《心理学报》审稿意见与作者回应

题目：基于 P300 的隐藏信息测试探测虚假身份信息
作者：成嘉予，王皓然，冯旺，傅根跃，赛李阳

第一轮

审稿人 1 意见：

该论文选题具有一定的新颖性，作者创新性地将测谎研究的关注点从传统的“否认型谎言”转向“虚构型虚假信息”的检测，并采用基于 P300 的 CIT（Concealed Information Test）范式对虚假信息进行检验。研究结果显示，虚假身份信息与真实身份信息在记忆编码强度及自我相关性上存在差异，ROC 分析结果表明，组间鉴别率能够较好地区分虚假组与真实组。这对于发展识别使用虚假身份的不法分子（如间谍）的测谎技术具有应用价值，同时拓展了 CIT 技术在司法与安全领域的适用范围，并在一定程度上为司法测谎与安全审查提供了实证依据。此外，文章还采用贝叶斯因子对假设检验进行了补充，这一做法值得肯定，有助于从多角度提升结果的可靠性。尽管如此，论文仍存在以下问题：

回应：感谢审稿专家对本研究选题创新性、方法设计及结果可靠性的积极评价和鼓励。这将有力推动我们进一步完善论文并开展后续研究。

意见 1：关于摘要表述。摘要中写道：“（2）个体鉴别分析显示 P300 在区分真实与虚假身份信息时表现出较高的鉴别效能（AUC 实验 1 = 0.82；AUC 实验 2 = 0.86）”。此处应明确区分“组间鉴别率”与“个体鉴别率”。前者对应 ROC 结果，而后者应基于 bootstrap 的个体分析。建议将摘要表述更正，以避免读者产生混淆。

回应：非常感谢您对摘要表述提出的宝贵建议。对于您指出的“组间鉴别率”与“个体鉴别率”需明确区分的问题，我们完全认同，并已在摘要中对相关内容进行了修订，以确保读者能够清晰理解两类指标的不同来源与含义。具体修改如下：

“P300 在区分真实与虚假身份信息方面具有较高的鉴别效能。实验 1 与实验 2 的 ROC 分析显示，其组间效率（AUC）分别为 0.82 和 0.86；在实验 3 中，间谍组基于 bootstrap 的个体鉴别率达到 90%。”

意见 2：关于实验操纵的精细化。在实验设计中，虚假组被试在实验前三天收到了包含姓名、年龄、生日、专业和籍贯等的虚假身份信息，但作者仅在实验当天通过填写虚假身份问卷来检验其掌握情况，并未进行更系统的操纵性检验（如每日线上反馈或记忆强化训练）。这种操作可能导致天花板效应及较大个体差异：部分被试可能充分内化了虚假身份，而另一些仅停留在表面记忆层面。这种差异可能解释了 bootstrap 个体鉴别率低于 ROC 组间鉴别率的结果，也体现在不同被试间 CIT 效应（Probe-irrelevant 差异）波动较大。未来研究可进一步加强操纵检验，以提升生态效度。

回应：非常感谢专家对实验操纵环节提出的深刻见解与建设性建议。我们在研究设计中确实考虑了虚假身份学习的操纵因素，并采取了一定的控制措施。具体而言，在实验开始前三天，我们向“间谍组”被试提供包含姓名、年龄、生日、专业与籍贯等信息的虚假身份资料，并要求其在之后三天内尽量在日常生活中使用该身份。实验当天，无提示地要求被试根据记忆完整默写其虚假身份信息。凡在默写过程中出现明显错误、较长停顿或查阅辅助资料的被试，

均判定为未达到操纵要求并停止实验。在三个实验中，共有 4 名“间谍组”被试因未能完成该环节而被剔除。

尽管上述程序在一定程度上确保了被试对虚假身份信息的掌握，但我们也承认，被试在虚假身份的内化深度与角色融入水平方面仍可能存在差异。本研究尚未建立系统的量化指标来评估这一差异。正如专家所指出，这种潜在个体差异可能在一定程度上导致了 bootstrap 个体鉴别率低于 ROC 组间鉴别率，并造成 CIT 效应（Probe–Irrelevant 差异）的波动。
我们高度认同专家提出的改进方向，并已在修订稿的总讨论部分补充相关说明。具体增加如下：

“虽然 P300 波幅在虚假身份信息识别中具有一定效用，但本研究中被试仅对虚假身份进行了三天的学习和识记……被试在身份内化和角色融入程度上可能存在显著个体差异。此类差异可能导致部分被试停留在表层记忆阶段，而另一些由于较强的角色代入充分内化了虚假身份，从而影响 CIT 效应的稳定性，也可能部分解释 ROC 组间与 bootstrap 个体鉴别率的差异。未来研究应考虑引入更系统化、分阶段的操纵性检验流程，如每日线上反馈、记忆强化训练或持续性角色沉浸任务 (Monaro et al., 2017; Slater & Sanchez-Vives, 2016)，以动态评估和强化虚假身份的学习与内化程度，从而提高操纵一致性与生态效度。”

意见 3：关于数据处理与电极参考。作者在在线数据采集时采用右侧乳突为参考，但在离线分析中似乎未进行重参考处理。多数相关研究（如 Rosenfeld 等人）常在离线阶段使用双侧乳突重参考，以增强研究间可比性。建议作者说明未进行重参考的理由。此外，在 BrainVision Analyzer 2.1 的 Raw Data Inspection 中，作者表述为“删除伪迹率大于 20%的通道”。此处措辞可能存在歧义，更合理的理解应是“将伪迹率大于 20%的通道排除在伪迹处理之外，以保留更多数据”，而非完全删除通道。建议更正表述。
回应：感谢审稿专家的宝贵建议。我们同意参考电极的选择可能会影响 ERP 幅值及统计效应量，从而影响与既往研究的可比性。我们在数据采集阶段采用右侧乳突在线参考。在离线分析中，我们继续延续该参考设定，主要基于以下考虑：（1）避免引入额外噪声并保持信噪比（SNR）：双侧乳突平均参考（linked/average mastoids）会将两个乳突通道的噪声共同引入参考项；若其中一侧乳突通道存在较多肌电/接触阻抗波动，平均参考可能会降低目标通道的稳定性，进而影响 P300 的估计精度与后续分类性能 (Rosenfeld, 2000)。（2）乳突并非“电静默点”，可能导致效应被“部分抵消”：P300 具有较广的头皮电场分布，乳突附近同样可能记录到一定的电位变化。将双侧乳突平均后作为参考，等效于在所有通道上减去这部分成分，从而使 P300 幅值整体变小，效应量与分类判别的距离指标可能随之略有下降 (Luck, 2014; Yao, 2001)。我们在复核分析中观察到的“效应变小、鉴别效率下降”与这一机制是一致的。

为回应审稿专家的建议并增强研究间可比性，我们新增了双侧乳突重参考的对比分析（详细可见补充材料）。我们将数据重参考为双耳乳突平均后，在不改变任何其他预处理步骤与统计/分类流程的前提下，完整重复了所有关键分析。结果显示两种参考方式的主要结果/结论基本一致，主要区别在于与单耳乳突参考相比，双耳乳图平均后的总体波幅下降，其主要效应的效应量减小，测谎鉴别率略有降低。

鉴于研究结论在不同参考策略下保持一致，而单耳参考在我们的数据中表现出略优的信噪比与鉴别效率。我们决定正文仍报告单耳乳突参考下的主分析结果，同时将双耳乳突重参考的全套复核结果作为补充材料提供，以便读者与后续研究进行横向对照与复现。在正文部分我们也进行了相应的说明和补充：

“EEG 采集时以右侧乳突为在线参考，离线分析中保持了右侧乳突参考设置(Luck, 2014; Rosenfeld, 2000; Yao, 2001)。同时，为检验结果的稳健性和可重复性，我们进一步采用双侧

乳突平均为参考对数据进行了复制分析，主要结果基本没有变化（详见附录 5 补充材料）。”

关于 Raw Data Inspection 中伪迹处理的表述： 感谢审稿专家指出措辞的歧义。我们已根据建议将原文修改为：

“将伪迹率大于 20%的通道排除在伪迹处理之外，以保留更多数据。”

这一表述更准确地反映了实际操作流程，也避免了将其误解为完全删除通道。 再次感谢审稿专家的严谨意见，这对提升稿件的清晰度与方法规范性具有重要价值。

意见 4：关于术语表述的统一性。在 ROC 分析中，作者将“犯罪组”作为状态变量，但在方法部分又表述为“间谍组”与“无辜组”。建议统一用语，以保持全篇的一致性。

回应：非常感谢审稿专家对术语表述一致性的细致指正。在修订稿中，我们已将相关内容统一为”间谍组“与“无辜组”，以确保全篇术语一致、表述规范。 再次感谢审稿专家的宝贵建议。

意见 5：关于个体分类标准的逻辑。论文中判断被试是否为“间谍”的标准为：“Bootstrap 抽样后无辜组真实姓名与间谍组虚假姓名的平均 P300 波幅差异减去无关刺激的差值（DMPI）”。该表述不够清晰，容易引发误解。根据作者的逻辑，作者指的是 bootstrap 抽样后，无辜组被试的 CIT 效应的均值与间谍组的 CIT 效应的均值的差值？作者认为真实身份信息会引发更强的 P300，而无关刺激在两组间差异不大，因此差值应为正值。如果某一被试在 1000 次抽样中有 800/900 次探测刺激–无关刺激差值小于该阈值，则判定为间谍。然而，该阈值的设定目前主要依赖经验性推理，缺乏严谨的数理依据。建议作者对其逻辑进行更详细的阐释，以增强方法论的说服力。

回应：感谢您的提问和建议。传统的 BSIT-P300 是指经过 Bootstrap 抽样后获得的 1000 个探测刺激和无关刺激的 P300 波幅差值大于零的次数。即在 1000 次 Bootstrap 抽样中, 如果探测刺激 P300 波幅大于无关刺激 P300 波幅的次数超过 900 次(BSIT-P300 > 900), 可认为该被试为“有罪者”，反之则被认为是“无辜者”。 然而在本研究中，不论是间谍组被试使用的虚假姓名还是无辜组被试使用的真实姓名都会诱发一个比无关刺激更大的 P300 波幅。因此，沿用传统的判别指标将会导致所有的被试都被判定为“间谍”。所以本研究中我们使用了 Bootstrap 抽样后无辜组被试的真实姓名与间谍组被试的虚假姓名之间的平均 P300 波幅差异减去无关刺激的差值(different mean of probe-irrelevant, DMPI)作为判断被试是否为“间谍”的标准，即您理解的无辜组被试的 CIT 效应的均值与间谍组的 CIT 效应的均值的差值。如果在 Bootstrap 抽样后被试的探测刺激所诱发的 P300 波幅和无关刺激所诱发的 P300 波幅之间的差异(PI-P300)小于 DMPI 的次数大于 900 次 (BSIT-P300 > 900), 则该被试被将被判定为“间谍”。

关于 900/1000 这一阈值，我们承认目前并不存在完全严谨的数理推导支撑，其本质上仍是一个经验性选择。然而，这一标准在理论直观和经验结果上都具有一定依据：首先，Rosenfeld (2020)指出，对于真正的无辜者被试，probe 在理论上可视为另一类 irrelevant，因此在多次重采样迭代下，probe > irrelevant 的期望比例应接近 50%。在此背景下，将判别阈值设定为 900/1000（即 90%），意味着只有当观测比例显著高于机会水平时才作出“知情”判定，从而在个体层面有效压低虚报率的同时，仍能保持较高的命中率。其次，该阈值与 P300-CIT 领域中 Rosenfeld 等及后续研究普遍采用的 bootstrap 判别标准保持一致，便于与大量现有工作进行直接比较 (Deng et al., 2016; Meijer et al., 2017; Rosenfeld et al., 2017; Ward et al., 2014)。同时，多项独立研究亦在该阈值下报告了较高的诊断效能。

意见 6：关于个体鉴别率偏低的问题。实验一和实验二虽在 ROC 分析中表现出较高的组间

鉴别率，但 bootstrap 的个体鉴别率（27/40）明显低于以往研究的水平。这一差异需要更多解释，并应讨论可能的改进方向，以提升个体水平的检测效度。

回应：感谢审稿专家提出的问题。我们同意在本研究中，虽然 ROC 分析表现出较高的组间区分效度，但基于 Bootstrap 的个体鉴别率（27/40）低于部分以往使用真实身份信息的 P300-CIT 研究。我们在修订稿的总讨论部分进行了补充，以解释这一差异并提出未来改进方向：

“虽然 P300 波幅在虚假身份信息识别中具有一定效用，但本研究中被试仅对虚假身份进行了三天的学习和识记……被试在身份内化和角色融入程度上可能存在显著个体差异。此类差异可能导致部分被试停留在表层记忆阶段，而另一些由于较强的角色代入充分内化了虚假身份，从而影响 CIT 效应的稳定性，也可能部分解释 ROC 组间与 bootstrap 个体鉴别率的差异。未来研究应考虑引入更系统化、分阶段的操纵性检验流程，如每日线上反馈、记忆强化训练或持续性角色沉浸任务 (Monaro et al., 2017; Slater & Sanchez-Vives, 2016)，以动态评估和强化虚假身份的学习与内化程度，从而提高操纵一致性与生态效度。”

意见 7：关于实验三的设计与解释。实验三采用被试内设计，样本量为 22 人。作者应补充 G*Power 的效能分析以与前两项研究保持一致。结果显示，控制个体差异后，个体鉴别率提高至 90%，这验证了虚假身份加工程度的个体差异确实影响检测效度。然而，这一结果也提示研究可能在现实中更具应用前景：在现实环境下，无论间谍还是普通身份者，对自身身份的加工都更深入，因而检测效度可能优于实验室研究。作者可在讨论中强调这一点，以凸显研究价值。

回应：首先感谢审稿专家对我们结果的肯定，以及提出的问题和具有启发性的建议。我们对实验三做了 G*Power 的效能分析，计算出所需要的样本量较少，仅为 12 名被试。为防止过大的抽样误差，所以我们增加了样本量保持了和前两项研究相同的样本量。目前已经在实验三的被试部分补充了对应的 G*Power 的效能分析：

“根据实验二的结果，P300 刺激类型主效应的效应量 $\eta p2 = 0.66$，我们将其转换为 f 值 ($f = 1.38$)，使用 G*Power3.1 计算研究三所需样本量，在显著性水平 $\alpha = 0.05$ 且效应量 $f = 1.38$ 时，预测达到 95%统计力水平 $(1-\beta)$ 的总样本需要 12 名被试(Faul et al., 2007)。考虑到该估计样本量较小，可能导致效应量估计不稳定并降低结果的稳健性，同时为了与前两个研究的样本规模保持一致，我们在研究三中共招募 22 名被试，其中两名被试由于没有记住自己的虚假身份信息而被排除在外。”

此外，我们也在总讨论中增加了对实验三结果的进一步深入讨论：

“值得注意的是，实验三在控制个体差异后个体鉴别率显著提升至 90%，这一结果不仅支持了前述关于虚假身份加工深度差异的解释，也为本研究的实际应用价值提供了重要启示。在现实情境中，无论是需要隐藏真实身份的间谍、犯罪嫌疑人，还是日常情境下的普通身份者，其对自身真实身份信息的编码、加工和内化程度均远高于实验室中短期构建的虚假身份。在真实环境中，身份信息往往与长期记忆、自我概念、情绪、甚至生理激励紧密耦合，因此其认知痕迹更为稳定、显著，难以在短期内进行抑制或替换。 基于此，实验三的结果可能反映了更贴近现实应用场景的上限：当身份加工深度得到保证时，P300 对真实与虚假身份的区分能力呈现高效且稳定的模式。这提示，在实际侦查、情报筛查或安全审查等应用领域中，个体在自然情境下对真实身份的强加工程度，可能使检测效度优于典型的实验室研究，从而增强基于脑电的 CIT 检测方法的外部效度与实用前景。未来研究可以在更贴近真实身份加工深度的环境下进一步验证这一推测。”

……

审稿人 2 意见：

该研究通过三个实验系统探讨了 P300 成分在识别虚假身份信息中的作用。实验设计合理，结果显示 P300 在区分真实与虚假身份信息方面表现出较高的鉴别准确率（达 90%）。我有以下几点建议与疑问：

意见 1：引言部分的理论基础有待加强。尽管文中阐述了研究虚构型虚假信息的现实意义，但未充分说明其与“否认型隐瞒”信息在认知加工过程中的异同。建议进一步阐释二者在信息处理机制上的差异与联系。

回应：感谢审稿专家的宝贵意见。我们已在引言中进一步强化了理论基础，特别是对“否认型隐瞒”与“虚构型隐瞒”在认知加工机制上的差异进行了阐释。具体修改如下：

“然而，现有研究主要集中于对真实存在、但被故意否认的信息进行检测(klein Selle et al., 2019; Olson et al., 2019; Zheng et al., 2022)，即“否认型隐瞒”。例如，在恐怖袭击案中犯罪嫌疑人通过否认真正的炸弹安装地点（如公园）以隐藏其真实记忆。而对“虚构型隐瞒”——如犯罪分子使用的虚假身份——的探测研究仍较为缺乏。虚构身份广泛出现在恐怖活动、间谍渗透、电信诈骗等多个对社会安全构成严重威胁的领域。在这些情景中，个体并非否认真实信息，而是学习并维持一个虚构的身份信息。”

意见 2：文中提到：“然而，现有研究大多集中于对真实存在但被故意否认的信息的检测，即‘否认型隐瞒’，而对虚构型虚假信息——如犯罪分子使用的虚假身份——的探测研究仍较为缺乏。”为便于读者理解，建议补充“真实存在但被故意否认的信息”的具体示例。

回应：感谢审稿专家提出的问题和建议。我们已在前言对应部分增加了具体示例以方便读者理解：

“……即“否认型隐瞒”。例如，在恐怖袭击案中犯罪嫌疑人通过否认真正的炸弹安装地点（如公园）以隐藏其真实记忆。……”

意见 3：文中指出：“因此近几年来研究者们开始探索采用心理测量的方法检测虚假身份信息。”建议作者进一步查阅相关文献，确认采用测量方法检测虚假信息的研究是否仅始于近年。

文中提到：“截至目前，仅有两项研究尝试对虚假身份信息进行检测。”但以下文献亦涉及身份信息伪装与隐藏的神经机制研究，建议核实是否遗漏相关研究。应该还有其它一些相关研究。

*Hu, X., Wu, H., & Fu, G. (2011). Temporal course of executive control when lying about self-and other-referential information: an ERP study. Brain research, 1369, 149-157.

*Ding, X. P., Du, X., Lei, D., Hu, C. S., Fu, G., & Chen, G. (2012). The neural correlates of identity faking and concealment: an FMRI study. PLoS One, 7(11), e48639.

*Zhang, P., Yan, L., & Wang, Q. (2023). Neural processes underlying faking and concealing a personal identity: An electroencephalogram study. Social Behavior and Personality: an international journal, 51(2), 1-12.

回应：感谢审稿专家对我们文献综述部分提出的细致意见。我们承认初稿中对“近几年来”以及“仅有两项研究尝试对虚假身份信息进行检测”的表述不够严谨，也未充分覆盖以往关于虚假身份与身份隐瞒之神经机制的研究。根据审稿专家提供的文献线索并重新检索核实，我们已在前言中补充并讨论了相关研究(Chen et al., 2023; Ding et al., 2012; Hu et al., 2011; Zhang et al., 2023)，以更准确地呈现该领域的发展脉络。但本研究聚焦于通过神经指标的差异作为身份检测的指标能够获得多高的准确率，将神经或行为指标作为“身份检测指标”进行

分类检测的研究，目前可检索到的工作主要集中于 Monaro 等人的两项研究。目前我们已经将涉及身份信息伪装与隐藏的神经机制的研究增加到前言部分，并说明了本研究与以往相关研究的差异：

“目前已有一些研究考察了虚构信息的神经机制 (Chen et al., 2023; Ding et al., 2012; Hu et al., 2011; Zhang et al., 2023)。例如，Hu 等人 (2011) 采用脑电技术发现，相比说真话，无论是否认真实身份信息还是虚构身份信息都会显著诱发更大的 N200 波幅和更小的 P300 波幅，表明否认和虚构信息的过程均需要更强的认知控制（如冲突监测），并伴随着更高的认知负荷；Ding 等人 (2012) 采用功能磁共振技术发现，相比说真话，否认个人信息激活与抑制控制相关的脑区（例如，腹外侧前额叶皮层，ventrolateral prefrontal cortex，vlPFC），而虚构个人信息则激活与工作记忆相关的脑区（例如，背外侧前额叶皮层，dorsolateral prefrontal cortex，dlPFC）。这些研究表明：相比说真话，虚构信息需要个体运用工作记忆和冲突检测等过程来保持自己的虚构信息，同时监测与事实相违背的反应。然而，这些研究大多只关注虚构信息的神经机制，没有考察如何有效地探测虚假身份。尽管目前已有多种生物识别工具，如指纹、人脸、视网膜等生理或行为特征来检查身份信息(Ashbourn, 2014)。但在恐怖主义或专业间谍等场景中，嫌疑人身份往往未知，其生物特征可能不在数据库中，从而难以仅凭生物特征确认其自曝身份的真实性 (Sabena et al., 2010)。因此，发展能够在“身份信息层面”评估真伪的神经指标与检测范式，具有现实必要性与方法学价值。

目前，仅有两项研究试图探索虚假身份信息的检测 (Monaro, Gamberini, et al., 2017b, 2017a). ……”

意见 4：文中提到：“真实身份组的被试在回答预期问题和意外问题时是快速并准确的。”请对“预期问题”与“意外问题”进行明确定义。

回应：感谢审稿专家的指正。很抱歉遗漏了对“预期问题”与“意外问题”的明确定义。根据审稿专家的建议，我们已在前言中补充了上述定义，以提高术语使用的清晰度和读者的理解度。

“Monaro 等人使用了一种意外问题范式来增加被试的认知负荷从而检测虚假身份的新技术。该技术包含三种问题：（1）预期问题（Expected questions），是实验前进行排练和记忆的信息。例如，姓名、出生日期、籍贯等常规身份信息均属于预期问题。（2）意外问题（Unexpected questions），是实验前没有经过排练但与预期问题密切相关的信息。例如，生肖、星座等。（3）控制问题（Control questions），是基于实际情况被试必须如实回答的一些个人信息。例如，“你现在正坐在电脑面前吗？”该理论认为，虚假身份持有者在应对意外问题时，需付出更多认知资源以抑制真实反应、提取虚构信息并监控回答一致性，从而在行为层面暴露异常。而真实身份持有者会对这些三类问题进行快速的自动化反应，但是虚假身份使用者则需要更多的时间对意外问题的答案进行计算和构建 (Monaro et al., 2018, 2021; Monaro, Gamberini, et al., 2017c; Monaro, Spolaor, et al., 2017)。”

意见 5：文中提及“BSDF-P300”与“BSIT-P300”，其中“BSDF”与“BSIT”的全称分别是什么？另，关于“BSIT-P300”的描述中是否重复出现该术语？请予以核实。

回应：非常感谢您对术语使用提出的细致指正。 关于您提及的“BSDF-P300“与“BST-P300”，其中：

- BSDF-P300 为 bootstrapped mean probe-minus-irrelevant differences (Deng et al., 2016) 的缩写，指在 Bootstrap 叠加平均后，探测刺激减去无关刺激的 P300 波幅的差值。
- BSIT-P300 为 bootstrapped significant iterations test (Deng et al., 2016)的缩写，是指经过 Bootstrap 抽样后获得的 1000 个探测刺激和无关刺激的 P300 波幅差值大于零的次数。

我们已全面检查文稿中两项指标首次出现的段落，并补充了各自的英文全称，以确保读

者能够 准确理解其含义。同时，也已按照您的建议重新审阅全文，确认“BSIT-P300“的表述无重复或 自我指代造成的歧义，并对可能引起混淆的部分进行了适当调整。 再次感谢审稿专家对细节的关注与专业建议，这对提升稿件的术语规范性和整体清晰度具有重要价值。

意见 6：请阐明 BSDF 与 BSIT 在方法学上的主要区别。
回应：BSDF-P300 是指 Bootstrap 抽样后探测刺激的平均 P300 波幅差异减去无关刺激平均 P300 波幅的差值。BSIT-P300 传统的 BSIT-P300 是指经过 Bootstrap 抽样后获得的 1000 个探测刺激和无关刺激的 P300 波幅差值大于零的次数。
因此，两者主要方法学差异可概括为：

1. 统计量类型不同： BSDF 使用“平均差值”，属于连续型效应量指标； BSIT 使用“差值为正的次数占比”，为离散型概率统计指标。

2. 关注点不同： BSDF 更强调差异的“幅度”； BSIT 更强调差异的“概率”。

3. 适用性略有不同：BSDF 可以清楚的看到被试在两种刺激上所诱发的 P300 幅度的差异，对波幅差异较大的被试更敏感； BSIT 可以对波幅差异较小，但方向稳定的情况仍可能表现较高的判别率。

意见 7：文中部分引用缺少页码信息，请补充完整。
回应：感谢审稿专家的提醒。我们已对全文的文献引用进行了逐一核查，并补充了此前缺失的页码信息。同时，需要说明的是，部分引用来源于在线优先出版的期刊文章，此类文献在正式排版前并不提供固定页码，因此无法补充对应页面范围。除上述情况外，所有需要标注页码的文献均已在修订稿中更正，以确保引用格式的准确性与完整性。 再次感谢审稿专家对细节的关注，这对提高稿件的规范性具有重要意义。

意见 8：作者在讨论部分提出了基于记忆机制的检测思路，建议在引言中对该理论框架进行简要介绍，以帮助读者更好地理解研究的理论依据与预期结果。
回应：感谢审稿专家的宝贵建议。我们同意，在讨论部分提出的基于记忆机制的检测思路，确实需要在引言中进行前置性的理论说明，以增强研究逻辑的完整性和预期效果的可理解性。根据建议，我们已在引言中新增了一段内容。再次感谢审稿专家提出的建设性建议，对进一步强化论文的理论框架具有非常重要的作用。

“除了说真话和虚构信息在认知负荷的差异外，真实信息与虚假信息在记忆层面也存在差异。虚假信息的加工更依赖新近学习、情境模拟、角色扮演以及自我相关性的人工构建。因此，相比真实信息，虚假信息往往存在缺乏深度编码、情绪意义和自传背景等特点，这种差异可能会导致其神经反应与真实自传信息存在差别。而 P300 波幅对刺激的意义性、熟悉度与再认强度高度敏感，这一敏感性来源于刺激在记忆系统中的编码深度(Karis et al., 1984；Johnson，1986；Polich，2007)。换言之，个体是否拥有某项信息的真实记忆表征，是决定其能否诱发显著 P300 的关键因素。因此：若真实身份信息与虚假身份信息在编码深度与熟悉性上存在系统差异，则其对应的 P300 反应也应呈现稳定差异。”

第二轮

审稿人 1 意见：

非常欣赏作者在本轮修改中所付出的大量努力，包括新增双侧乳突为参考条件下的对比分析等重要补充。可以清楚地看到作者对稿件进行了许多调整，以及对此前提出的每一条意见和问题所作出的详细的回应。总体而言，这一轮修改使得研究设计与论证逻辑更加合理，

稿件的整体质量得到了明显提升。为进一步提升稿件的学术价值与方法学透明度，现仍有以下两点建议供作者参考：

意见 1：关于 BSIT 个体分类阈值（“间谍”判定标准）的文献依据说明

在论文中，作者将个体是否被判定为“间谍”的标准设定为：在 bootstrap 抽样后，探测刺激所诱发的 P300 波幅与无关刺激所诱发的 P300 波幅之间的差异（PI–P300）小于 DMPI 的次数超过 900 次（即 BSIT–P300 > 900）。在回复信中，作者已说明该阈值与 P300–CIT 领域中 Rosenfeld 等及其后续研究普遍采用的 bootstrap 判别标准保持一致，且便于与既有研究进行直接比较（Deng et al., 2016; Meijer et al., 2017; Rosenfeld et al., 2017; Ward et al., 2014），同时多项独立研究也在该阈值下报告了较高的诊断效能。建议作者将上述文献依据与理由直接补充进正文中有关阈值设定的相应部分，以增强该“900”这一判别标准的理论与经验支撑，也有助于读者更好地理解该阈值选择的合理性。

回应：感谢您的细致审阅与宝贵建议。我们已根据您的意见，将有关阈值“900”设定的文献依据与理由补充到正文中对应部分。具体修改如下：

“阈值设定为 900/1000，即要求该效应在 90% 以上的重抽样迭代中保持一致方向，可作为较为保守的个体层面稳定性判据，有助于降低抽样波动导致的误判风险，并便于与既有 P300-CIT 研究进行比较(Deng et al., 2016; Meijer et al., 2017; Rosenfeld et al., 2017; Ward et al., 2014)。”

意见 2：关于 BSIT 分析中最优阈值判别的进一步探索

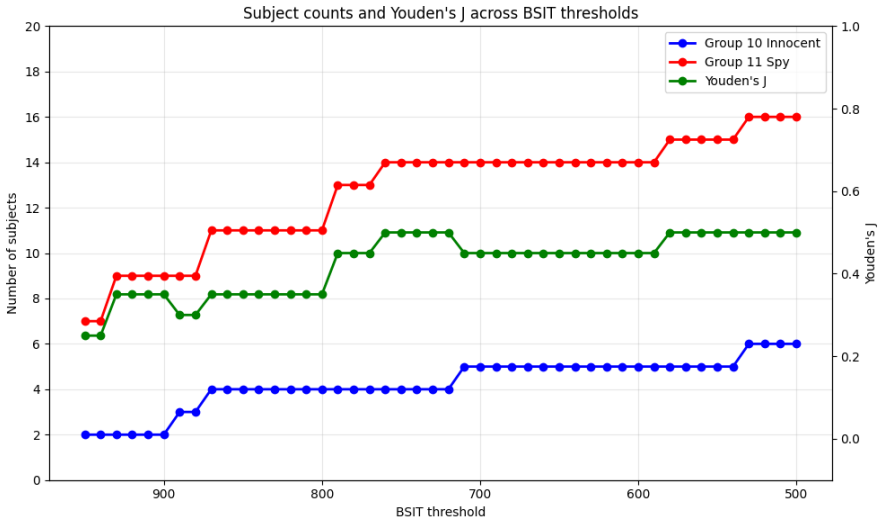
在 BSIT 分析中引入最优阈值判别（optimal cutoff）将具有更重要的意义。具体而言，作者可考虑在 probe–irrelevant > DMPI 的前提下，对 DMPI 阈值在 500–1000 区间内的个体鉴别率进行探索，以确定在何种阈值设定下可以获得最高的个体鉴别率。这一分析有助于进一步评估在虚构型虚假身份信息检测情境下，BSIT 的最优判别阈值与传统否认型谎言检测中常用的 900 阈值之间的关系，并为后续相关研究在阈值设定与方法学推广方面提供参考依据。

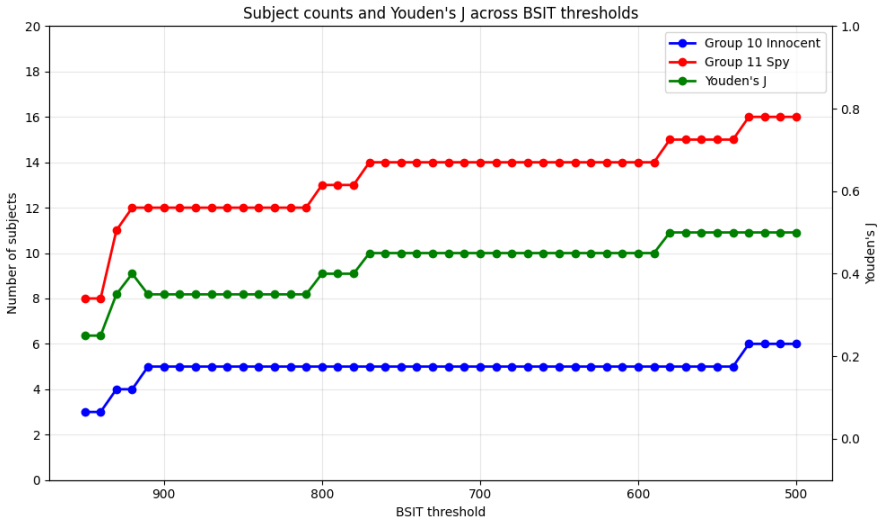
回应：感谢审稿专家的建议。我们同意，在 BSIT 分析中进一步探索不同判定阈值下的个体鉴别表现，有助于评估本研究情境下 BSIT 判定标准与传统 P300-CIT 研究中常用 900/1000 阈值之间的关系。根据专家建议，我们新增了基于最优阈值的探索性分析。具体而言，我们将 BSIT 判定阈值从 500 至 1000 以 10 为步长进行变化。在每一个阈值下，若个体的 BSIT 次数达到或超过该阈值，则将其判定为“间谍组”。对于同时包含无辜组和间谍组的研究 1 和研究 2，我们分别计算了间谍组命中率、无辜组误报率，并进一步计算 Youden’s J 指数，即 击中率− 误报率(Hassanzad & Hajian-Tilaki, 2024; Youden, 1950)。这样可以避免仅依据间谍组命中率选择较低阈值而导致无辜组误报率升高的问题。对于仅包含间谍组的研究 3，由于缺少无辜组参照，我们仅报告不同阈值下的间谍组检出人数和命中率，而不计算误报率和 Youden’s J。

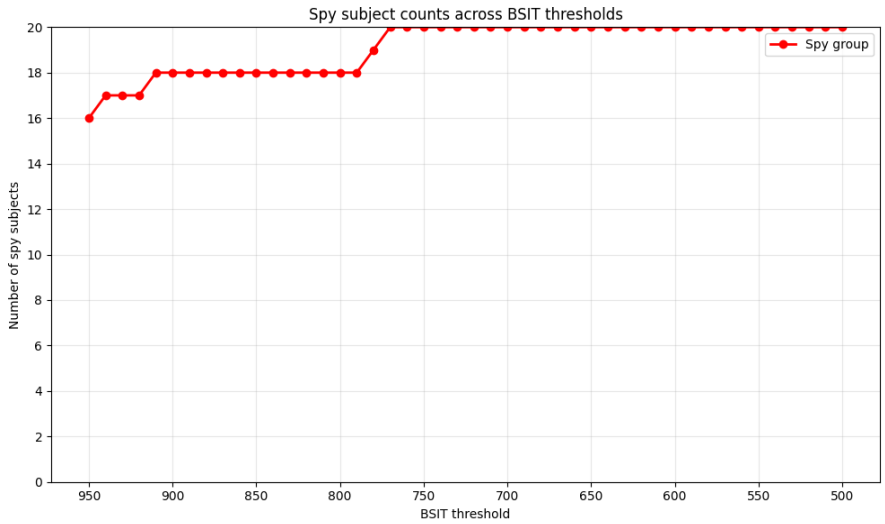
分析结果显示，在研究 1 中，传统 900 阈值下的命中率为 45.0%，误报率为 10.0%，Youden’s J 为 0.35；当阈值降低后，Youden’s J 的最大值为 0.50。其中，阈值 720–760 可在较低误报率的同时获得较好的命中率，例如阈值 750 时命中率为 70.0%，误报率为 20.0%，Youden’s J 为 0.50。在研究 2 中，传统 900 阈值下的命中率为 60.0%，误报率为 25.0%，Youden’s J 为 0.35；阈值 540–580 时 Youden’s J 达到最大值 0.50，对应命中率为 75.0%，误报率为 25.0%。研究 3 中，在传统 900 阈值下，20 名间谍组被试中有 18 人达到判定标准，命中率为 90.0%；当阈值降低至 770 及以下时，所有 20 名被试均达到判定标准。

总体而言，该探索性分析提示，在本研究的虚构型虚假身份信息检测情境中，传统 900 阈值具有较强的保守性，可以保持较低误报率，但可能牺牲部分间谍组检出率。相较之下，较低的经验阈值能够提高个体检出率，并在研究 1 和研究 2 中获得更高的 Youden's J。不过，由于该最优阈值分析是在当前样本内完成的，我们将其定位为探索性补充分析，而不以其替代主要分析中的传统阈值判定。我们已将该部分结果加入补充材料 2 中，并以补充图展示不同阈值下的命中率、误报率和 Youden's J 变化趋势。并在正文相应位置加上了引导语：

“此外，我们进一步对 BSIT-P300 的判定阈值进行了探索性补充分析。具体而言，我们将判定阈值从 500/1000 至 1000/1000 以 10 为步长进行变化，并在不同阈值下计算间谍组命中率、无辜组误报率及 Youden's J，以评估个体鉴别表现与误报控制之间的平衡，完整结果请见补充材料 2。”







补充图 1. 不同 BSIT 判定阈值下的个体鉴别表现，图上中下分别是实验一、实验二和实验三的结果。研究 1 和研究 2 中，红线表示间谍组中被正确判定的人数，蓝线表示无辜组中被误判为间谍组的人数，绿线表示 Youden's J，即命中率减去误报率。实验三仅包含间谍组，因此仅展示不同阈值下间谍组达到判定标准的人数。所有分析均以 BSIT 阈值从 500 至 1000、步长为 10 进行探索。

审稿人 2 意见：

感谢作者对先前审稿意见的回复。然而，相关问题尚未得到充分解决，具体如下：

意见 1：本人此前建议作者加强对"否认型隐瞒"与"虚假型隐瞒"在认知加工机制层面的差异分析。但当前回复仍主要停留在对两类隐瞒形式的描述性介绍，缺乏对其潜在认知机制差异的实质性讨论。尤其是在理论层面，两者在工作记忆负荷、冲突监测等关键认知过程上可能存在何种差异，仍未得到明确阐述。建议作者对此进行更为系统和深入的分析。

回应：感谢审稿专家的进一步建议。我们同意审稿专家的意见，即"否认型隐瞒"与"虚构型隐瞒"在认知加工机制层面存在差异。我们在上一版修订稿中的前言部分已经对其认知加工层面的差异进行补充。例如，在前言部分第一页最后一段关于虚构信息神经机制的介绍中讲到"Hu 等人 (2011) 采用脑电技术发现，相比说真话，无论是否认真实身份信息还是虚构身份信息都会显著诱发更大的 N200 波幅和更小的 P300 波幅，表明否认和虚构信息的过程均需要更强的认知控制（如冲突监测），并伴随着更高的认知负荷；Ding 等人 (2012) 采用功能磁共振技术发现，相比说真话，否认个人信息激活与抑制控制相关的脑区（例如，腹外侧前额叶皮层，ventrolateral prefrontal cortex，vlPFC），而虚构个人信息则激活与工作记忆相关的脑区（例如，背外侧前额叶皮层，dorsolateral prefrontal cortex，dlPFC）。"针对审稿专家的本次意见，我们又在描述两者的后面进行了相关补充：

"与"否认型隐瞒"不同，虚构型隐瞒不仅需要抑制真实反应，还需要主动生成虚假信息，并持续监控虚假信息与情境一致性，因此相比真实陈述，虚构型隐瞒可能涉及更高的认知负荷和更强的冲突监测(Ding et al., 2012; Hu et al., 2011)。"

意见 2：修改稿中增加了以往有关虚构信息的神经机制研究。尽管这些研究并未直接以虚假信息识别为研究目标，但其相关发现可能对本研究具有一些理论启示。目前稿件尚未充分讨

论这些研究与本研究之间的潜在联系。建议作者补充探讨这些相关对本研究理论基础的支持作用。

回应：感谢审稿专家的宝贵意见。正如专家所指出的，关于虚构信息的神经机制研究确实为虚假信息识别提供了重要的理论启示，尤其是其中指出虚构信息相较于真实陈述往往伴随着更高的工作记忆负荷和更强的冲突监测，这一特征为基于认知负荷差异的虚假信息识别与检测方法提供了理论依据。然而，本研究的核心切入点并非记忆层面，而是立足于从真实信息与虚假信息在记忆加工过程中的差异出发来探讨虚假信息的识别问题。因此，尽管上述神经机制研究对具有重要的参考意义，但其并不直接构成本研究的理论基础。

鉴于专家的建议，我们已在修订稿中补充这些研究如何为从认知资源角度探讨虚假信息检测的研究提供重要的理论支持的论述：

“尽管上述关于虚构信息神经机制的研究尚未直接探讨虚假信息的检测，但其研究结果为后续相关研究提供了重要的理论基础。例如，这些研究表明，相比真实陈述，虚构信息涉及更高的工作记忆负荷和更强的冲突检测过程，因此会消耗更多的认知资源。在此基础上目前有两项研究试图根据认知负荷探索虚假身份信息的检测 (Monaro, Gamberini, et al., 2017b, 2017a).”

第三轮

审稿人 1 意见：感谢作者的回复。我没有其他问题。

编委意见：同意录用。

主编意见：同意发表。