模型（generalized partial credit model, GPCM）不同，IRTree 模型不假设所有类别差异由同一潜在特质 TOI 及其阈值决定，而是将多级反应分解为多个二分决策节点，每个节点可分配不同的被试参数。尽管信息被分解至不同的伪题目上，整套伪题目作为一个整体仍完整保留了原始有序作答中的全部信息。这一信息完整性由 IRTree 的似然分解结构直接保证，在树结构正确指定的前提下，各伪题目的条件独立似然乘积等于原始作答的似然函数(Li, 2022)。此外，IRTree 的树结构也为分离不同认知过程提供了基石，各节点参数（如难度）均具有与具体决策过程直接对应的直观解释(Khorramdel & Von Davier, 2014)。”

其次，我们精简了关于 MPT 与 GLMM 的论述篇幅，将其合为一段后移至 1.2 节末尾，仅简要提示 IRTree 的树状表示与 MPT 的认知决策树结构相通，而其统计表达与估计则可借助 GLMM 框架实现。这样既保留了对框架间联系的概述，又避免在读者尚未掌握核心对象时增加概念负荷，同时也按照审稿人的建议明确区分了 MPT 侧重“树状表示”、GLMM 侧重“统计表达与估计实现”的不同作用。再次感谢专家对于提升章节可读性的细致指导。

意见 5：在 1.3 章节，作者在总结 IRTree 在反应风格建模中的优势和理论特色时，存在论证偏结论性表述和证据支撑不足的问题。首先，关于 IRTree 模型能够“完整保留原有信息”“分离不同心理过程”以及“参数具备直观的解释性”等优势的阐述方面，文中主要停留于概念层面的陈述，没有进一步说明这些判断成立的条件、边界及其方法学依据。其次，在将 IRTree 模型和混合 IRT、混合因子模型以及 MNRM 等方法进行比较时，作者虽然指出了一些重要差异，但对于这些差异为何能够构成理论优势、又在何种情境下可能带来额外代价等内容，讨论并不充分。

回应：非常感谢专家的宝贵意见。我们针对专家指出的意见对 1.3 节进行了以下修改。

首先，我们将 1.3 节的开头调整到 1.2 节的开头作为 IRTree 树结构特征的阐述（上一条意见 4 的回复）。其中“分离不同心理过程”以及“参数具备直观的解释性”源于 IRTree 统计结构的天然特征(Khorramdel & Von Davier, 2014)，成立的前提是研究者能基于理论或经验提出合理的节点分解方案。至于“完整保留信息”这一特征也以树结构合理指定为前提，此时各伪题目的条件独立似然乘积等于原始作答的似然函数，Li(2022)提供了这一推导，我们已在修改稿中补充了该文献：

Li, Z. (2022). *Filling gaps in the research on IRTree theoretically and practically: A comprehensive taxonomy and a user-friendly R package* [Unpublished doctoral dissertation]. The Ohio State University.

其次，关于方法比较，本文的本意并非全面评判孰优孰劣，而是将 IRTree 置于 RS 建模方法谱系中，帮助读者理解其独特假设。初稿的表述确实容易误导读者，在修改稿中我们删去了小标题中的“优势”，在正文仅指出关键方法差异：潜类别方法与潜特质方法对 RS 的概念化不同；IRTree 与 MNRM 在 RS 是否全局影响类别选择概率上存在区别。同时，尽管可以指出文献中更倾向于潜特质方法，但目前对潜特质方法下两类模型（IRTree 与 MNRM）尚无共识性结论。我们在 1.3 节末尾补充了如下讨论，本文后续仍聚焦 IRTree 自身框架演进，不展开方法间比较：

“目前两类模型孰优孰劣尚无共识性结论，二者均为主流 RS 建模方法。考虑到两者同属验证性框架，研究者在实践中需重视理论建构，并依据理论假设选择合适的方法：当缺乏明确的节点分解理论或者研究者不关心作答过程成分时，MNRM 的一步全局建模可能更为简便；当理论明确指向 RS 影响特定决策阶段时，IRTree 的结构化分解更为匹配。”

意见 6：在阐述“2.1 从节点内单维到节点内多维”章节内容时，存在几个值得进一步强化论证的地方：（1）“节点间多维”和“节点内多维”属于不同层面的概念，文中对二者关系的阐释尚不够清晰，导致从基础 IRTree 模型过渡到节点内多维模型时显得比较跳跃。（2）在阐述节点内多维 IRTree 模型的优势时，作者关于“节点内多维更贴近现实”“能够消除模仿效应”“提供无偏估计”等的判断，主要依据模型拟合指标、仿真结果和个别应用研究加以支撑，但这些证据尚不足以直接推出如此强的理论和方法学结论。（3），文中对节点内多维模型的优势阐述较多，但对其复杂度提升、识别约束、估计稳定性及实际应用门槛等代价的讨论则不足，容易造成论证上的失衡，即会让人觉得节点内多维在所有条件下都要优于单维。

回应：非常感谢专家的宝贵意见。针对 2.1 节中“节点间多维”与“节点内多维”关系阐释不清、优势论证过强且缺乏对代价的讨论等问题，我们进行了以下系统修改。

首先，为厘清“节点间多维”与“节点内多维”的概念关系，我们增加了更多铺陈论述，尽可能让从“节点间多维”到“节点内多维”的过渡逻辑更清晰。在 1.1 节末尾明确点出“节点间多维”结构（“这里的‘三维’指每道原始题目所分解出的三个节点伪题目分别对应于 MRS、TOI 和 ERS 三个维度，从而构成节点间的多维结构”），并在第 2 章节的引入段指出

基础 IRTree 模型在每个节点内部仍维持“节点内单维”假设（“即每个节点的反应仅由单一潜变量驱动”）。进而在 2.1 节开头概述为“基础 IRTree 模型虽整体呈现‘节点间多维’结构，但在每个节点内部仍维持‘节点内单维’假设”，然后从理论出发过渡到节点内多维的必要性：“为刻画这种类似双因子（Bifactor）模型的‘TOI 全局、RS 局部’假设，TOI 负载于所有伪题目，在原本仅 RS 作用的节点形成‘TOI+RS’的补偿性多维结构”（具体见 2.1 小节第 1 段）。

其次，关于节点内多维模型优势的论证，我们弱化了结论性表述，例如改为“其优势也已得到了初步验证”，并补充提醒读者“上述优势证据来自有限的模拟研究和应用案例，其普遍性和稳健性仍有待更多研究检验”，以避免过度概括导致结论过强。

最后，为使论证更加平衡，我们专门增加了对节点内多维模型局限与挑战的讨论，从“节点内多维结构提升了 IRTree 模型的复杂度，带来模型识别与估计稳定性方面的挑战”起，用两段文字展开论述其识别约束、估计稳定性等问题。同时，在总结段（第 10 页最后一段）给出了更加平衡的应用建议：

“综上所述，尽管节点内多维结构增加了模型复杂度，其在纠正偏差、分离特质、降低 TOI 估计标准误方面展现出比较明确的优势。概括而言，该结构尤其适用于以下场景：研究者有明确的理论预期，认为 TOI 对作答过程具有全局影响（而不仅限于方向决策）；数据可能存在反应分布不对称（如临床样本、高利害选拔）；研究目的侧重于准确分离 RS 与 TOI，而非追求模型简约性。在样本量较小或理论假设尚不清晰时，节点内单维模型仍是更稳妥的选择。”

意见 7：在对“2.2 从优势过程到理想点过程”章节内容进行阐述时，部分内容的论证链条存在不足。具体来说，作者没有充分说明为什么要构建理想点过程的 IRTree 模型。现有论述更多是从“理想点模型在部分非认知测量中可能更合适”直接推进到“有必要将其纳入 IRTree 模型”，中间缺少针对节点层面建模需求的具体论证。其次，缺少关于理想点和优势过程共存模型的使用边界的相差阐述。

回应：衷心感谢专家的宝贵意见。我们针对专家指出的意见对 2.2 和 2.3 节做了如下修改。

首先，为充分说明为什么要构建理想点过程的 IRTree 模型，我们在综述文献时加强了对各研究理论建构的表述。具体如下（第 12 页第 2 段）：

“Jin 等(2022)率先提出了理想点-优势两阶段树模型（IDtree），其核心论证是在方向决策上，被试的反应机制应遵循理想点过程而非优势过程。他们指出，对于社会不平等态度之类的构念，个体只有在自身构念水平与题目描述的位置接近时才会‘同意’，过高或过低均会‘不同意’，这恰是理想点过程的单峰特征。……Li 等(2025)提出更一般的展开 IRTree（UIRTree）模型，并将节点层面的论证表述为：IRTree 中所有受 TOI 影响的节点，均应考虑采用理想点过程，而明确受 ERS、MRS 等影响的节点保留优势过程。……Zeng 等人(2025)的研究进一步探索了新的理论假设，认为方向决策与强度决策应由两个不同的理想点 TOI 驱动……”

同时，我们在 2.2 节总结段做了回顾与补充（第 12 页最后一段）：“……研究者对理想点过程在 IRTree 中应用的认识不断发展，凸显了在节点层面审视反应函数设定的必要性。从理论假设看，纳入理想点过程的核心前提是：个体在作答时是‘匹配判断’而非‘水平判断’，其反应取决于自身 TOI 与题目位置间的距离而非 TOI 的绝对高低。该假设在测量态度、人格等典型行为时通常比优势假设更合理。在建模前可先基于理论确定 TOI 在方向决策上的反应函数，……”以此强化论证链条。

其次，针对理想点与优势过程共存模型的使用边界问题，我们在 2.3 节总结段补充了相关阐述：“然而，目前对这些结构的研究依然十分有限。对于 5 点量表，尚无系统的模拟研

究评估如公式(19)-(22)所示的复杂模型在有限样本量下的参数返真性及与现有简化模型的比较。而对于方向决策已采用理想点过程的 5 点量表，如何在折中反应节点引入节点内多维结构仍是方法学研究的空白。”从而明确了当前研究的局限以及模型的应用边界。

意见 8：文章将“节点内单维到节点内多维”和“优势过程到理想点过程”归入“框架演进 I”，而将作答过程异质性的纳入归入“框架演进 II”，这种组织方式从逻辑层级看不够整齐。前两者分别涉及节点内部潜变量结构的扩展和节点反应函数表达形式的扩展，而“异质性纳入”则对应于更高层面的作答过程变异建模，三者并不处于同一个分类维度。当前的写法容易给人一种分类标准在不同部分之间切换的感觉。

回应：非常感谢专家的意见。原写作意图并非将第 2 章节中的两个扩展与第 3 章节的过程层面异质性并列为三个平行部分，但客观上确实造成了分类维度不统一的问题。为避免误解，我们将这两章的标题调整为“节点层面的框架演进：潜变量结构与反应函数的拓展”和“过程层面的框架演进：作答过程异质性的建模”，从而突出潜变量结构和反应函数的拓展均从属于节点层面的拓展，而节点层面的拓展与过程层面的拓展才是两条主线的综述逻辑。同时，我们也在第 3 章节引入段开头增加了“本章将视野从节点内部抬升到整个作答过程，聚焦于更宏观的问题”，以体现两条主线的递进关系。再次感谢您的指导。

意见 9：在第 3 章节中，作者将混合 IRTree 模型、路径不确定性 IRTree 模型和动态 IRTree 模型统一归入“作答过程异质性的纳入”的论证依据不足。3.1 到 3.3 三个不同子节所讨论的东西并不是同一层面的“异质性”，作者没有充分说明它们为何能够被视作同一条演进主线；而各部分之间的递进关系是什么，作者也没有说清楚。

回应：非常感谢专家的意见。3.1 至 3.3 节讨论的三类异质性确实处于不同层面，但构成了一条逐层递进的演进主线，且均是对“同质作答过程”假设（所有被试遵循相同的决策路径）的突破。其中，个体间异质性（3.1 节）关注不同被试间反应策略的稳定差异；作答内异质性（3.2 节）进一步允许同一观测作答源于多种心理路径，突破了“一种结果只能对应一条路径”的假设；个体内动态异质性（3.3 节）则将视角延伸至时间维度，捕捉反应策略随任务进程或测量时点的变化。由于对动态变化的建模直接挑战了 RS 跨时间稳定的传统假设，故置于最后，作为前两层静态分析（稳定的个体差异、单次作答的路径混合）的延伸与补充。三类异质性的递进不仅体现了模型复杂度的逐步提升，更是研究逻辑从“不同人有策略差异”到“一种作答背后可能由不同策略产生”再到“策略如何随时间进程变化”的层层深入。

为清晰呈现这一递进，我们在第 3 章节的引入段以工作满意度调查为例，展示了三类异质性所对应的典型研究问题，帮助读者直观认识“三层异质性逐层深入”。同时，3.2 节开头强调了与个体间异质性的承接关系（“此类单结果多路径模型在作答层面刻画了过程异质性，即同一观测作答可能源于不同的心理决策过程，从而将异质性的研究从稳定的个体差异延伸至单次作答的路径不确定性”），3.3 节开头则点明了与前面两层的推进关系（“在系统探讨了认知过程在个体间的稳定差异以及在单次作答中的不确定性之后，研究议题进一步转向 RS 是否会在个体内随着任务进程或者时间推移而发生系统性变化。IRTree 模型的应用也从识别相对静态的群体或个体模式，逐步扩展到捕捉个体自身反应策略的动态异质性”）。

意见 10：文章主要阐述了传统 IRTree 模型的三个拓展或演进（节点内单维到多维、优势过程到理想点过程、作答过程异质性引入），但对于三个拓展或演进之间的内在逻辑关系的阐述不清晰和不具体，给人的感觉是三个并列的拓展。

回应：非常感谢专家的意见。正如在意见 8 的回复中所述，三者并非并列关系，而是存在清晰的层次与路径：从节点内单维模型到引入节点内多维结构、从所有节点为优势过程到引入

理想点过程节点，分别对应节点层面“潜变量结构”和“反应函数形式”两个不同的建模精细化方向，二者可独立也可结合使用；而作答过程异质性的引入则将分析单元从节点内部提升至整个作答路径，是在前两者所奠定的节点模型基础上，进一步考虑决策路径本身的变异。我们已通过调整章节标题、增加第 3 章节引入段的衔接文字（见意见 8 的回应）阐述了两条主线的内在逻辑；至于第 2 章节内部的两个不同层面的拓展，我们也在第 2 章节的引入段交代了两种拓展的并行关系，以及二者可以结合使用，以帮助读者更好地认识到第 2 章节的综述逻辑：

“基础 IRTree 在节点层面隐含了两个关键假设：一是节点内单维结构，即每个节点的反应仅由单一潜变量驱动；二是反应函数均为优势过程，即假定每个节点的反应概率均随对应潜变量单调变化。……本章节即从这两个方向梳理节点层面的拓展：首先讨论节点内多维结构的纳入，允许一个节点同时受多个被试参数影响；其次讨论理想点过程的纳入，允许节点采用非单调反应函数；最后讨论两类拓展的结合使用。”

意见 11：从文章对 IRTree 模型及拓展模型的描述来看，文章更多的只是阐述了模型发展历史，对于各拓展模型解决了什么问题，又带来了什么代价（如拓展模型适用于什么样的数据和情境，对样本量、模型识别、估计稳定性等内容有什么样的要求，相较于传统 IRTree 模型的解释优势又是什么）等内容的阐述非常少，使得文章看起来像是综而不述。

回应：衷心感谢专家的宝贵意见。我们在保留模型演进脉络的基础上，基于已有文献可获取的信息，系统补充了各模型适用情境、样本量需求、识别约束与估计稳定性等方面的评述。

具体而言，在 2.1 节末尾增加了节点内多维结构的适用场景（如 TOI 全局影响、反应分布不对称、侧重分离 RS 与 TOI）及样本量较小时的替代建议，并讨论了参数约束与估计稳定性挑战；在 2.2 节末尾补充了理想点过程的理论前提（匹配判断）与操作建议（理论不充分时同时拟合竞争模型）；在 2.3 节末尾明确指出当前研究的局限（5 点量表缺乏参数返真性研究、缺乏模型选择准则与软件实现指导）；3.1 节以表格形式系统比较了各种混合 IRTree 模型的分析目标、模拟研究表现、主要贡献等，并补充了样本量局限与类别数确定建议；3.2 节明确了纳入 ARS 的单结果多路径模型需足够数量正向题与反向题的局限；3.3 节补充了动态模型研究尚处起步阶段、缺乏系统模拟评估的评述，并区分了单次测量动态性与纵向动态性的不同假设。此外，在第 4 章节我们新增了 4.1 节“模型回顾”，以表格形式对综述涵盖的主要模型变体从理论基础、结构特征与适用场景角度进行了系统总结对比。

意见 12：“4.1 流程建议”是本文非常重要的内容，审稿人认为该流程存在若干值得商榷的地方。首先，阶段一被界定为“确定是否存在 RS 以及 TOI 的反应机制”，这一表述偏强，因为普通的 IRT 和单维 IRTree 模型之间的比较至多能提供是否有必要采用过程分解建模的初步证据，并不足以严格判定 RS 是否存在。其次，M0 到 M1 到 M2 的推进逻辑偏线性化，并且一些关键设定（如先在优势过程下调试方向不变性约束、在 M2 中沿用 M1 的 TOI 反应过程假设）更多体现为实际操作上的简化策略，而不是已经被充分论证的一般规则。再者，对于 5 点计分数据，作者提出先扩展极端反应节点、再考虑折中节点的安排，这个阐述同样比较依赖经验判断，建议进一步说明其理论依据和适用条件。最后，异质性建模放在了最后，并且作为“非必须步骤”，这一安排可能弱化其在模型诊断和结构选择中的实际重要性。

回应：衷心感谢专家指出值得商榷之处。我们已据此对 4.2 节“流程建议”做了充分修改。

首先，通过比较普通 IRT 模型和节点内单维 IRTree 模型来判断 RS 是否存在确实是已有文献中的做法(Khorramdel et al., 2019)，我们也同意审稿人对表述偏强的担忧，已将相关表述修改得更为保守。例如，在 4.2.1 节中，阶段一的核心目标调整为“判断是否有必要采用

IRTree 建模”，并说明“参考 Khorramdel 等(2019)的做法，通过比较 M_1 与 M_0 的拟合表现判断 IRTree 模型的过程分解建模是否合理：若 M_1 优于 M_0，则可认为数据中存在 IRTree 可捕捉的 RS。”又如 4.3.2 节的实证分析部分，相应修改为：“在考察的所有信息准则上 M_1 均优于 M_0，表明 IRTree 模型的过程分解建模更为合理。”

其次，我们认同“M_0 到 M_1 到 M_2 的推进逻辑偏线性化”的问题，已删去原线性化的流程图，避免让读者认为当中每一小步都需要严格循序执行。目前仅保留三个阶段的流程建议（增设 4.2.1 至 4.2.3 三个小节对应三个阶段），并且在 4.2 节引入段强调这是“递进式的一般流程”，从 M_0 到 M_1 到 M_2 更多体现从简单到复杂的一般建模思路，而 4.2 节侧重在这个简明的三阶段流程内提出建模注意事项。

至于当中的一些设定，关于“先在优势过程下调试方向不变性约束”和“在 M_2 中沿用 M_1 的 TOI 反应过程假设”这两条简化策略的确缺少充分论证。不过，由于修改后 4.2 节的流程中 TOI 反应函数（优势过程 vs 理想点过程）的确定更加突出理论优先，最迟在拟合 IRT 模型选出 M_0 时就应明确，并贯穿整个分析，因此已不需要在后续拟合 IRTree 模型阶段专门讨论 TOI 反应过程，也就不再需要向读者呈现这些简化建议。

而对于 5 点计分数据“先扩展极端反应节点、再考虑折中节点”的安排，主要考虑到极端反应节点内的多维结构已有较充分的模拟与实证支持，相对成熟；而折中反应节点的复杂结构仍属前沿，方法学验证尚不充分（这点已在 2.3 节说明），因此优先尝试前者符合从简到繁的建模逻辑，且文中已向读者说明仅在 M_2 脱颖而出后再考虑进一步的 M_3，因为如果拟合了 M_2 发现相比 M_1 没有实质性的理论与拟合收益，也就没有必要拟合 M_3。

最后，对于阶段三的异质性建模，我们不再将其概括为“非必要步骤”，而是改用更平衡更充分的论述：“在确定了最佳同质模型后，研究者可进一步探索作答过程异质性。……需注意的是，此类异质性建模方法往往面临参数众多、估计困难的挑战，对样本量、题量和模型设定要求较高。建议仅当具备明确的理论动机（如预期存在‘无 RS 作答者’亚群体）且样本量充足（通常需 1000 人以上）时再谨慎尝试。同时，应始终以阶段二确立的同质模型为基准，通过比较异质性模型是否带来实质性的理论增益与拟合改善来评估其必要性。”这样既不会弱化异质性建模在建议流程中的重要性，也指出了其适用条件，并且也强调应基于最佳同质模型进行探索，避免不必要的模型复杂化。此外，出于篇幅考虑，我们无法在 4.3 节再给出对阶段三的详细分析示例；但为了更完整地呼应 4.2 节的三阶段流程，我们在分析代码文件中额外提供了建立其中一种异质性模型（单结果多路径 IRTree 模型）的示例语句，供感兴趣的读者参考（见 4.3.1 节第 2 段的脚注）。

意见 13：实证数据的应用方面，文章直接说使用大五人格问卷中宜人性分量表的数据，但为何选这个数据，这个数据是否具有代表性，是否能够较好地展示反应风格问题等，文章并没有给出科学说明。

回应：非常感谢专家的意见。我们已在正文中补充了对数据选取的科学说明，明确指出：“作为成熟的人格测量工具，大五人格问卷已被广泛用于 RS 研究，例如 Khorramdel 和 Von Davier (2014)曾分析过该问卷作答中的 ERS 和 MRS。……为便于演示 IRTree 分析流程，选取宜人性分量表的 10 题，作为典型的 5 点计分、单构念问卷数据分析情境”。

意见 14：文章在方法流程建议中指出，在进行模型选择时，不仅需要看拟合指标，还需兼顾理论合理性，但在应用案例的展示方面，却只是依据拟合指标进行模型选择，而理论审视则被完全忽略。

回应：非常感谢专家的意见。我们首先在 4.2 节流程建议中更加明确了不同步骤的判断依据：在阶段一，判断是否以及确定参数约束形式时，主要依赖模型拟合比较，而 TOI 反应过程

（优势 vs. 理想点）的确定则必须坚持理论优先，只有理论不够充分时才同时拟合两种反应过程模型；在阶段二，M_1 与 M_2 的比较、M_2 与 M_3 的比较，都需要综合理论预期、拟合指标与参数可解释性来判断。

在应用示例（4.3 节）中，我们更加详细地阐述了每一步模型选择的依据：

（1）TOI 反应过程的确定（IRT 模型阶段）：虽然 10 道题目表述方向明确，理论上支持优势过程，但考虑到人格测验中理想点过程也很常用，故同时拟合 GPCM 与 GGUM。最终未仅依据拟合指标（GGUM 虽然 AIC 略优，但其他信息准则（BIC、SABIC）均支持 GPCM），更重要的是 GGUM 对第 3 题的位置参数估计极不稳定（点估计与 95%置信区间为 6.45[-1.81, 14.71]），而 GPCM 无此问题。因此，综合理论、拟合指标与参数估计稳定性，选择了更简洁的 GPCM 作为 M_0。

（2）M_1 与 M_2 的比较：理论预期 TOI 也影响极端反应节点，而 M_2 成功收敛，且宜人性特质在极端反应节点上的斜率估计值方向符合理论预期，再结合 M_2 同时所有信息准则均优于 M_1，故选择 M_2。

（3）M_2 与 M_3 的比较：理论预期 TOI 也影响折中反应节点，从 M_2 到 M_3，信息准则进一步下降，同时 M_3 参数估计没有明显问题，故选择 M_3 作为最终模型。

意见 15：在应用示例中，对结果进行解释时，对于很多的论断，作者只是基于数据结果进行解释，审稿人觉得仅基于数据进行解释还不够，至少需要结合数据和理论进行解释。就如对图 6 结果的解释来说，文章指出“…当节点 3 和节点 4 由单维改为多维结构时，MRS 的斜率估计基本稳定，而 ERS 在部分题目（例如题目 1、4、7）上的斜率明显下降，表明有部分变异实际由宜人性解释”，斜率变得稳定或明显下降为什么就可以用于说明变异由宜人性给予了解释？它的科学依据是什么？作者需要对文中涉及结论性的结果进行科学合理的说明。

回应：非常感谢专家的提问。我们认同仅基于数据结果进行解释是不够的，必须结合理论与数据共同说明。为此，我们在 4.3.2 节再次简述 2.1 节介绍的模仿效应（mimicry effects）作为科学依据说明，并明确将参数变化与理论机制联系起来。首先是在报告被试参数间相关的估计结果时指出，当数据存在反应分布不对称或特质与 RS 关联时，节点内单维模型会将本应由 TOI 解释的变异错误地归因于 RS，导致 RS 与 TOI 的相关被高估，而引入多维结构后，该部分变异被重新归于宜人性特质，因而 RS 与宜人性的净相关下降，这正是模仿效应的直接证据。然后在解释图 6 的斜率下降结果时，补充了如下论证：“当 RS 方差固定为 1 时，斜率衡量了 RS 对被试在题目上局部决策的贡献。当宜人性特质作为全局因子被纳入模型后，若 RS 在某题上的斜率显著下降，说明原单维模型对该斜率的估计中混入了本应由宜人性解释的变异。”

意见 16：5.1 第一句话，文章指出“在 RS 研究中，始终存在着两种既相互区别又彼此关联的分析视角”，紧接着文章分别阐述了两个视角。但是并没有明确说明两个视角究竟是如何相互区别又彼此关联的。另外，文章在引言和演进章节也没有说明 IRTree 的演进模型如何详细地体现这两个视角，只在讨论部分进行了简单地陈述，换言之，这两种视角在前文并没有被充分地建立起来。

回应：非常感谢专家的宝贵意见。我们同意原文对“两种视角”的区别与关联阐述不够清晰，且在前文中未充分建立。为此，我们在引言第一段末尾明确点出了 RS 研究中两种并存的分析视角：“一种将其视为需要控制的测量偏差，另一种将其视为值得探究的认知过程。两种视角各有侧重，亦可相互补充。”紧接着在第二段指出 IRTree 模型“能够兼顾偏差控制和过程理解的双重分析视角”。这样在文章开头就建立了两个视角的区分与关联。

过程理解视角侧重对作答策略的分析，而这基本贯穿了本文框架演进的两条主线；至于偏差控制视角，在框架演进部分主要在节点内多维结构抵消模仿效应等处简略提及，更多留到讨论 5.1 节来总结。5.1 节对 IRTree 框架演进如何体现这两个视角做了更加充分的讨论，并且回扣标题“RS 建模的深化”来进一步强调 IRTree 对两种视角的兼顾：

“IRTree 框架演进清晰呈现了 RS 研究的一种深化趋势：从早期将 RS 仅仅视为需要清理的噪声，逐步转向将其同时视为需要理解的认知策略成分。这种深化也并非否定偏差控制的必要性，而是主张在理解过程的基础上进行控制。……以上两种思路服务于不同的研究目标：若研究核心在于精确测量特质，可优先选用简洁的偏差控制方法；若旨在探究作答心理机制或检验 RS 与 TOI 的实质性关系（如跨文化研究中的群体策略差异），则 IRTree 的过程解释力难以替代。总之，研究者需根据自身研究目标，在两种视角之间做出审慎的选择。”

意见 17：5.2 所列问题虽然比较具体，但这些问题和前文之间的对应关系并不是很明确。比如 4 点和 5 点量表在李克特量表研究中非常常见，而 6 点及以上的量表则比较少，其使用情境也相对有限，如此将其视为待解决问题是否科学和有逻辑依据，这需要进一步说明。再者，测量不变性问题的阐述也较为突兀，前面只是提及了方向不变性问题，文章对不变性问题并没有明确进行说明，最后冒出测量不变性的问题阐述就显得突兀。另外，作者将项目功能差异（DIF）和测量不变性视为同一个概念，这其实犯了概念混淆的错误，测量不变性和 DIF 并不是同一个东西，作者需要说明 DIF 和测量不变性之间的内在关系，才能将 DIF 放在测量不变性框架下进行阐述。

回应：非常感谢专家的宝贵意见。我们已据此对 5.2 节做了充分修改。

首先，对于 6 点和 7 点李克特量表，我们引用了文献支持其同样应用广泛：“虽然 4 点和 5 点量表在实证研究中最为常见，但 6 点及 7 点量表同样得到了广泛应用(Böckenholt, 2017; Henninger et al., 2025)”，且由于其关联着 Meiser 等(2019)提出的多级计分节点问题，因而与 IRTree 结构的扩展联系紧密(“但都是现有 IRTree 模型中二分极端反应过程的进一步延伸”)，值得被视为待解决问题。

最后，“测量不变性”的表述确实易造成费解，和项目功能差异（differential item functioning, DIF）混为一谈也确实不够严谨。目前将第三段的首句删去“测量不变性”，整体改为：“最后，在 IRTree 框架下，关于参数的等价性，考察范围可从树结构内部拓展至跨群体情境，呈现出独特的复杂性”。修改后第三段的逻辑可表述为：从树结构内部的方向不变性，到协变量（如措辞方向为代表的题目特征）对参数等价性的影响，再到多组比较情境的 DIF 问题。

次要问题：

意见 18：标题中的“深化”没有被概念化，作者使用了一个很强的判断词，但没有在引言中明确其内涵和评判标准。究竟是什么的深化？内容的深化？生成机制的深化？建模方式的复杂化？

回应：非常感谢专家的提问。我们已在引言中阐述综述必要性时（见意见 2 的回复）增加了如下论述：“第一，围绕 IRTree 开展的方法学研究持续积累，推动了 RS 研究从单纯的偏差控制转向兼顾过程理解，这一研究取向的转变值得阶段性回顾。第二，IRTree 模型的一系列拓展也并非零散的技术更新，而是围绕作答过程表征这一核心命题的建模精细化，这构成了 RS 研究深化的另一重要层次。”由此，我们将标题的“RS 建模的深化”明确为两个层次，一是研究范式的深化（从单纯偏差控制到兼顾过程理解），二是建模方式的深化（即建模精细化。前者是后者的方向指引，后者是前者的技术支撑。

意见 19：对已有的文献综述的阐述过于简略。
回应：非常感谢专家的意见。我们在引言部分专门增加了对现有综述不足的评述，明确指出："已有的关于 IRTree 模型的教程或综述，或仅关注某一类拓展(Plieninger, 2021; Xue & Chen, 2023)，或在综述各种 RS 建模方法时概括性地论及 IRTree 模型(Bolt & Meng, 2025; 张缨斌, 王烨晖, 2019)；尤其在国内，相关综述由于时间较早，仅介绍了 IRTree 发展初期的基础模型(张缨斌, 王烨晖, 2019)。"在此基础上，我们进一步说明本文将从框架演进的角度系统梳理 IRTree 模型的发展，以弥补上述缺口。此外，在第 2 章节和第 3 章节的各小节中，我们系统补充了各模型适用情境、样本量需求、识别约束与估计稳定性等方面的评述（见意见 11 的回复），确保文献综述的广度和深度。

意见 20：表 1 上面的一句中的"题目间多维"具体指什么？各维分别代表什么？
回应：感谢专家的提问。此处的"题目间多维"是"伪题目间多维"，具体指的是"每道原始题目所分解出的三个节点伪题目分别对应于 MRS、TOI 和 ERS 三个不同维度，从而构成节点间的多维结构"。已将原文该处表述修改得更加清晰。

意见 21：文中存在错字和别字情况，如 1.3 章节第二段话"理论分野首先体现为将 RS 概念化为潜类别还是潜在特质。"中的"分野"。
回应：感谢专家的意见。已删去"分野"的表述。并仔细检查全文，排除错字和别字情况。

意见 22：文中多次出现了"方向不变性问题"这一专业术语，作者需明确说明该术语表达什么意思，在 IRTree 模型中又有什么作用。
回应：感谢专家的意见。我们已在 1.2 节进一步明确"方向不变性"的定义："IRTree 的方向不变性通常指在'同意'与'不同意'两侧 ERS 的作用完全一致，这意味着模型中只使用一个 ERS 潜变量和一套伪题目参数来同时刻画两种态度方向的极端反应。"然后也补充论述讨论方向不变性问题的作用："如果方向不变性成立，则支持将 ERS 视为一种跨方向的一般性极端反应倾向，便于模型简化与解释。但现实中，被试可能对正向极端和负向极端表现出不同偏好，例如倾向于选择'非常同意'但避免'非常不同意'；某些题目也可能更容易引发正向极端反应而非负向极端反应。此时，放松方向不变性假设，分别用两个潜变量（正向 ERS、负向 ERS）和两套伪题目参数建模，能更精细地捕捉这种差异。"

意见 23：从 2.2.1 和 2.2.2 章节内容来看，IRTree 模型在处理 RS 时的发展并不是从优势过程到理想点过程（A 到 B），而是不同节点、甚至同一节点内同时容纳两类反应过程的混合建模（A 到 AB）。
回应：感谢专家的意见。我们同意专家的观点，节点内反应函数的拓展更准确的表述应为"从所有节点为优势过程到纳入理想点过程节点"。这也启发我们注意到潜变量结构的拓展也并非得到"节点内多维 IRTree 模型"（因为 TOI 的方向决策依然是单维结构）。由此我们调整了全文的措辞，将"节点内多维 IRTree 模型"都调整为"含节点内多维结构的 IRTree 模型"，将 2.1 节和 2.2 节的小标题分别改为"潜变量结构的拓展：节点内多维结构的纳入"和"反应函数的拓展：理想点过程的纳入"。

意见 24：由表 4 可知，对于 GPCM 和 GGUM 两个模型，GGUM 在对数似然和 AIC 方面要优于 GPCM，而 GPCM 则在其他三个指标上的表现优于 GGUM，对于这种两个模型在不同指标上的表现不相同的情况下，文章最后选择了 GPCM，其理由是什么？
回应：感谢专家的提问。我们已经修改了 4.3.2 节，更详细地阐明了 TOI 反应过程确定（IRT

模型阶段）中模型选择的依据（见意见 14 的回复）：虽然 10 道题目表述方向明确，理论上支持优势过程，但考虑到人格测验中理想点过程也很常用，故同时拟合 GPCM 与 GGUM。最终未仅依据拟合指标（GGUM 虽然 AIC 略优，但其他信息准则（BIC、SABIC）均支持 GPCM），更重要的是 GGUM 对第 3 题的位置参数估计极不稳定（点估计与 95%置信区间为 6.45[-1.81, 14.71]），而 GPCM 无此问题。因此，综合理论、拟合指标与参数估计稳定性，选择了更简洁的 GPCM 作为 M_0。

意见 25：在 M0 的模型使用方面，作者使用了 GPCM 和 GGUM，据审稿人所知，对于李克特量表数据，GRM 的拟合效度在绝大多数情况下要优于 GPCM，这种情况下文章为何选用 GPCM？选择 GPCM 而不是 GRM 的科学依据是什么呢？

回应：感谢专家的提问。我们认为在实证分析中采用 GPCM 和 GRM 都是可以的，优势过程下选择哪一个模型进行拟合并不是我们所关注的问题。这里我们主要出于前文比较过 GPCM 和 GGUM 的结构共性（“除总”结构）和差异（反应过程假设不同）所以应用二者来作为优势过程和理想点过程的代表。专家的意见也启发我们在 4.3.1 节第 2 段更明确地补充交代：“（以 GPCM 和 GGUM 分别作为优势过程和理想点过程的示例模型，不纠结于同一反应过程下多种 IRT 模型的比较）”。

意见 26：在 4.2.2 章节，文章指出“对于节点内单维 IRTree 模型，先确定了两个方向的极端反应节点仅约束斜率等价”，审稿人的疑问是如何确定？为何仅约束斜率等价？

回应：感谢专家的提问。在 1.2 节首次提到方向不变性问题时我们明确了文献的做法：“Jeon 和 De Boeck(2019)建议从基准模型开始逐步释放方向不变性的假设，并通过比较模型拟合的方式来选择解释力更高的模型结构。”针对专家的疑问，我们在 4.2.1 节明确了节点内单维 IRTree 模型的参数约束确定方式：“参照 Jeon 和 De Boeck(2019)的建议，从斜率和截距均相等的模型开始逐步释放约束，根据拟合变化与参数可解释性来确定最终形式。”并且在 4.3.2 节的实证演示中更详细地叙述了选择仅约束斜率等价的原因：“在其两侧极端反应节点上尝试了斜率与截距相等、仅斜率相等和无约束三种参数约束，除 AIC 外的信息准则均支持仅约束斜率相等，故将采用该约束方式的节点内单维模型选为 M_1”。

此外，我们也在 2.1 节谈到节点内多维结构下的参数不变性问题：“研究者倾向于设定两侧极端节点的斜率参数绝对值对应相等（即 ERS 与 TOI 的作用强度在两侧相同），而截距参数自由估计。由于节点内多维结构的截距不再对应明确的‘基线难度’含义，自由估计截距是合理做法；至于斜率绝对值跨方向相等的假设是否成立，可沿用 Jeon 和 De Boeck(2019)的模型比较方法检验。”并在 4.2.2 节明确“建议先按照文献中的惯例，约束两侧极端节点上 ERS 斜率相等、TOI 斜率互为相反数，以保证模型识别。研究者亦可尝试放松该约束，通过模型比较来评估是否要设定斜率约束”。

希望以上说明能解决专家的疑问。

意见 27：在建议流程和应用示例中，文章都在优势模型中引入理想点过程，但对于理想点过程的引入顺序，作者没有讲清楚。如在应用示例中，作者先后在节点 2、节点 1 引入了理想点过程，但并没有讲清楚为什么是先节点 2 再节点 1 这样一种顺序，它的依据是什么？换言之，文章对何时引入理想点过程，何时不引入理想点、依次在哪些节点引入理想点过程等并没有清晰的说明。

回应：感谢专家的提问和意见。首先，我们认为要区分两种不同性质的理想点过程使用：（1）节点 2（方向决策）：这是真正意义上的反应函数拓展，即判断 TOI 对方向决策的影响遵循优势过程还是理想点过程。引入顺序必须理论优先：研究者应基于构念本身的测量理论

做出判断，当理论不明确时再同时拟合两种模型进行比较（见 4.2.1 节及 4.3.2 节）。
（2）节点 1（折中反应决策）：此处引入理想点过程并非讨论 TOI 的反应函数形式，而是为了在节点 1 同时纳入 MRS 与 TOI 而借用理想点过程的数学形式（因其单峰特性可自然刻画 TOI 在中间选项上的作用）。我们已在 2.3 节和 4.3.2 节中将此类应用明确表述为“借用理想点过程刻画”，以区别于反应函数的实质拓展。

其次，针对“为何先节点 2 后节点 1”的顺序：在应用示例中，节点 2（方向决策）的理想点过程属于反应函数拓展，应在阶段一就明确；而节点 1 的理想点过程仅作为节点内多维结构（MRS+TOI）的一种实现手段，属于节点内两种反应过程共存的建模（2.3 节），其前提是节点内多维结构已被论证为必要（即 M_2 优于 M_1 后再考虑 M_3）。因此，先处理方向决策的反应函数选择，后借用理想点过程处理折中节点的多维结构更符合逻辑顺序。

最后，正如我们在回应意见 12 和意见 14 时提到的，在修改 4.2 的过程中，我们已经明确说明了在阶段一即需判断 TOI 的反应函数，并且需坚持理论优先，综合理论、拟合指标与参数估计稳定性来判断；在阶段二考虑节点内多维结构时，也说明了只有 5 点计分且 M_2 优于 M_1 才考虑借用理想点过程实现折中反应节点的多维结构。

意见 28：4.2.2 章节第一段最后一句话“…给出了更可靠的相关估计：MRS 与宜人性水平无显著相关，而 ERS 与宜人性水平存在显著但较弱的正相关”，作者做出该论断的科学依据是什么？得到无显著和弱相关的结果就可以说明可靠吗？依据是什么？
回应：感谢专家的提问。和对意见 15 的回复一致，我们在 4.3.2 节再次简述 2.1 节介绍的模仿效应作为科学依据说明，并明确将参数变化与理论机制联系起来。除了意见 15 的回复中提到的修改，我们也针对“更可靠的结论”做了更多阐述：“最终的模型 M_3 通过控制宜人性特质在所有节点上的全局影响，在统计上分离了共同变异与局部特定变异，最大程度排除了因模仿效应造成的虚假共变，提供了相对更可靠的结论……”。因此，M_3 给出的相关估计之所以“相对更可靠”，并非因为结果本身显著或不显著，而是因为它是在控制了模仿效应之后获得的，相比估计被扭曲的单维模型 M_1 和仅控制部分模仿效应的 M_2 更接近真实参数值。

意见 29：5.3 展望部分，文章对 IRTree 模型迈向整合建模框架的展望过高，导致其证据支撑性不足。综合 5.3 章节的内容来说，这些发展情况只能说明 IRTree 模型具有较好的发展潜力，至于能不能实现整合，文章并没有提供充分的证据。
回应：感谢专家的意见。我们已对 5.3 节的最后一段进行了修改，弱化了相关前瞻性论断，并明确补充了“这一思路尚无文献探讨，其可行性与适用条件尚需后续研究检验”等谨慎性表述，以体现这一节的展望只能说明有发展潜力，尚无明确文献依据，有待更多探索。具体修改见 5.3 节末段。

……

审稿人 2 意见：

论文对 IRT 中反应风格识别和控制的 IRTree 模型进行了系统性梳理，总结了 IRTree 的核心逻辑和发展方向，体现出较高的专业性。但在论文结构、表述的严谨性和实践建议部分依然有优化的空间。

意见 1：在引言中，只指出反应风格是一种特异性倾向不足以解释清楚，应该参照已有文献对其进行清晰定义。
回应：衷心感谢专家的宝贵意见。我们参考了其他经典文献，选择了更清晰的 RS 定义：“RS

是指被试脱离题目内容、基于无关因素作答问卷题目的系统性倾向(Baumgartner & Steenkamp, 2001; Paulhus, 1991)”。新增文献：

Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). Academic Press. https://doi.org/10.1016/B978-0-12-590241-0.50006-X

意见 2：在 IRTree 之前，研究者已经进行过量化反应风格的诸多探索，应该在引言中进行简要阐述，体现 IRTree 在 RS 测量和控制中的方法学位置。

回应：衷心感谢专家的宝贵意见。我们已在引言第二段增加了对 RS 测量方法的简要回顾（从早期计数法到基于 IRT 的模型法，再到 IRTree 模型的提出）。

意见 3：“IRTree 从单维到多维、从优势过程到理想点过程，以及纳入作答过程异质性”。总结这些的实践意义是什么，需要在引言部分加以说明。

回应：衷心感谢专家的宝贵意见。正如回复审稿人 1 的意见 1、2 和 3 时所述，我们已完善了引言对于问题意识、综述必要性以及研究贡献的论述。当中从实践层面提到了当前研究现状“对研究者构成一定障碍：方法学研究者因缺乏对 IRTree 演进脉络的系统梳理，难以进一步改进和拓展模型；应用研究者因缺乏清晰的理论假设说明和实践指引，面对众多变体时容易无所适从”。而本文的目标是通过梳理和总结框架演进来解决上述问题，具体是“首次系统梳理 IRTree 从基础模型到前沿变体的演进历程，结合已有研究成果总结 IRTree 主要变体模型的统计结构、理论假设与适用边界，并提供递进式分析流程与实证演示”，为研究者提供“兼顾理论深度与实践操作的中文指引”。

意见 4：在许多 IRTree 开发的论文中，研究者明确声明对于作答过程的顺序假设，而 1.2 部分第一段写到“虽然 IRTree 模型通过条件概率表征结构，似乎隐含某种顺序，但不应过度解读…”。作者前人研究者的普遍性认知如何看待，应该给予更加清晰和具有说服力的论述。

回应：衷心感谢专家的宝贵意见。我们认识到，1.2 节中原有的不应过度解读顺序的观点与 IRTree 文献中普遍存在的顺序假设之间存在张力，论述不够清晰有力。为此，我们在 1.2 节中做了如下折中与补充：首先依然指出 IRTree 模型的统计结构核心在于通过条件概率分解作答过程，其本身对节点顺序并不严格敏感，更本质的是描述了包含哪些决策过程以及每个过程受哪些被试参数影响。但是，在实践层面，研究者可以采用“顺序解释”来陈述建模假设，以符合认知直觉；但在解释结果时，采用比“顺序解释”更弱的“成分解释”（即仅强调各节点所代表的心理成分及其参数含义）更为稳妥，避免对顺序的过度依赖。

意见 5：论文不同模块之间缺乏必要的衔接。如为何会从节点内单维发展到节点内多维，应该在第一章和第二章之间做好衔接，减少罗列感。其他部分类似。

回应：衷心感谢专家的宝贵意见。正如在回复审稿人 1 的意见 6（1）时所述，我们在第一章和第二章都增加了铺垫性文字（1.1 节即点出“节点间多维结构”，第 2 章节的引入段里点出“节点内单维结构”，然后 2.1 节进一步过渡到“节点内多维结构”）以实现更好的衔接。其他部分我们也做了类似的努力，例如增加了第 3 章节的引入段，在 2.3 和 3.3 节的开头都强化了其与前两节的联系。

意见 6：论文 3.1 部分，“同质作答过程”的假设过于简化，请举例说明理由。

回应：非常感谢专家的意见。我们已在第 3 章节的引入段增加了一个“工作满意度调查”的小例子，用来说明“同质作答过程”可能如何被违背，并由此引出本章要梳理的三类异质性。

意见 7：4.1 实践建议部分，对于选择理想点还是优势模型的决策，应该根据量表所测量的内容、数据形态等综合考虑，而非完全基于拟合指标进行数据驱动。
回应：衷心感谢专家的宝贵意见。正如回复审稿人 1 的意见 7 和 14 时所述，我们已在 2.2 节和 4.2 节都强化了如何纳入理想点过程的流程建议的相关论述，进一步强调了 TOI 反应过程的设定需要坚持理论优先（结合量表内容等判断），只有在理论不充分时才拟合两种过程的竞争模型，并且依然要综合理论、拟合和参数可解释性来判断。

意见 8：虽然提供应用示例可以帮助读者理解模型的应用步骤，但模型的复杂性限制了实际的理解和应用。作者是否可以在补充材料中提供一个 tutorial，完善从数据导入到结果导出与可视化的全流程指南。
回应：衷心感谢专家的宝贵意见。我们已将原补充材料中的 R 语句转为 rmd 语句，并提供相应渲染出的 html 文件作为可读性更强的全流程 tutorial，其中提供了从数据导入到阶段一、阶段二每一步的代码和结果输出，包括最后参数估计的可视化。

意见 9：表 4 中的模型名称难以理解，不同模型是否有规范的命名加以区分
回应：衷心感谢专家的宝贵意见。我们在修改后的 4.3 节已换掉基于 Merhof 和 Meiser(2024) 的较难理解的符号表示，统一用 M_1、M_2、M_3 表示 IRTree 模型，或者用文字叙述模型结构。

意见 10：实证分析中，不同 IRTree 模型之间的拟合结果差异不大，是否有必要使用更复杂的模型？建议报告不同模型的参数估计用时情况，结合使用成本和模型增益综合阐述模型选择建议。
回应：衷心感谢专家的宝贵意见。按照专家的建议，我们考察了不同 IRTree 模型的用时。在 4.3 节的示例中，M_1, M_2 和 M_3 用同一台计算机（操作系统为 Windows 11（版本 25H2），处理器为 AMD Ryzen 7 H 260 w/ Radeon 780M Graphics (3.80 GHz)，内存为 32.0 GB RAM。分析软件为 R 4.4.0 (x64)）的估计用时分别为 127.55 秒、743.58 秒、1961.69 秒。模型复杂度的上升的确造成参数估计用时明显延长。然而，我们在修改 4.2 流程建议时认为模型选择应优先基于理论合理性、拟合改善与参数可解释性，用时本身不作为主要判定依据。在本示例中，M_3 虽然在时间成本上最高，但在理论假设上更为精细（宜人性特质全局影响），拟合改善明显，且参数估计合理（模仿效应明显削弱），因而综合收益超过了其时间成本。同时，我们已在 5.2 节补充了对时间成本的讨论，指出该问题在特定场景下（如快速分析、大规模模拟）确实构成应用阻碍，可能需通过估计工具优化或树结构设计来缓解。具体修改见 5.2 节。

意见 11：作者对 IRTree 过程树是否有助于理解被试作答认知过程存在一定的前后矛盾。前文阐述 IRTree 决策树不能过分解读为作答过程假设。后文阐述“IRTree 贡献对过程的理解”，这是否隐含着 IRTree 天然具有作答反应过程假设，可以通过这种树结构理解作答过程？如果不能理解，IRTree 对理解作答过程就没有贡献？IRTree 过程树究竟能否显式理解为作答过程，还是作答过程的其中一种形式？如果只是其中一种形式，为何？建议在补充材料中以数学论证的形式阐明这一观点，并在正文中注意表述的一致性。
回应：衷心感谢专家的宝贵意见。针对专家的意见，我们修改了 1.2 节，主要做了以下澄清。

首先，我们通过链式乘法法则（公式(9)与公式(10)的对比）从数学上证明：IRTree 模型对节点顺序不敏感，联合概率在任意调换节点顺序时能写成不同的条件概率乘积的形式，因此树结构并不预设心理加工的时间顺序，只提供其中一种可能的顺序形式。这一论证已在正文中给出。

其次，基于这一数学论证，我们明确区分了“顺序解释”与“成分解释”：IRTree 的树结构不应被理解为被试依次经历各决策节点的时间顺序，而应理解为对作答结果的边际概率进行分解，将一次完整反应拆解为若干心理成分的组合。每个节点代表一种可能的心理过程（如是否折中反应、是否同意、是否极端反应），但节点之间的条件概率关系仅表示逻辑依赖，而非时间先后。正如对意见 4 的回应所述，我们也对此问题做了适当的折中论述。但这不影响我们综合已有文献观点后所持有的“成分解释”立场。

基于上述澄清，我们已修改正文中可能引起歧义的表述，确保全文在“IRTree 是否代表作答过程”这一问题上保持一致：IRTree 提供的是对作答结果的成分分解，而非对心理加工时间顺序的刻画。具体修改见 1.2 节及全文相关表述。

意见 12：作者提出的“验证性的递进式实证分析流程”可能在大部分数据上展示出复杂模型的最佳拟合，模型越复杂，拟合越好。然而，数据分析不应单纯以模型拟合指标为唯一指导，而应考虑成本、数据情况、测验内容等多种因素。建议作者弱化对于分析流程的建议，在实践建议中考虑多种实际因素。

回应：衷心感谢专家的宝贵意见。正如对审稿人 1 意见 14 的回应，我们完全认同数据分析不应以模型拟合为唯一指导，而应综合考虑多种因素。为避免“模型越复杂拟合越好”的误导，我们在 4.2 节中删去流程图，仅保留三个阶段作为一般的分析流程，转而侧重于各阶段的分析注意事项。在模型选择依据方面，我们在给出了更平衡、更综合的阐述。例如在阶段一（判断是否有必要采用 IRTree 建模）中，明确强调 TOI 反应过程的确定坚持理论优先，仅当理论依据不充分时才同时拟合优势与理想点模型进行比较，而非盲目追求复杂模型。还例如在阶段三（异质性建模）中，“建议仅当具备明确的理论动机（如预期存在‘无 RS 作答者’亚群体）且样本量充足（通常需 1000 人以上）”，并强调“应始终以阶段二确立的同质模型为基准，通过比较异质性模型是否带来实质性的理论增益与拟合改善来评估其必要性”。

此外，在第 2 和第 3 章节综述各类模型变体时，我们也多次提醒模型并非越复杂越好。例如 2.1 节末尾明确提到：“在样本量较小或理论假设尚不清晰时，节点内单维模型仍是更稳妥的选择”。

第二轮

审稿人 1 意见：

经过一轮修改后，论文质量得到较大提升，问题意识明显加强，论文的逻辑性得到提升，研究的价值和意义得到凝练。不过论文还存在一些问题，供作者参考。

主要问题：

意见 1：引言第三段，作者指出“不同模型各自隐含怎样的理论假设、又适用于怎样的分析场景，这一基本问题尚未得到系统回答”，这一表述略显过强。实际上，各类 IRTree 模型在其原始提出时通常都会说明自身的理论假设及大致适用范围，因此相关问题并非“尚未回答”，而是其答案分散存在于不同模型论文之中，尚缺乏系统性的梳理、比较与整合。但本文的这种表述给人的感觉是各模型的原始文献没有回答这些问题，而本文尝试替前人回答这些问题。

回应：衷心感谢专家的宝贵意见。我们完全认同原文表述容易引起误解。已按照专家的建议，将有关表述修改为：“然而，尽管 IRTree 模型变体日益丰富，不同模型各自隐含怎样的理论

假设、又适用于怎样的分析场景，有关这一基本问题的讨论虽散见于各模型的原始文献中，至今仍缺乏系统的梳理、比较与整合。”

意见 2：论文内容似乎出现了前后矛盾的情况。作者在 1.3 章节（第一段）强调相较于潜类别方法，连续特质方法更符合 ERS 个体差异呈连续分布的现实，并称其在多数应用场景中更为适用；但在 3.1 章节又将混合 IRTree 作为个体间异质性建模的重要扩展进行介绍。当前这两部分之间的关系说明不足，容易给读者造成前后不一致的感觉。

回应：感谢专家敏锐地指出这一问题。需要说明的是，1.3 节讨论的是反应风格（RS）本身应被概念化为连续特质还是潜类别，并指出连续特质方法可建模个体间 RS 强度的差异，而潜类别方法则假定类别内部 RS 水平同质，因此许多研究倾向于采用连续特质方法（该表述已根据意见 8 对文献引用的指正进行了修正）。而 3.1 节的混合 IRTree 模型并非将 RS 本身视为类别，而是在各亚组内部仍保持 RS 的连续特质假设，仅允许不同亚组在树结构或参数上存在差异，这与 1.3 节的表述并不矛盾。

为避免读者产生误解，我们已在原文中做了如下修改：（1）在 1.3 节明确区分：将 RS 概念化为潜类别的方法，是“允许不同类别下的 IRT 模型或因子分析模型参数各异”，即类别内部不引入 RS 连续特质；（2）在 3.1 节相应位置补充说明：“……混合 IRTree 模型也得到开发与应用，通过潜类别结构从数据中识别具有不同反应策略模式的亚组。该模型在各亚组内部仍将 RS 作为连续潜在特质进行建模，仅允许亚组间在参数或树结构上存在差异。”

意见 3：对个体内动态异质性建模的内容阐述相对简略，作者只是简单说了一下哪些研究者做了什么，至于具体怎么做的，并没有给予阐述。这和第 2 章节中的三个内容阐述相比，显得相对简单和粗糙。

回应：感谢审稿人的宝贵意见。3.3 节对个体内动态异质性建模的阐述确实相对简略。这主要是因为该方向的研究尚处于起步阶段，相关文献数量有限，且动态 IRTree 模型的数学细节较为复杂。为保持全文的详略平衡，避免因过度展开细节而冲淡框架演进的主线，我们最初仅做了概括性介绍。

针对专家的意见，我们已在修改稿中对 3.3 节的三篇代表性文献各补充了其核心公式及相应说明性文字，具体包括：（1）Merhof 和 Meiser (2023)中斜率随题目位置变化的轨迹函数及其参数解释；（2）Ames 和 Leventhal (2022, 2023)的纵向 IRTree 模型伪题目反应概率公式中对被试参数均值漂移的处理；（3）Hasselhorn 等(2024)的多水平 IRTree 模型中，TOI 或 RS 在个体内水平的分解方程。这些补充使读者能够大致了解模型的技术要点，同时仍保持综述的概括性。再次感谢专家的细致指正。

意见 4：既然前文第 2 章节将节点层面的演进概括为三类重要扩展，为何“4.2 流程建议”只是系统展示了节点内多维结构，而没有将“理想点过程的纳入”以及“两种反应过程共存的多维结构”纳入同等清晰的流程框架？

回应：感谢专家的提问。关于 4.2 节为何未将三类扩展同等清晰地纳入流程框架，我们做如下说明。

（1）关于理想点过程的纳入：本文将其视为与节点内多维结构同等级别的节点层面拓展，只是关于二者的流程建议所处的阶段有所不同。是否纳入节点内多维结构的考量集中于阶段二，而是否纳入理想点过程的讨论则集中于阶段一。为此，我们在 4.2.1 节阶段一中明确指出，在拟合基础 IRT 模型时即应优先确定 TOI 反应机制（优势过程还是理想点过程），并强调在后续 IRTree 建模时，TOI 所在的方向决策与普通 IRT 模型的反应过程应当保持一致，而 RS 作用的非方向决策默认优势过程。为了方便读者更清晰地理解，我们已调整 4.2.1

节阶段一的表述结构，将是否纳入理想点过程的讨论单独作为一段，以增强有关这一问题的流程清晰度。

（2）关于两种反应过程共存的多维结构：本文实际并未将其列为与节点内多维结构、理想点过程并列的第三类重要扩展。如 2.3 节所述，它是前两类扩展的整合形态，即在节点内多维结构的基础上，允许不同节点采用不同的反应函数（优势过程或理想点过程）。在流程建议中，该结构并不构成一个独立的建模步骤：若阶段一确定 TOI 应采用理想点过程，则阶段二在构建节点内多维结构时自然面临优势与理想点过程共存的问题（如极端反应节点同时受 ERS 优势过程和 TOI 理想点过程影响）；若阶段一确定 TOI 采用优势过程，则阶段二仅在 5 点量表情境下考虑在折中反应节点借用理想点过程建模（如公式(22)所示）。因此，我们没有将其与前两种扩展以同等清晰的方式专门列出。

希望以上说明能够澄清我们的考量。再次感谢您的宝贵意见。

意见 5：引言第一段对第二种视角的阐述过于简略和抽象，只使用一句话“从认知与跨文化视角看，RS 又是作答策略的关键成分”，建议再拓展一下这部分内容。

回应：衷心感谢专家的建议。我们已在引言第一段对第二种视角进行了拓展，补充了如下内容：“然而，从认知与跨文化视角来看，RS 又是反应策略的关键要素：在认知层面，RS 可体现为被试应对问卷认知负荷时采取的满意化（satisficing）策略，可能受认知资源节省倾向的驱动(Merhof & Meiser, 2023)；在跨文化层面，RS 则是值得挖掘的文化心理现象，例如 ERS 与移民群体的文化适应策略相关联，并反映在论证风格及其代际差异上(Morren et al., 2012)”。

次要问题：

意见 6：摘要指出“…从节点均为优势过程到尝试引入理想点过程，对作答理论机制的表征能力不断提升…”。审稿人认为无论是优势过程还是理想点过程，都只是从不同视角对李克特数据进行表征，至于二种方法是否具有递进关系（即理想点的表征能力优于优势过程），这需要强有力的科学证据。

回应：感谢专家的意见。我们同意优势过程和理想点过程只是从不同理论视角刻画反应机制，各有适用情境，不存在普遍的优劣或递进关系。已根据专家建议，将摘要的有关表述修改为“从节点均为优势过程到尝试引入理想点过程，对作答理论机制的表征方式更加丰富”，以体现视角的拓展而非能力的提升。

意见 7：摘要只是简单呈现了“两种视角”这四个字，没有说明究竟是哪两种视角，建议将“控制测量偏差和理解作答过程”加入到摘要中，如此，读者通过摘要便知晓两种视角的大致意思。

回应：感谢专家的建议。我们已在摘要中明确写入了“控制测量偏差与理解作答过程两种视角”，使读者通过摘要即可清晰了解其含义。

意见 8：3.1 章节第一段，作者指出“相比于强制类别内部同质的潜类别方法，连续特质方法更符合 ERS 个体差异呈连续分布的现实，因而在多数应用场景中更为适用(Schoenmakers et al., 2024)”，审稿人阅读了作者列出的参考文献，发现作者可能存在过度推论情况，在 Schoenmakers 等(2024)研究中，他们阐述的是由于类别 ERS 模型（即混合 IRT 模型）无法对类别内 ERS 强度差异进行建模，因此他们聚集于将连续 ERS 建模（pp. 147）。从 Schoenmakers 等(2024)的表述来看，他们并没有说“连续特质方法更符合 ERS 个体差异呈连

续分布的现实，因而在多数应用场景中更为适用”。

Schoenmakers, M., Tijmstra, J., Vermunt, J., & Bolsinova, M. (2024). Correcting for extreme response style: Model choice matters. *Educational and Psychological Measurement, 84*(1), 145-170.

回应：衷心感谢专家严谨的文献核查。在引用 Schoenmakers 等(2024)时表述确实过强，超出了该文献的支持范围。该句表述已改为“相比潜类别方法假定类别内部 ERS 水平同质，连续特质方法可建模个体间 ERS 强度的差异，因而更为灵活(Schoenmakers et al., 2024)”。

意见 9：2.1 章节第一段，作者写到“…对 TOI 与 RS 的作用提出了更完善的假设”，建议作者谨慎使用“更完善的假设”这类表述。Meiser et al. (2019) 和 Merhof 和 Meiser (2023) 的主要贡献更适合概括为对传统节点内单维 IRTree 假设的细化和扩展，而不是作为“更完善”的证明；细化和扩展并不意味着就更完善。

回应：感谢专家的建议。我们已将 2.1 节中对应位置的表述修改为“进一步拓展了对 TOI 与 RS 作用的假设”。

意见 10：从节点内单维和多维的数学表达式（公式 6-8 和公式 12-14）来看，可以较为清晰地知晓单维和多维之间的差异，但从图 1 和图 2 来看，感觉很难看出二者之间有什么差异。

回应：感谢专家的宝贵意见。图 1 和图 2 的树状结构确实难以直观区分节点内单维与多维的差异，对此，我们做如下说明。

我们认为，在 IRTree 模型中，树状图、模型方程与伪题目编码三者从不同侧面共同提供了对模型的理解，彼此互补（见 1.1 节）。树状图的核心功能是展示作答过程的成分分解，因而不一定要求在图中标注每个节点对应的潜变量结构（正如伪题目编码也不需要体现单维和多维的差异）。节点内单维与节点内多维结构的差异主要通过模型方程体现（如公式(6)-(8)为节点内单维，公式(12)-(14)为节点内多维），而二者可以共用一个树状图。这一处理方式也常见于已有文献 (如 Bolt & Meng, 2025; Meiser et al., 2019; Merhof & Meiser, 2024; Schoenmakers et al., 2023)。为了更明确树状图中不标注潜变量结构的考虑，已在正文 2.1 节添加如下脚注：

“参考已有文献惯例(Bolt & Meng, 2025; Meiser et al., 2019; Merhof & Meiser, 2024; Schoenmakers et al., 2023)，本文在树状图中仅展示作答过程成分分解，不标注节点内的潜变量结构是单维还是多维，节点内的潜变量结构通过具体模型方程来体现。因而同一树状图可以对应不同的模型方程。”

意见 11：在正文中对图 4 进行说明时，建议明确标志出新节点“是否默认肯定作答”是图 4 当中的哪个节点。从文章当前的阐述和图来看，应该是“节点 1”。

回应：感谢专家的建议。我们已在新节点“是否默认肯定作答”后加括号并注明为“图 4 中的节点 1”。

意见 12：4.2.1 到 4.2.3 三个三级标题内容基本上没有提供任何信息，建议将对应阶段主要做的事或任务给予概括并作为三级标题。

回应：衷心感谢专家的建议。现已根据各阶段的核心任务对三级标题进行了修改，使读者能够一目了然地了解每个阶段的主要工作。具体修改如下：

4.2.1 阶段一：判断 IRTree 建模必要性并确定 TOI 反应函数；

4.2.2 阶段二：评估是否纳入节点内多维结构；

4.2.3 阶段三：探索作答过程异质性。

意见 13：5.1 章节第一段话，作者阐述到“…IRTree 模型在 RS 研究中的独特价值与定位：它不仅能控制测量偏差，更能帮助探究作答背后认知过程。这也是 IRTree 区别于传统偏差控制方法的核心特征”，该表述过于绝对化。虽然文章在前面确实显示 IRTree 具有同时服务于偏差控制和作答过程刻画的潜力，但目前还没有强有力的证据来支持这么强的总结性判断。尤其是“更能帮助探究”“区别于传统偏差控制方法的核心特征”等表述带有很强的比较和独特性判断。

回应：感谢专家的宝贵意见。我们已弱化相关表述，修改为“本文重点梳理了 IRTree 在节点层面和过程层面的框架演进，共同揭示了 IRTree 模型在 RS 研究中的潜在价值，即它能够一定程度上同时服务于测量偏差控制和作答过程刻画两种视角的分析。这使得 IRTree 相较于传统偏差控制方法具有其独特之处。”

意见 14：部分语句的表达过于武断和绝对，如，第 2 章节第一段落，“…构成了框架演进的第一条主线”，这是作者对整个领域演进结构的概括性判断，而不是一个中性的章节说明。建议对全文进行仔细阅读，避免再出现此类表述。

回应：感谢专家的宝贵意见。首先，针对审稿人指出的“构成了框架演进的第一条主线”这一表述，我们将原文修改为“本文在梳理框架演进时将其作为第一条主线”，将判断限定于本文自身的梳理框架和组织逻辑，而非对整个领域演进的客观断言。相应地，第 3 章节开头的“构成了 IRTree 框架演进的第二条主线”也修改为“这一追问即引出本文梳理 IRTree 框架演进的第二条主线”，以保持全文语气一致。

我们也对全文进行了仔细检查，并对其中过于武断或绝对的表述做了弱化，例如将 2.1 节最后一句“节点内单维模型仍是更稳妥的选择”中“仍是”改为“可能是”，将 2.3 节结尾段的第一句“代表了 IRTree 模型在节点层面精细化建模的最新进展”中“代表了”改为“可视作”等等。

审稿人 2 意见：

作者本轮对稿件做出大幅度修改，全面回应并妥善解决了上一轮提出的所有评审问题；全文行文逻辑、论证力度与可读性均有明显提升，稿件质量满足刊发要求，同意录用发表。

回应：衷心感谢审稿专家对我们上一轮修改的肯定与认可。再次感谢专家付出的时间与专业支持。

第三轮

审稿人 1 意见：

作者经过两轮修改，论文质量得以进一步提升，作者较好地回答和解决了审稿人问题、意见和建议，论文逻辑更加通顺、内容丰富、论证更加科学。但论文还存在需要修改的地方。

意见 1：对于表 1 和表 2 中的缺失值，作者用 NA 表示，这种表示方法通常出现在 R 语言这类统计软件中，至于期刊论文中的缺失值，似乎并不会用 NA 表示，建议作者采用教育与心理统计领域权威期刊论文（或者 APA 手册）对缺失值的表示方法。

回应：衷心感谢专家的宝贵意见。我们经过调研确认，尽管部分文献如 Schoenmakers 等(2024)使用 NA 标记 IRTree 模型伪题目编码的计划缺失值，但教育与心理统计领域绝大多数权威

期刊论文，如 *Psychological Methods* 期刊的论文 Böckenholt (2012)、*Organizational Research Methods* 期刊的论文 Plieninger (2021)、*Behavior Research Methods* 期刊的论文 Merhof 和 Meiser (2024)等确实普遍采用短横线（dash）“—”来表示伪题目编码中的计划缺失值。我们认为这种表示方法更为规范，已将表 1 和表 2 中的“NA”全部修改为“—”，并增加表注：“—”表示该节点不适用（计划缺失），即在对应的作答类别下该伪题目无需进行判定。

意见 2：在对表 5 结果进行解读时，存在以下问题：

（A）. “模仿效应的直接证据”这一结论偏强。作者发现，随着极端反应节点和折中反应节点被扩展为节点内多维结构（M2），ERS、MRS 与宜人性的相关有所下降。该结果模式确实和模仿效应的理论预期相一致，但尚不足以视为模仿效应的“直接证据”。 Merhof、Böhm 和 Meiser（2023）在界定模仿效应时，不只关注“相关变小”，而关注 RS 因子错误吸收 TOI 引起的变异，由此造成 TOI—RS 协方差估计偏差，并伴随 RS 方差高估；而在本研究的真实数据中，TOI 和 RS 之间的真实相关及参数真值未知，因此无法确定节点内单维模型中的相关是否确实被高估，也不能确认多维模型中的相关已更接近真实值。

（B）不同模型中的潜变量相关未必具有完全相同的含义。从 M1 到 M2、再到 M3，各模型的节点载荷、潜变量作用路径及方差分解方式发生了变化。相同名称的 ERS、MRS 和 TOI 潜变量在不同模型中的测量结构并不完全一致，因此其相关参数未必能够被直接视为同一参数在不同模型中的简单变化。作者需说明不同模型之间潜变量尺度和实质含义是否具有可比性，同时应避免将相关下降直接解释为虚假共变被清除。

（C）研究并没有直接检验不同模型中相关参数的差异。作者主要依据相关系数点估计的变化作出判断，例如 ERS 和宜人性的相关从.32 降至.16，再降至.12；MRS 和宜人性的相关从−.54 降至−.07。但文中未直接检验这些相关差异是否达到统计显著。需要注意，一个相关显著而另一个相关不显著，并不意味着两个相关之间存在显著差异。建议采用参数自助法、Delta 方法或贝叶斯后验差值分析，对模型间相关参数差异进行直接检验。

（D） “净相关”“最大程度排除虚假共变”等用语缺乏充分依据。作者将节点内多维模型中的相关称为“净相关”，并认为 M3“最大程度排除了因模仿效应造成的虚假共变”。这些表述隐含了模型已正确识别并完全分离真实共同变异与局部变异，但当前分析尚不能排除树结构错设、反应函数错设、遗漏作答过程、参数约束不当或作答异质性等其他可能性。

（E） M3 的拟合优势不能直接等同于其实质解释更可靠。表 5 显示，M3 在当前候选模型中具有更优的信息准则，这可以支持其相对拟合更好，但不能据此直接认定 M3 对真实作答机制的刻画更准确，或其参数结论必然更可靠。模型复杂度提高也可能带来过拟合、参数不稳定或解释依赖性增强。建议作者补充收敛情况、标准误、参数稳定性、重抽样或敏感性分析，并将结论限定在当前数据与候选模型范围内。作者在撰写内容时，一定要保持客观和公正，尽量不要过度推论；再次对论文全文进行仔细审查。

回应：衷心感谢专家的宝贵意见。在结论严谨性和措辞准确性上，这一部分的表述的确需要提升，我们逐条认真学习并逐一修改。

（A） 关于“模仿效应的直接证据”表述偏强

如专家所指出的，在真实数据分析中 TOI 与 RS 的真实相关及参数真值未知，确实无法确认单维模型中的相关是否被高估，也无法断言多维模型中的相关已更接近真实值。因此“直接证据”的表述确实过强。我们已将原文这句话调整为：“这一模式与模仿效应的理论预期一致(Merhof et al., 2024)”，同时删去“导致相关高估”等有关表述。斜率估计的部分也删去了类似“原单维模型中 ERS 斜率的估计受到了模仿效应的干扰”这样明确指向模仿效应的表述。

（B） 关于不同模型间潜变量的可比性

我们认同同一名称的 ERS、MRS 和宜人性特质在不同模型中的测量结构并不完全等同，因此不宜将跨模型的相关下降直接解读为同一潜变量间关系的真实变化。此处我们关注的也并非是对同一参数进行跨模型直接比较，而是观察同一名称参数在不同模型设定下估计值的变化方向（例如相关估计值是增大还是减小、斜率估计值在各题目上如何变化），并将这种变化趋势与模仿效应的理论预期方向进行对照。为了避免表述产生歧义，我们在 4.3.2 节第 2 段和第 3 段将相关和斜率的结果全部强调为“相关估计值”和“斜率估计值”，并在第 2 段补充一句：“也就是说，跨模型的相关系数估计的变化反映出不同模型设定下参数估计模式的差异”。我们也认同不宜将相关下降直接解释为虚假共变被清除，因此也删掉了相关表述。

（C） 关于模型间相关估计差异的统计检验

我们同意在未进行直接差异检验的情况下，不宜将相关系数的点估计下降解读为显著变化。而跨模型相关变化的统计检验并非本研究分析的核心关注点，因此我们在正文中弱化了有关叙述，避免做出未经统计检验的推断。已在正文中删除“显著下降”“明显下降”等隐含统计比较的表述，改为仅描述点估计的变化趋势（如“ERS 与宜人性的相关估计值有所下降”），并结合(B)的修改明确指出跨模型相关变化反映的是不同模型设定下参数估计模式的差异，避免将其解读为同一参数在不同模型间的真实变化。

（D） 关于“净相关”“最大程度排除虚假共变”等用语不妥

结合(B)的修改，我们已在文中删除上述缺乏充分证据的措辞，将有关表述修改为：“在本文所考察的候选模型范围内，M_3 表现出相对最优的拟合，为宜人性与两种 RS 的关系提供了可供参考的估计结果”。斜率估计的部分也相应修改为“上述对比显示，在本样本数据中，节点内单维模型下 MRS 节点受宜人性特质的影响可能较 ERS 更为严重，从而提示在折中反应节点也引入多维结构的潜在意义”。在这两段整体避免使用“校正”、“清除”等暗示模型已正确分离所有变异的用语。

（E） 关于 M_3 的拟合优势不等于实质解释更可靠

首先我们同意 M_3 的拟合优势不能直接等同于其实质解释更可靠，或者对真实作答机制的刻画更准确。我们已在文中调整上述措辞，正如（D）中提到的，强调参数估计结果可供参考，且仅在本文所考察的候选模型范围内。由于模型选择的过程已在 4.3.2 节第一段有明确说明，而第二段和第三段的内容聚焦于参数估计结果的解读，为避免偏离该部分的核心分析焦点，我们未在修改稿中再额外展开收敛情况、参数稳定性、重抽样或敏感性分析等讨论。

我们也遵照专家的意见仔细检查了全文，尤其关注结果解释的部分，修改了其他偏强或过度延伸的表述，例如将“这说明模仿效应也存在于 MRS 的测量中”修改为“该趋势提示模仿效应也可能存在于 MRS 的测量中”。

意见 3：论文引用了一些在线发表的期刊论文，但到目前为止，有些论文其实已经正式发表，但作者依旧引用的是在线发表的时间，建议作者仔细核对各期刊正式刊载的年份，如 Merhof et al.（2023）的论文于 2023 年在线发表，但已经于 2024 年正式发表了。

回应：感谢专家的细致指正。我们确认 Merhof 等(2023)的论文确实已于 2024 年正式发表。此外，经过逐一核对，我们发现参考文献列表中另有其他 3 篇在线发表论文也已于近期正式发表。现已将这 4 篇文献的引用信息统一修改为正式发表版本，在参考文献列表中全部更新完毕，其中 2 篇的出版年份有变更，正文中涉及这些文献的引用年份也已一并修正。再次感谢专家的严谨把关！

第四轮

审稿人 1 意见：

经过新一轮修改，作者较好地回答了审稿人的意见、问题和建议，论文质量较上一版得到提升，论文文献综述全面、问题提出合理、逻辑通顺、结果和结论可靠，达到了期刊发表的要求。建议发表。

编委 1 复审意见：

该文经过多轮审稿和修改，质量得到提升。同意审稿人意见，建议发表。

回应：衷心感谢编委专家对本研究及多轮修改工作的认可与支持！

编委 2 复审意见：

该综述精准把握了心理计量学方法学前沿，项目反应树框架从节点内单维到多维、从静态到动态、从确认性到探索性的演进脉络梳理清晰，对反应风格研究的方法论深化具有重要示范作用。这一理论、技术与应用整合的路径，与《心理科学进展》所倡导的前沿方法引介与本土问题关怀并重的取向高度契合，建议录用，并在修改中进一步以中国常用量表为例强化跨文化讨论的意义。

回应：衷心感谢编委专家对本研究的充分肯定与宝贵建议。我们完全认同《心理科学进展》所倡导的前沿方法引介与本土问题关怀并重的取向。在本次修改中，我们在原文章的基础上，于以下四处增添了以中国常用量表为例的跨文化讨论，以进一步强化本土问题关怀：

（1）引言第一段：在概述 RS 的跨文化视角时，补充了中国样本在相同量表上 ERS 更少、MRS 更多的实证证据(Guo et al., 2019; Harzing et al., 2012)，在文章开篇就将中国常用量表的跨文化差异问题作为 RS 研究的重要背景引入。

（2）实证数据分析 4.3.1 节：在交代数据来源时，除引用国际文献外，补充指出该量表也为国内 RS 研究(郭庆科 等, 2007)所关注，说明以大五人格问卷为演示示例对中国研究者具有直接的参考价值。

（3）讨论 5.1 节：在总结偏差控制与过程理解两种视角之后，针对中国本土量表的开发与验证，阐述了 IRTree 模型同时回应两类关切（偏差威胁与文化信息）的现实意义。

（4）讨论 5.2 节：在讨论参数等价性向跨群体情境的拓展时，专门以大五人格量表在中西文化群体中的施测为例，区分了方向决策节点上的 DIF（内容理解差异）与极端反应节点上的 DIF（ERS 使用差异），揭示了中国常用量表跨文化应用时 IRTree 框架下 DIF 检测的独特复杂性。

主编意见：

文章经过多轮评审和修改，已经达到《心理科学进展》的要求，建议录用。