

《心理科学进展》审稿意见与作者回应

题目：空间导航能力性别差异的三水平元分析

作者：薛笑然，崔伟，张丽

第一轮

审稿人 1 意见：

空间导航是认知心理学研究的重点领域，特别是在诺奖授予之后，更是得到了极大的关注。本研究选题具有很强的时效性，并分析了多种影响导航能力的因素，为该领域的进一步研究提供了很好的参考方向。在数据分析方法上，相较于传统的单水平元分析，研究采用三水平元分析，能够更细致地识别并区分不同来源的误差，使得研究结果更加可靠，能够更全面地反映空间导航能力的性别差异。进一步，研究考察了年龄、任务环境、测试场景和辅助装备等系列调节变量，为未来的测量任务设置提供了新的思路。文章写作详实，语言得体。推荐在修改后发表。

意见 1：发表偏倚。文章报道称“变量在 Begg 上的 p 值不显著($t = 0.07, p = 0.086$)”。 p 值已经边缘显著。请汇报 meta 各研究间异质性 I^2 和 Egger 检验的结果。如果 p 值显著，请分析去掉调节因子后的发表偏倚。

回应：感谢您的宝贵建议。根据您的意见，我们对发表偏倚和异质性进行了详细的补充分析，具体结果和解释如下：

(1) 发表偏倚的评估：我们首先进行了 Egger 偏倚检验。结果显示发表偏倚显著($t = 4.39, p < 0.001$)，且限制估计为 $b = 0.17, 95\% \text{ CI } [0.102, 0.246]$ 。这表明可能存在发表偏倚，进而影响效应量的总体估计。为进一步探讨偏倚来源，我们还去掉了调节因子，但 Egger 检验的结果保持不变，仍然显著，说明调节因子对发表偏倚影响较小，偏倚可能是系统性的。

为了更全面地评估发表偏倚，我们应用了 Duval 和 Tweedie 的 Trim-and-Fill 方法，通过该方法的调整以更准确地估计总体效应量。该方法估计存在 93 项可能缺失的研究，主要集中在漏斗图的左侧（效应量较小的一侧）。补充这些缺失的研究后，总体效应量从 0.39 降至 0.23 ($p < 0.001, 95\% \text{ CI } [0.180, 0.276]$)。尽管调整后效应量有所减小，但依然显著，表明性别在空间导航能力上的差异具有稳健性，支持了研究结果。综上，尽管可能存在一定的发表偏倚，但对空间导航能力与性别关系的总体效应影响是有限的，元分析结果依然具有稳健性。

(2) 异质性分析：为了进一步探讨异质性的来源，我们进行了异质性检验和方差分解。使用随机效应模型对异质性进行了分层分析，结果显示，总异质性 I^2 为 83.80%，主要来自研究间的异质性。进一步的方差分解显示，抽样方差占总方差的 18.66%，研究内方差占 11.94%，研究间方差占 69.40%，说明研究间的差异性对总体变异性的贡献最大。Q 检验结果显示异质性显著 ($Q(df = 464) = 2231.05, p < 0.001$)，支持使用随机效应模型的合理性。

我们在论文的“2.3.4 发表偏倚检验”部分对以上内容进行了相应的文字修改，并对“图 2 效应量 d 分布的漏斗图”做出相应修改。以进一步增强研究的透明性和稳健性。此外，我们在“4.3 研究意义、局限性与未来展望”部分提到了发表偏倚的潜在影响，即：尽管我们采用了 Trim-and-Fill 方法进行了发表偏倚的调整，但偏倚可能仍对效应量估计存在一定影响，这一点将作为本研究的一项局限性加以说明。

2.3.4: 漏斗图上代表效应量的点基本分布在平均效应量的两侧且位于顶端, 但 Begg 检验($t = 0.07, p = 0.086$)和 Egger 检验($t = 4.39, p < 0.001$)都表明可能存在一定程度的发表偏倚。为了进一步评估发表偏倚的影响, 采用了 Duval 和 Tweedie(2000)的 Trim-and-Fill 方法, 该方法估计可能存在 93 项缺失研究, 补充这些研究后效应量略微降低(从 0.39 降至 0.23), 但调整后效应量仍然显著($p < 0.001, 95\% \text{ CI } [0.180, 0.276]$), 即发表偏倚对空间导航能力与性别关系的总效应影响是有限的, 元分析结果依然具有稳健性(见图 2)。

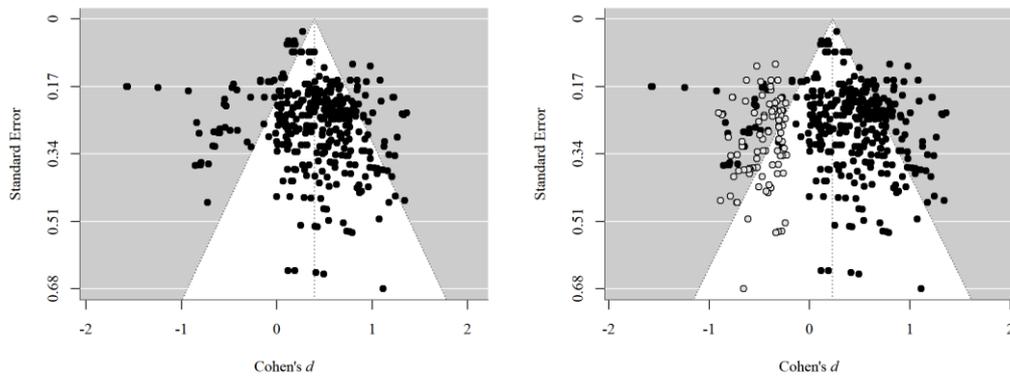


图 2 效应量 d 分布的漏斗图

注: 图中 x 轴为 Cohen's d 效应量, y 轴为每个效应量对应的标准误。左图为原始漏斗图, 表示未调整的效应量分布; 右图为经过 Trim-and-Fill 方法调整后的漏斗图, 其中空心点代表通过调整方法补充的潜在缺失研究。

4.3: 其次, 本研究使用了多种检验评估发表偏倚, 并使用 Duval 与 Tweedie(2000)的 Trim-and-Fill 方法来减少偏倚的影响, 但发表偏倚仍可能对效应量估计带来影响, 导致系统性高估。尽管调整后效应量仍然显著, 提示研究结论具有稳健性, 但未来研究有必要纳入未发表或难以获得的数据, 以进一步降低偏倚的影响。

参考文献:

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56(2), 455–463.

意见 2: 正文中采用随机效应模型对空间导航能力的性别差异进行了主效应检验。请阐述不使用固定效应的理由。

回应: 感谢您的宝贵意见。我们选择使用随机效应模型而非固定效应模型来分析空间导航能力的性别差异, 主要由于本研究存在显著的研究间异质性。固定效应模型假设各研究的效应量代表同一个总体效应, 适用于各研究之间不存在实质性异质性的情况。然而异质性检验表明, Q 检验和 I^2 指数 (83.80%) 均显示研究间存在显著的效应量变异性, 在这种情况下, 固定效应模型假设效应量一致将不适用, 而随机效应模型更能合理反映这些差异性。随机效应模型假设, 每个单独研究的效应量并非来源于一个完全相同的总体, 而是围绕一个总体效应的分布。这意味着个体研究的效应量可能因种种差异 (如被试特征、研究方法、测量工具、研究情境等) 而有所不同, 随机效应模型允许这些差异作为系统变异的一部分进入模型, 基于以上考虑, 本研究选择了随机效应模型进行主效应检验。

在使用 Begg 和 Egger 检验发现发表偏倚后, 我们使用了 Duval 和 Tweedie 的 Trim-and-Fill 方法进行了调整, 调整后的效应量 (从 0.39 降至 0.23) 仍然显著 ($p < 0.001, 95\% \text{ CI } [0.180, 0.276]$), 表明性别在空间导航能力上的差异具有稳健性。这一结果表明随机效应模型在存在偏倚和异质性的情况下依然提供了稳健的效应量估计, 证实了本研究使用随机效应模型的合理性。

意见 3: 空间导航的神经研究是目前热点。包括位置细胞、网格细胞在定位中的作用，以及它们的序列放电 replay 如何支持路线规划等。请在讨论中加入讨论，指出本文对相关神经科学研究的启发和联系。

回应: 感谢您的宝贵建议。根据您的建议，我们在讨论部分的“4.1 空间导航能力的性别差异”中补充了关于空间导航系统的神经机制，特别是海马、内嗅皮层中神经元的相位进动对于个体导航过程的作用。此外，我们还引入了 Long 和 Zhang (2021) 的最新成果，该研究在躯体感觉皮层中发现了并行导航系统，说明大脑会通过多个区域应对复杂导航任务。通过上述补充，更全面地探讨了空间导航的神经机制，希望能够为该领域的后续研究提供一定启发。

4.1: 除了进化学中两性劳动分工可能带来的影响，神经生物学的研究也揭示了男女性在大脑结构和功能上的差异可能造成空间导航能力的差异。大脑的不同区域存在多种并行的导航系统，用于应对复杂的空间环境和任务。海马-内嗅皮层导航系统是大脑中核心的空间导航环路(王琳, 王亮, 2017)，海马和内嗅皮层的神经元能够通过相位进动(phase precession)表征空间位置，并参与非空间任务的信息编码。其中，海马作为核心脑区，整合和协调序列信息，其内部的位置细胞会在个体处于特定位置时激活，以形成环境的认知地图(O'Keefe, 1976; O'Keefe & Dostrovsky, 1971; Marchette et al., 2011)。内嗅皮层中的网格细胞提供空间距离的度量，以六边形网格形式激活，为空间定位提供精确支持(Hafting et al., 2005)；头方向细胞则作为“内部指南针”，在运动过程中提供方向信息(Taube et al., 1990; Long et al., 2022, 2024; Zhang, 1996)。除了经典的海马-内嗅皮层导航系统，Long 和 Zhang(2021)还在大脑的躯体感觉皮层中发现了一套全新的空间导航系统。该系统中也存在位置细胞、头方向性细胞、边界细胞、网格细胞和联合细胞等类型的神经元，与海马和内嗅皮层中表征空间信息的经典空间定位细胞相似，并具有类似的放电特性。目前关于不同性别个体在导航过程中的认知神经差异主要以动物为被试，而一项人类实验通过 fMRI 发现，男女性在完成空间导航任务时均大面积使用大脑，但男性比女性更多依赖海马体。进一步地，研究向女性注射睾酮素以增强其海马体激活，结果发现女性对迷宫布局的理解有所加深，也表明睾酮素水平与空间导航的性别差异密切相关 (Pintzka et al., 2016)。这一结论为探索性别因素如何影响大脑空间导航能力提供了重要的神经生物学依据。未来研究有必要进一步利用神经成像技术，深入探讨不同空间导航系统在男女性中的功能差异，从而加深对大脑如何通过多种并行系统应对复杂导航任务的理解，为探究空间导航能力的性别差异提供新的视角。

参考文献:

Long, X., & Zhang, S. J. (2021). A novel somatosensory spatial navigation system outside the hippocampal formation. *Cell Research*, 31(6), 649–663.

意见 4: 引言中“它是一种相对复杂、较为综合的空间能力，涉及多种认知能力，包括空间定位、空间记忆、空间感知能力等”。有重复嫌疑，建议修改为“它是一种相对复杂、较为综合的空间能力，涉及多种认知能力，包括情景记忆、决策、感知觉等”。

回应: 感谢您的建议。我们已根据您的意见，将引言第一段中的相关表述修改为“它是一种相对复杂、较为综合的空间能力，涉及多种认知能力，包括情景记忆、决策、感知觉等”。这一修改减少了表述上的重复，使内容更加凝练准确。

意见 5: 讨论中“有研究通过 fMRI 发现男女性在完成空间导航任务时，均大面积地使用大脑，但男性比女性更多地使用海马体。”请附上参考文献。

回应: 感谢您的建议。关于“男女性在完成空间导航任务时均大面积使用大脑，但男性比女性更多地使用海马体”的描述，与后一句“向女性注射睾酮素以增强其海马体的激活后，女性对迷宫的布局有了更深入的了解”确实引用自同一篇文献 (Pintzka et al., 2016)，由于我们

的表述不够清晰，导致了这一误解。我们已在“4.1 空间导航能力的性别差异”部分对这段表述进行了优化，使前后两句的逻辑关系更为清晰，避免引起读者误解。

4.1: 而一项人类实验通过 fMRI 发现, 男女性在完成空间导航任务时均大面积使用大脑, 但男性比女性更多依赖海马体。进一步地, 研究向女性注射睾酮素以增强其海马体激活, 结果发现女性对迷宫布局的理解有所加深, 也表明睾酮素水平与空间导航的性别差异密切相关 (Pintzka et al., 2016)。这一结论为探索性别因素如何影响大脑空间导航能力提供了重要的神经生物学依据。

意见 6: 讨论中“Coutrot 等人(2022)的大规模研究证实了这一点, 发现城市以外地区长大的个体由于活动范围更广, 因此展现出了更强的空间导航能力。”文章关键的发现是环境结构的复杂度影响空间导航能力。请扩展对这篇文章的讨论。

回应:感谢您的宝贵意见, 使我们对环境因素的讨论更加深入和全面。基于 Coutrot 等人(2022)的研究, 成长环境结构的复杂度(而不仅仅是城乡间的简单区别)会影响个体的空间导航能力。我们在文章中补充了对成长环境结构复杂度如何影响空间导航能力的详细讨论, 因此我们在讨论部分进一步分析了复杂环境如何通过提供更丰富的空间刺激影响空间导航能力的发展, 并进一步指出男性与女性在成长过程中对环境的不同利用方式可能导致了性别差异的产生。我们已在“4.2 空间导航能力性别差异的调节因素”部分修改了相关表述, 以更全面地阐述成长环境的结构复杂度在空间导航能力性别差异中的潜在作用。

4.2: Coutrot 等人(2022)的一项全球性大规模研究发现, 成长环境的结构复杂度显著影响个体的空间导航能力。成长于乡村等复杂环境的人群比城市居民更擅长导航, 这一现象在现代化国家中尤为明显。这可能是因为乡村地区的非规则地形和复杂布局为个体提供了更多的空间挑战, 使他们在应对复杂环境时发展出更强的空间导航能力。相比之下, 现代化城市多采用简单的网格状结构(如芝加哥的街道布局), 个体无法获得丰富的寻路经历, 因此空间导航能力发展受限。这种成长环境中的复杂度不仅直接影响个体的空间导航能力, 还可能与性别共同作用, 影响男女性在空间导航能力上的发展。男女性在成长过程中的活动范围往往存在差异, 男性往往更有可能被家长允许外出游戏, 且更倾向于开展具有探索性的活动, 因此在童年期具有更高的外出频次、更远的出行距离和探索广度, 更有可能接触到复杂的空间环境(Schug, 2016a, 2016b), 这种探索经验使男性在成长过程中积累了丰富的空间线索, 有助于空间导航能力的发展。相对而言, 女性在成长过程中接触复杂空间环境的机会较少, 活动范围更为受限, 因此缺乏相应的探索机会。这些差异凸显了生活环境在塑造空间导航能力性别差异中的潜在作用。

.....

审稿人 2 意见:

作者采用元分析方法, 系统分析了 173 篇空间导航相关论文的数据, 旨在探究空间导航能力的性别差异及其调节因素。根据数据分析结果, 作者认为, 他们的研究确认了男性在空间导航能力上的优势, 并识别了影响性别差异的重要调节变量。整体而言, 研究设计合理, 数据分析和结果解读科学合理, 论文撰写规范。以下是对论文内容的具体建议。

意见 1: 作者根据分析结果得出结论: “确认了空间导航能力的男性优势现象”、“证实了男性的空间导航能力优于女性”。然后, 这一结论与“发现了影响性别差异程度的调节变量”可能存在逻辑上的冲突。如果存在显著的调节变量, 可能意味着在特定条件下性别差异并不显著。建议进一步明确在哪些条件下性别差异显著, 哪些条件下不显著, 以更精确地界定空间导航能力的性别差异。

回应：感谢您对本研究的深入建议，这一建议对于我们完善文章结构具有重要意义。正如您所言，调节变量确实表明在某些特定条件下性别差异并不显著，因此，我们根据您的建议，分别在摘要、结果、讨论部分进一步明确了空间导航能力的性别差异在不同条件下的显著情况。具体而言，在摘要部分，我们将表述调整为：“结果表明，在大多数条件下，空间导航能力存在性别差异，男性的空间导航能力强于女性，且二者关系受年龄、表征方式、时间限制、任务环境、测试场景、辅助装备的调节，在婴幼儿期（0至4岁）和成年晚期（65岁以上）的人群中，和室内-室外双重测试条件和个体不使用任何辅助设备参与测试的条件下，空间导航能力的性别差异不显著。”通过调整表述，进一步明确了空间导航能力与调节因素的交互作用，强调了空间导航能力的性别差异并不是普遍且固定的，而是依赖于特定的任务情境和个体特征。在结果部分，修改了“3.2 空间导航能力性别差异的调节效应检验”的表述，明确阐述了在哪些调节变量的何种水平下，空间导航能力不存在性别差异。在讨论部分第一段，将表述调整为：“总体而言，空间导航能力存在性别差异，男性的空间导航能力更强，然而在婴幼儿期（0至4岁）和成年晚期（65岁以上）的人群中，未发现显著的性别差异。在室内-室外双重测试条件和个体不使用任何辅助设备参与测试的条件下，也未观察到性别差异。”并在“4.1 空间导航能力的性别差异”部分的开头明确指出了空间导航能力的性别差异在何种情况下存在，以及在哪些特定条件下并不显著。上述所有修改的目的都是更精确地界定空间导航能力的性别差异，确保对性别差异的描述与调节变量的影响相一致。

3.2: 空间导航能力性别差异的调节效应检验结果如表1所示。结果表明：(1)年龄具有显著的调节效应， $F(6, 361) = 2.91, p = 0.009$ ，在4至65岁的群体及跨年龄研究中，空间导航能力的性别差异显著存在，男性优势明显。在婴幼儿期(0至4岁)和成年晚期(65岁以上)，未发现显著的性别差异；(2)表征方式具有边缘显著的调节效应， $F(1, 370) = 3.44, p = 0.064$ ，与自我中心的表征方式相比，男女性被要求使用环境中心的表征方式时的空间导航能力差异更大；(3)时间限制具有边缘显著的调节效应： $F(1, 370) = 3.42, p = 0.065$ ，与不限时完成空间导航任务相比，男女性在有时间限制的空间导航任务中表现出更大的能力差异；(4)任务环境具有显著的调节效应， $F(3, 368) = 3.28, p = 0.022$ ，与室内测试相比，男女性在完成水迷宫任务时表现出更大的空间导航能力差异。而在室内-室外双重测试条件下，未发现性别差异。(5)测试场景具有显著的调节效应， $F(1, 370) = 9.74, p = 0.002$ ，与真实场景相比，男女性在视频场景中的空间导航能力差异更大；(6)辅助装备具有显著的调节效应， $F(2, 369) = 7.41, p < 0.001$ ，男女性在使用电子设备、纸笔工具时表现出性别差异，而不使用辅助设备时则无性别差异。而在以下变量中并没有发现显著的调节效应：(1)地区： $F(6, 365) = 0.53, p = 0.786$ ；(2)任务类型： $F(7, 364) = 1.17, p = 0.322$ ；(3)测量指标： $F(5, 366) = 1.92, p = 0.091$ 。

4: 总体而言，空间导航能力存在性别差异，男性的空间导航能力更强，然而在婴幼儿期(0至4岁)和成年晚期(65岁以上)的人群中，未发现显著的性别差异。在室内-室外双重测试条件和个体不使用任何辅助设备参与测试的条件下，也未观察到性别差异。

4.1: 本研究结果首先证实了空间导航能力存在性别差异，男性的空间导航能力优于女性，这一现象在4至65岁个体中普遍存在，且在大多数测试条件下具有显著体现。

意见2：作者在论文中特别强调了“三水平元分析方法”的使用，并在标题和引言部分提及。如果这是论文的创新点，建议在引言部分简要介绍该方法的独特性和优势。目前方法部分的介绍尚显不足，未能充分展现该方法的创新和必要之处。此外，建议作者通过合适的方式公开共享元分析用到的代码。

回应：感谢您的建议。我们在方法部分的“2.3.1 三水平元分析的效应量计算”补充了对三水平元分析方法的详细介绍，说明了该方法在处理复杂数据结构、解决效应量不独立问题以及提升统计效率和稳健性方面的优势，也详细阐述了该方法如何通过分解方差来源来控制效应

量之间的依赖性，从而更好地展示三水平元分析在本研究中的必要性和应用价值。同时，感谢您关于代码共享的建议。我们理解数据和代码共享对于研究透明性的重要性，后续将通过合适的代码共享平台公开本元分析所用的代码，以便其他研究者能够复现和验证分析结果，促进相关领域的进一步研究。

2.3.1: 本研究选取标准化均数差(standardized mean difference, SMD)Cohen's d 作为效应量。由于空间导航能力的测量方法和指标种类颇多，许多文献包含多个效应量，导致效应量之间存在依赖性，与传统元分析方法认为效应量相互独立的假设不符，故采用三水平元分析方法，通过补充研究内方差来考虑效应量之间的依赖性(Cheung, 2014)，将方差来源分解为抽样方差(水平 1)、研究内方差(水平 2)、研究间方差(水平 3)(Hox et al., 2017)。相较于传统方法，三水平元分析更适合复杂数据结构的处理，不仅解决了效应量不独立的问题，还在保留原始文献信息完整性的同时，有效提升了统计效率和稳健性(Cheung, 2019)。

意见 3: 建议在正文中明确介绍所使用的检索数据库，而不仅仅是在图表中展示。这将有助于读者了解研究的全面性和系统性。

回应: 感谢您的宝贵建议。我们已在“2.1 文献检索与筛选”部分补充了所使用的检索数据库，以便读者更清晰地了解研究的全面性和系统性。具体来说，我们通过多个数据库进行检索，包括中文数据库：中国知网 (CNKI)、万方、维普；英文数据库：Web of Science、PubMed、EBSCO、PsyArXiv，并结合手动检索以补充遗漏的相关文献。

2.1: 通过多个数据库中进行检索，以确保文献的全面性和系统性，所用中文数据库包括：中国知网(CNKI)、万方、维普，英文数据库包括：Web of Science、PubMed、EBSCO、PsyArXiv，并结合手动检索以补充遗漏的相关文献。

意见 4: 鉴于作者广泛调研了大量的文献并进行了复杂的编码工作，为了提高论文的可重复性和数据的可复用性，建议将编码结果和提取的效应量已公开可获取的在线文档形式共享，方便读者查阅和引用。

回应: 感谢您的建议。我们已在投稿后将编码结果和提取的效应量数据发送至编辑部邮箱，供读者查阅和引用。如有进一步需求，我们也非常愿意配合提供更多信息。

意见 5: 在 2.3.4 部分，报告的统计值“ $t = 0.07, p = 0.086$ ”可能存在问题，建议进一步核查。同时，建议仔细检查论文中报告的所有统计值，确保其准确性和可靠性。

回应: 感谢您的宝贵意见。根据您的建议，我们对“ $t = 0.07, p = 0.086$ ”这一统计值进行了核查，确认确实存在发表偏倚的可能性。为进一步评估偏倚的影响，我们在核实偏倚的基础上补充了其他分析。

具体来说，我们首先进行了 Egger 偏倚检验。结果显示发表偏倚显著($t = 4.39, p < 0.001$)，且限制估计为 $b = 0.174, 95\% \text{ CI } [0.102, 0.246]$ 。这表明可能存在发表偏倚，进而影响效应量的总体估计。为进一步探讨偏倚来源，我们还去掉了调节因子，但 Egger 检验的结果保持不变，仍然显著，说明调节因子对发表偏倚影响较小，偏倚可能是系统性的。为了更全面地评估发表偏倚，我们应用了 Duval 和 Tweedie 的 Trim-and-Fill 方法，通过该方法的调整以更准确地估计总体效应量。该方法估计存在 93 项可能缺失的研究，主要集中在漏斗图的左侧（效应量较小的一侧）。补充这些缺失的研究后，总体效应量从 0.39 降至 0.23 ($p < 0.001, 95\% \text{ CI } [0.180, 0.276]$)。尽管调整后效应量有所减小，但依然显著，表明性别在空间导航能力上的差异具有稳健性，支持了研究结果。综上，尽管可能存在一定的发表偏倚，但对空间导航能力与性别关系的总体效应影响是有限的，元分析结果依然具有稳健性。我们在论文的“2.3.4 发表偏倚检验”部分对以上内容进行了相应的文字修改，并对“图 2 效应量 d 分布的

漏斗图”做出相应修改。以进一步增强研究的透明性和稳健性。此外，我们在“4.3 研究意义、局限性与未来展望”部分提到了发表偏倚的潜在影响，即：尽管采用了 Trim-and-Fill 方法进行了发表偏倚的调整，但偏倚可能仍对效应量估计存在一定影响，这一点将作为本研究的一项局限性加以说明。

同时，为确保研究的严谨性，我们对论文中其他统计结果进行了全面检查，所有数据均已确认准确可靠。

2.3.4: 漏斗图上代表效应量的点基本分布在平均效应量的两侧且位于顶端，但 Begg 检验($t = 0.07, p = 0.086$)和 Egger 检验($t = 4.39, p < 0.001$)都表明可能存在一定程度的发表偏倚。为了进一步评估发表偏倚的影响，采用了 Duval 和 Tweedie(2000)的 Trim-and-Fill 方法，该方法估计可能存在 93 项缺失研究，补充这些研究后效应量略微降低(从 0.39 降至 0.23)，但调整后效应量仍然显著($p < 0.001, 95\% \text{ CI } [0.180, 0.276]$)，即发表偏倚对空间导航能力与性别关系的总效应影响是有限的，元分析结果依然具有稳健性(见图 2)。

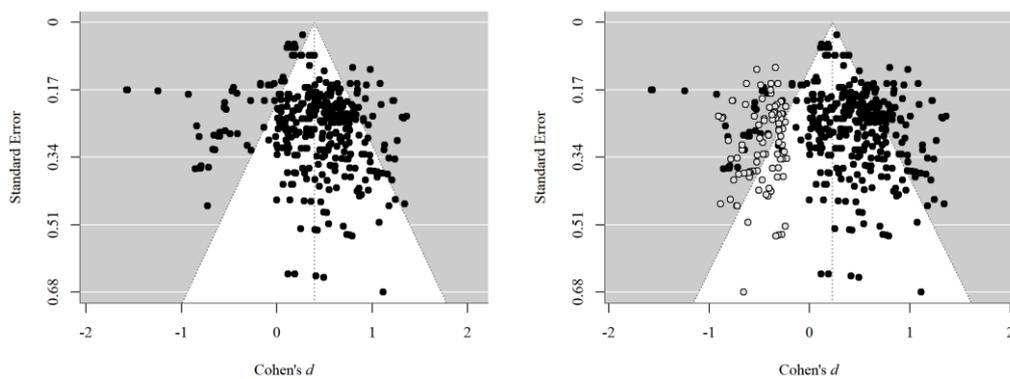


图 2 效应量 d 分布的漏斗图

注：图中 x 轴为 Cohen's d 效应量，y 轴为每个效应量对应的标准误。左图为原始漏斗图，表示未调整的效应量分布；右图为经过 Trim-and-Fill 方法调整后的漏斗图，其中空心点代表通过调整方法补充的潜在缺失研究。

4.3: 其次，本研究使用了多种检验评估发表偏倚，并使用 Duval 与 Tweedie(2000)的 Trim-and-Fill 方法来减少偏倚的影响，但发表偏倚仍可能对效应量估计带来影响，导致系统性高估。尽管调整后效应量仍然显著，提示研究结论具有稳健性，但未来研究有必要纳入未发表或难以获得的数据，以进一步降低偏倚的影响。

意见 6: 在 3.2 部分，作者提到“结果显示至少有一个调节变量的回归系数显著偏离零”。建议进一步详细阐述这部分的结果和发现，为读者提供更丰富的信息。

回应：感谢您的建议。多重回归的目的是，在识别出显著的调节变量后，进一步确认这些变量的独立影响。这一步可以控制调节变量之间的共线性问题，从而提供更精确的效应估计。这一过程确保每个显著调节变量对因变量的影响是独立的，而非其他变量的共同作用。我们在“3.3 调节变量的多重回归分析”部分补充了对结果的说明：“表明这些变量在控制其他调节变量的共线性影响后，对因变量产生了独立的显著影响，进一步验证了其在空间导航能力中的独立调节作用。”，以便为读者提供更丰富的信息。

3.3: 为排除调节变量之间的共线性，根据 Assink 和 Wibbelink(2016)的方法，对显著的调节变量进行了多重回归分析。以年龄(婴幼儿期, 0 至 4 岁)、任务环境(室内)、表征方式(环境中心)、时间限制(不限时)、测试场景(真实场景)、辅助装备(电子设备)为参照变量，结果显示至少有一个调节变量的回归系数显著偏离零(见表 2)，表明这些变量在控制其他调节变量的共线性影响后，对因变量产生了独立的显著影响，进一步验证了其在空间导航能力中的

独立调节作用。

第二轮

审稿人 1 意见:

作者已经回答了我的问题。我没有其他意见了。

回应: 感谢外审专家对我们稿件的细致审阅和宝贵意见, 这对于文章的完善具有重要意义。如果您还有其他建议或意见, 我们将随时改进。再次感谢您的支持和指导!

审稿人 2 意见:

意见 1: 作者在回复信中详细阐释了使用随机效应模型的理由, 建议在正文中如 3.1 部分加入简单的阐述, 以使得读者对统计方法的选择更加明确。

回应: 感谢您的建议。我们在“3.1 空间导航能力性别差异的主效应检验”部分加入了使用随机效应模型的理由, 以更清晰地阐述主效应检验所使用统计方法的依据。希望这一修改能够更好地回应您的建议并提升文章的可读性。

3.1: 随机效应模型假设, 每个单独研究的效应量并非来源于一个完全相同的总体, 而是围绕一个总体效应的分布, 这意味着个体研究的效应量可能因种种差异(如被试特征、研究方法、测量工具、研究情境等)而有所不同, 而随机效应模型允许这些差异作为系统变异的一部分进入模型。因此, 本研究选择了随机效应模型进行主效应检验。

意见 2: 关于空间导航的性别差异研究, 原文增加表述为“目前关于不同性别个体在导航过程中的认知神经差异主要以动物为被试”, 但人类的研究目前也较多, 原文中只补充了一篇性别差异的神经机制的研究, 请进一步进行补充和阐释, 以下是一些参考示例文献, 如:

Grön, G., Wunderlich, A. P., Spitzer, M., Tomczak, R., & Riepe, M. W. (2000). Brain activation during human navigation: gender-different neural networks as substrate of performance. *Nature neuroscience*, 3(4), 404-408.

Kong, X. Z., Huang, Y., Hao, X., Hu, S., & Liu, J. (2017). Sex-linked association between cortical scene selectivity and navigational ability. *Neuroimage*, 158, 397-405.

Noachtar, I., Harris, T. A., Hidalgo-Lopez, E., & Pletzer, B. (2022). Sex and strategy effects on brain activation during a 3D-navigation task. *Communications biology*, 5(1), 234.

回应: 感谢您的建设性建议。针对您提出的关于补充人类研究内容的建议, 我们已在“4.1 空间导航能力的性别差异”中补充了有关人类空间导航能力性别差异的神经机制研究。希望这些修改能够提升文章的学术质量。

4.1: 关于不同性别个体在导航过程中的认知神经差异, 研究发现, 男女性在完成空间导航任务时均大面积使用大脑, 但男性比女性更多依赖海马体(Grön et al., 2000; Kong et al., 2017)。Noachtar 等人(2022)的研究表明, 男性和女性的脑区激活模式与导航策略显著相关。男性在环境中心的导航中表现出更多的顶叶和前额叶激活, 而女性在自我中心的导航中表现出更强的后部脑区激活。睾酮素也是影响空间导航能力性别差异的可能因素。有研究发现, 内源性睾酮素水平与女性的空间导航能力呈显著正相关, 而在男性中未观察到睾酮水平与导航能力的显著相关性(Burkitt et al., 2007)。进一步地, 研究通过向女性注射睾酮素, 发现与成功导航相关的内侧颞叶活性显著增强, 该组女性被试对迷宫布局的理解更深入, 并且在方

向测试中的表现更优。这些结果表明，睾酮素可能通过调节 MTL 的活动，影响空间导航能力的性别差异(Pintzka et al., 2016)。上述结论为探索性别因素如何影响大脑空间导航能力提供了重要的神经生物学依据。

意见 3: Pintzka et al. (2016)的研究以女性为被试，无法说明“男性比女性更多地使用海马体”，请正确引用文献，如上述第一个文献: Grön, G., Wunderlich, A. P., Spitzer, M., Tomczak, R., & Riepe, M. W. (2000). Brain activation during human navigation: gender-different neural networks as substrate of performance. *Nature neuroscience*, 3(4), 404-408.

回应: 感谢您的建议。我们认真阅读了 Pintzka et al. (2016)的文献，确实不能得出该结论。我们认真查阅了您推荐阅读的研究，并在“4.1 空间导航能力的性别差异”中明确注明引用了 Grön et al. (2000) 和 Kong et al. (2017) 的研究，这两项研究均提供了男性在空间导航任务中更多依赖海马体的直接证据。同时，我们删除了对 Pintzka et al. (2016) 的不准确引用。再次感谢您对我们研究的细致审阅和建设性意见，这对提高稿件的严谨性和科学性有重要帮助。

4.1: 关于不同性别个体在导航过程中的认知神经差异，研究发现，男女性在完成空间导航任务时均大面积使用大脑，但男性比女性更多依赖海马体(Grön et al., 2000; Kong et al., 2017)。

意见 4: 文章在讨论中提到社会文化的影响，可以酌情适当引用这两篇文献的内容:

Bartlett, K. A., & Camba, J. D. (2023). Gender differences in spatial ability: A critical review. *Educational Psychology Review*, 35(1), 8.

Hults, C. M., Francis, R. C., Clint, E. K., Smith, W., Sober, E. R., Garland Jr, T., & Rhodes, J. S. (2024). Still little evidence sex differences in spatial navigation are evolutionary adaptations. *Royal Society open science*, 11(1), 231532

回应: 感谢您的建议。我们认真阅读了上述研究，并在“4.1 空间导航能力的性别差异”中引用了上述研究，从而更好地从社会文化的角度对空间导航能力的性别差异进行解释，希望所做修改能够提升文章的学术质量。

4.1: 此外，一些传统的空间能力测试设计可能在无意间放大性别差异，强化了被试对性别空间刻板印象的认同。这种刻板印象在教育和社会互动中不断被重复和强化，进一步塑造并加剧了性别分化，而这种性别分化反过来又作为新的预测性结果对社会文化施加影响，形成一种恶性循环(Bartlett & Camba, 2023)。.....而 Hults 等人(2024)综述了多项研究，指出在那些不存在“男主外，女主内”分工模式的狩猎采集文化中，当男性和女性从事类似的空间任务时，性别差异几乎消失(Jang et al., 2019; Trumble et al., 2016)，印证了上述设想，即性别差异并非普遍存在，而是高度依赖于社会化和文化背景。

第三轮

编委 1 意见:

同意发表，不过建议作者对文字进行精简，现在前言与讨论很拖沓。

回应: 感谢您的建议。目前的引言和讨论部分确实存在语言冗杂、重点不够清晰的问题，已根据您的建议精简了引言和讨论部分的表述，使语言更加简洁明了，突出研究的核心内容和关键结论。具体而言，在引言部分，删除了与文章主题关联性较弱的文献综述，强化了研究问题的聚焦；在讨论部分，对重复、冗杂的论述进行了语言上的精简凝练，突出了对研究结果的解释。希望通过这些修改，使文章的逻辑更加紧凑，核心观点更加突出。再次感谢您的

宝贵建议！

编委 2 意见：同意发表。

主编意见：根据编委和审稿专家的意见，建议发表。