

《心理学报》审稿意见与作者回应

题目：任务状态动态调节汉字 N200 增强效应：来自贝叶斯阅读器模型的证据

作者：胡赛赛 伏泽青 周爱保 尹玉龙(通讯作者)

第一轮

审稿人 1 意见：

目前的研究通过两个 ERP 实验考察任务状态对汉字识别过程中 N200 增强效应的调节作用，并尝试在贝叶斯阅读器模型框架下解释相关结果。作者通过操纵任务类型、重复概率以及干扰水平，探讨这些因素如何影响 N200 成分，从而为理解视觉词汇识别中的正字法加工机制提供新的证据。整体而言，研究问题具有一定理论意义，将任务情境因素与 ERP 指标结合起来考察汉字识别过程，也具有一定的方法学价值。不过，在当前版本中，文章在理论概念界定、实验逻辑以及方法报告完整性等方面仍存在需要进一步修改和澄清之处。以下提出一些具体意见，供作者参考。

意见 1：文中部分核心概念在不同位置使用方式不一致，且理论关系界定不够清晰，需要统一和明确。

1) 文中多次交替使用“动态调节”与“动态优化”。两者在理论含义上并不等价：前者强调系统随情境变化而发生变化，而后者则包含明确的目标导向含义；

2) 研究问题中将“自动化神经反应”与“动态优化机制”表述为对立关系，但从理论上讲，两者并非必然互斥。例如，动态优化过程也可能通过自动化神经机制实现。

回应：非常感谢审稿专家指出本文核心概念使用不一致及理论关系界定不够清晰的问题。针对两个具体意见，我们分别回应如下。

关于“动态调节”与“动态优化”的区分。我们认同审稿人的判断，原稿中两个术语的交替使用缺乏明确的理论界定。在修改稿中，我们对二者的理论内涵进行了严格区分。“动态调节”是一个描述性概念，指认知加工参数随情境变量发生系统性改变，强调情境敏感性，但不预设这种改变是否朝向特定目标函数收敛。“动态优化”则是一个更强的理论主张，在情境敏感性的基础上进一步假定这种改变服务于以最小认知代价获取充分感知证据的目标。二者存在逻辑蕴含关系，动态调节是动态优化的必要而非充分条件。基于这一区分，修改稿在描述实验现象时使用“动态调节”，在贝叶斯模型框架下进行理论解释时使用“动态优化”。

关于"自动化神经反应"与"动态优化机制"的关系。我们认同审稿人的观点，原稿将二者表述为对立关系在理论上不够准确。经过梳理，我们认为原有表述未能准确反映两种理论的本质分歧。两种理论的关键区别不在于加工过程是否涉及自动化神经机制，而在于早期正字法加工是否对任务状态变量表现出系统性敏感性。模块化前馈理论主张早期正字法加工具有刺激驱动的前馈特性，加工参数主要由刺激物理属性决定，不随先验概率、任务复杂度等任务状态变量发生系统性改变。贝叶斯阅读者模型则预测，早期正字法加工的深度会随任务状态进行调谐性调整。这种调谐性加工并不排斥刺激驱动的前馈加工。

据此，修改稿对两种理论立场的关系进行了重新定位，不再将二者表述为对立关系。刺激驱动的前馈加工描述了信息沿视觉层级逐级传递的基本通路，构成正字法加工的基础架构。任务状态驱动的调谐性加工则描述了这一通路上的加工深度如何随任务状态参数进行灵活调整。两种理论的差异在于，模块化前馈理论认为前馈通路上的加工参数主要由刺激物理属性决定，任务状态不对其产生系统性影响。贝叶斯阅读者模型则进一步主张，在前馈加工基础上，加工深度还受任务状态的调谐性调整。本研究通过操纵先验概率、任务复杂度与干扰水平，发现 N200 增强效应对上述三个任务状态变量均表现出系统性敏感性，为调谐性加工的存在提供了神经电生理证据。这一发现拓展了对 N200 增强效应功能意义的理解，但并不否认前馈加工作为基础通路的作用。修改稿据此将研究问题重新表述为，重复对早期正字法加工的影响是否在刺激驱动的前馈加工外，是否还反映了任务状态驱动的调谐性加工。具体修改详见第 1 页摘要第 5 到 7 行，第 2 页第 9-10 行，第 4 页第 4-5 行和第 23 页第 1-9 行。

意见 2: 摘要中提及的研究目的“通过两个脑电实验考察任务状态对汉字 N200 增强效应的调节作用，验证贝叶斯阅读者模型在汉字识别中的适用性。”与研究问题“大脑如何动态优化字形识别过程？汉字重复呈现引发的 N200 增强效应为回答这一问题提供了窗口，但该效应究竟反映自动化神经反应还是动态优化机制尚存争议。”之间的逻辑联系不甚清晰，建议作者进一步说明，为何检验“任务状态调节”能够区分这两种理论解释。

回应: 感谢审稿人的细致审阅与建设性意见。审稿人准确指出了原摘要中研究问题与研究目的之间逻辑衔接的不足，即未能清晰阐明"任务状态调节"为何能够区分两种理论解释。我们已对摘要进行了相应修改。

两种理论立场的关键区别在于，早期正字法加工在重复呈现过程中是否对任务状态变量表现出系统性敏感性。模块化前馈理论主张，早期字形识别是刺激驱动的前馈加工过程，重复引发的神经反应变化源于神经元适应或稀疏编码等自动化机制，N200 增强效应的幅度主

要取决于刺激的物理重复特性，不随任务状态发生系统性改变。贝叶斯读者模型则在承认前馈加工作为基础通路的前提下，进一步主张正字法加工深度会根据先验概率、任务复杂度和干扰水平进行调谐性调整，因此 N200 增强效应应随任务状态的变化而系统性地增强或减弱。据此，考察任务状态能否系统性地调节 N200 增强效应，构成了检验早期正字法加工是否在前馈加工之外还存在调谐性加工的关键实证检验。

在修改后的摘要中，我们在研究问题之后补充了两种理论预测，将原文"该效应究竟反映自动化神经反应还是动态优化机制尚存争议"修改为"重复对早期正字法加工的影响是否在刺激驱动的前馈加工外，还包含任务状态驱动的调谐性加工，目前仍存争议。前者预测这种效应不受任务状态影响，后者则预测其随任务状态动态优化"，从而明确了从研究问题到研究目的的推理逻辑。修改详见摘要部分，第 1 页第 5 到 7 行。

意见 3：两个实验在多个方面存在差异，包括：实验 1 为双字词，实验 2 为单字；实验 1 为真假词判断任务，实验 2 为语义判断任务；实验 1 操纵重复概率，实验 2 操纵干扰水平；与实验 1 中重复概率和任务类型对重复词对的影响方式不同，实验 2 中干扰字对与重复字对并不重叠。这些差异使两个实验在逻辑上并非严格的延续设计，因此在整合解释时需要明确区分各自的证据范围。请在引言中明确指出两个实验之间的逻辑关系。

回应：感谢审稿人的建议。我们理解审稿人的关切，即两个实验在刺激类型、任务形式、操纵变量以及干扰字对与重复字对的关系等方面存在差异，需要在引言中更清晰地阐明两者的逻辑关系与各自的证据范围。

我们认同两个实验并非严格的延续设计，而是服务于同一理论框架下的互补性检验。贝叶斯读者模型包含先验概率、任务复杂度与干扰水平三个核心参数，两个实验分别从不同参数组合的角度检验其对 N200 增强效应的调节作用。

具体而言，实验 1 旨在检验先验概率与任务复杂度对 N200 增强效应的调节作用。鉴于双字词是 N200 增强效应研究中最经典且效应最为稳定的刺激类型，实验 1 选用双字词作为实验材料。实验 1 的关键发现在于，高重复概率条件下相同-不同判断任务中 N200 增强效应完全消失，先验概率与任务复杂度的交互作用在一定程度上支持了任务状态对正字法加工的动态调节。然而，通过重复比率操纵先验概率时，物理特征层面的视觉重复率与字词水平的出现概率同步升高，尚无法完全排除前馈加工机制对 N200 增强效应消失的贡献。

实验 2 据此引入贝叶斯读者模型的第三个参数，即干扰水平，以进一步验证任务状态对字形识别过程的优化作用。由于干扰水平的提升要求识别系统积累更多特异性感知证据，

理论上提高干扰水平应能提高 N200 增强效应。与此同时，实验 2 需要在操纵干扰水平的过程中控制物理特征层面的视觉重复性，以排除前馈加工因素的替代解释。为此，我们引入正字法相似字来操纵干扰水平。这些相似字形与完全重复字形一样，与目标汉字共享主要视觉部件，在物理特征层面保持了高度的视觉特征重复性，从而控制了刺激驱动的前馈加工因素。与此同时，相似字形在整字水平上与目标字完全不同，构成了词汇识别层面的干扰。若 N200 增强效应在视觉特征重复概率基本不变的条件下，因干扰水平提升而恢复，则可排除前馈加工的解释，支持任务状态动态调节的观点。

在刺激类型方面，由于双字词水平上难以构建足够数量且严格匹配的正字法相似材料，实验 2 改用单字作为刺激。现有研究证明单字同样能够引发显著的 N200 增强效应。相应地，为适应单字，任务从真假词判断调整为语义判断，两者均要求被试通达语义表征。这些调整均服务于在单字水平上精确控制正字法相似性这一核心目标，同时保持语义通达的任务复杂度不变。

根据审稿人的建议，我们在修改稿引言末段对两个实验逻辑关系进行了进一步地明确阐述，补充说明了两个实验各自的证据功能、实验 2 的设计动机，以及刺激类型与任务形式调整的理由(第 4 页第 18 行到第 30 行)。此外，我们在讨论部分的"理论意义与局限"中进一步补充了对两个实验设计差异的说明(第 23 页第 20 行到第 24 页第 2 行)。

意见 4：实验 2 结果中，“提升干扰水平成功恢复了原本在高重复概率条件下消失的 N200 增强效应”，但实验 2 并未设置重复概率条件，无法与实验 1 形成对比。请在结果与讨论中删除或修改上述表述，将实验 2 的结果限定在实验 2 自身设计框架内解释，并在讨论部分再说明其与实验 1 结果之间的潜在联系。

回应：感谢审稿人指出这一表述问题。我们完全同意，实验 2 的结果应首先在其自身设计框架内进行解释，在总讨论中再明确两个实验之间的潜在联系。

考虑在实验 2 中，高干扰水平条件，在保持物理特征层面的视觉重复概率与实验 1 中高重复比例条件保持一致的同时，增加了词汇水平的干扰，在修改稿中，我们做了以下调整。在实验 2 的结果部分，我们删除了"提升干扰水平成功恢复了原本在高重复概率条件下消失的 N200 增强效应"这一跨实验比较的表述。从实验 2 干扰水平这一变量的特点出发，我们将这一表述修改为“即使物理特征层面的视觉重复概率维持较高水平，高干扰水平条件仍诱发了显著的 N200 增强效应，且其幅度与低干扰水平条件无显著差异。这一结果表明，词汇水平干扰的引入能够调节汉字字形重复识别的早期正字法加工”(第 17 页 6 到 9 行)。这一表

述限定在实验 2 自身设计的框架内。

在实验 1 的讨论部分，我们相应调整了由实验 1 引出实验 2 的过渡段落。在阐述高重复概率条件下 N200 增强效应消失的理论意义后，增补了实验 1 的方法论局限，即重复比率操纵导致物理特征层面的视觉重复率与字词水平的重复率同步升高，无法完全排除前馈加工的贡献。据此引出实验 2 的设计动机，即在维持视觉特征重复概率不变的前提下单独提升词汇水平的干扰，考察干扰水平能否调节 N200 增强效应，从互补角度为调谐性加工提供进一步的证据。该过渡段落的表述聚焦于实验 2 自身的变量逻辑，避免预设跨实验比较的结论(第 12 页 27 行到第 13 页第 3 行)。

在讨论 4.1 节，我们重新组织了论证结构，明确了两个实验之间的潜在联系。在讨论实验 1 的发现后，我们增补了对实验 1 局限性的说明，指出通过重复比率操纵先验概率时，物理特征层面的视觉重复率与字词水平的重复率同步升高，使得高重复概率条件下 N200 增强效应的消失无法完全排除前馈加工机制的贡献，并据此引出实验 2 的互补性检验角色。在讨论实验 2 的发现时，我们对比了实验 1 高重复概率条件与实验 2 高干扰条件在视觉物理特性与词汇水平上的异同，指出二者在物理特征层面的视觉重复概率上保持一致，但在词汇水平上存在关键差异。在此基础上，我们论证仅凭刺激驱动的前馈加工不足以解释实验 2 中 N200 增强效应的变化模式。此外，我们补充分析了 N200 增强效应幅度未随干扰水平线性增加这一现象，提示前馈加工与调谐性加工可能协同决定了该效应的最终表现。最终通过两个实验的互补关系，我们得出两个实验的结果共同支持贝叶斯阅读者模型关于汉字字形重复识别早期正字法加工过程中存在任务状态驱动的调谐性加工机制的主张(第 20 页 5 行到第 21 页第 6 行)。

意见 5：结论部分指出，“本研究支持将视觉词汇识别视为贝叶斯推断过程”，但似乎一些相关的模型，例如，预测编码模型，也可以从一定程度上解释目前的结果。也就是说，贝叶斯阅读者模型并不是目前数据的唯一解释。建议作者在结论部分适当收缩理论推断的范围，以避免过度外推。

回应：感谢审稿人的的审慎建议。我们认同该意见。我们从贝叶斯阅读者模型的视角出发，考察了任务驱动的调谐性加工在汉字字形重复识别早期正字法加工中的作用，但这一理论选择并不排斥其他同样强调自上而下调节作用的模型，如审稿人提出的预测编码模型。当前数据所揭示的任务状态驱动的调谐效应并非贝叶斯框架所独占的预测。

根据审稿人的建议，我们进行了如下修改。第一，在理论意义与局限部分，将理论定位

从"验证了贝叶斯阅读者模型在汉字加工领域的适用性"调整为"以贝叶斯阅读者模型为理论框架,为汉字字形重复识别早期正字法加工阶段存在任务状态驱动的调谐性加工提供了实证支持",从而将结论限定在特定理论视角下的证据贡献,而非对模型排他性的验证。同时,明确指出预测编码等框架亦可在一定程度上容纳本研究发现,并指出未来研究可通过设计针对性实验范式来区分不同模型的差异化预测(第 23 页第 1 行到第 12 行)。第二,在结论部分,删除了原文"本研究支持将视觉词汇识别视为贝叶斯推断过程"这一具有排他性暗示的表述,将理论贡献收窄为"将贝叶斯阅读者模型的核心预测从拼音文字拓展至汉字加工领域"。同时,将结论调整为"上述发现揭示了视觉词汇识别不仅涉及刺激驱动的前馈加工,还包含根据任务情境灵活调整加工深度的调谐性过程",以降低推断幅度,使结论的外推范围与实验数据的支撑力相匹配(第 24 页第 4 行到第 12 行)。

意见 6: 摘要中提及“本研究采用重复范式”,重复范式应被更准确表述。例如,“本研究采用视觉重复启动范式”。

回应: 感谢审稿人的建议。修订稿中已将“重复范式”明确表述为“视觉重复启动范式”,以提高术语准确性。

意见 7: 被试部分(涉及实验 1 和 2),请说明目前样本量的选择依据。

回应: 感谢审稿人的建议。本研究在数据采集前使用 GPower 3.1 进行了先验功效分析。考虑到本研究的核心假设预测任务状态变量与重复类型之间存在交互作用,我们以重复测量方差分析中 2×2 交互效应作为关键检验,设定中等效应量 Cohen's $f = 0.25$ (对应 $\eta_p^2 = 0.06$),显著性水平 $\alpha = 0.05$,统计检验力 $1 - \beta = 0.80$,重复测量间相关系数 $r^* = 0.50$ 。计算结果表明所需最小样本量为 24 人,据此实验 1 和实验 2 均以 24 名有效被试的数据进入最终分析。此外,两个实验中观测到的关键交互效应量均远超先验假设的中等效应量水平。具体而言,实验 1 中重复类型与任务类型的交互作用效应量 $\eta_p^2 = 0.42$,重复类型、重复概率与任务类型的三重交互作用效应量 $\eta_p^2 = 0.25$,实验 2 中重复类型与任务类型的交互作用效应量 $\eta_p^2 = 0.13$ 。上述效应量表明当前样本量具有充分的统计检验力。修订稿已在两个实验的被试部分补充样本量依据(第 5 页第 4 到第 7 行,第 13 页第 7 行)。

意见 8: 实验设计部分,建议进一步阐述高、低重复概率条件的设定依据。例如,是否有参考研究?是否有预实验数据支撑?

回应：感谢审稿人对实验设计细节的关注。高、低重复概率条件的设定综合考虑了范式结构的内在约束、已有研究的实证基础以及数据采集的可行性。

首先，本实验采用启动范式，且严格限制每个汉字仅重复呈现一次，以排除多次重复带来的混淆效应。这一设计决定了单个 block 内重复试次的理论上限为 50%。在此基础上，实验还需纳入一定比例的反应刺激与填充刺激以维持被试的任务参与度和注意水平，因此高重复概率条件最终确定为 41.67%。

低重复概率条件设定为 20.83%，这一取值具有明确的实证支撑。我们此前的研究在约 20% 的重复概率下观察到了稳定且显著的 N200 增强效应，表明该概率水平足以可靠地诱发目标 ERP 成分(Yin et al., 2020; Yin & Zhang, 2021)。因此，20.83% 的设定既有实验室内部数据的验证，也与该领域既有文献的参数选择相吻合。

两种概率条件之间约 2:1 的比例关系还兼顾了实验效率的考量。在保证每个条件下不少于 50 个试次的前提下，若进一步降低重复概率，总试次数将大幅增加，从而延长实验时长，可能引发被试疲劳和注意力衰减等问题，反而不利于数据质量。当前的概率设定在效应的可检测性与实验可行性之间取得了合理的平衡。

我们已在修改稿方法部分补充了上述设计依据及相关文献引用(第 5 页第 19 到 25 行)。

意见 9：材料部分，“填充刺激、非重复词对和假词的随机插入打破了刺激呈现的完全可预测性”表述中，建议避免使用“打破”这类较为绝对的表述，可考虑替换为“降低”以弱化过强的逻辑表达。

回应：非常感谢审稿人的建议。在修订稿中，“填充刺激、非重复词对和假词的随机插入打破了刺激呈现的完全可预测性”修改为“填充刺激、非重复词对和假词的随机插入降低了刺激呈现的完全可预测性”(第 7 页第 2 行到第 3 行)。

意见 10. 建议报告实验材料视角。

回应：非常感谢审稿专家提出我们的疏漏。在修改稿中，我们已经补充每个实验刺激材料对应的视角信息。实验 1 中，双字词以宋体形式呈现，视角大小为 $2.05^{\circ} \times 0.83^{\circ}$ 。实验 2 中，单字是以宋体形式呈现，视角大小为 $1.03^{\circ} \times 0.83^{\circ}$ 。上述信息已分别添加至修改稿实验 1 和实验 2 的“实验设备和程序”部分(第 6 页第 28 行和第 13 页第 28 到 29 行)。

意见 11：相同-不同判断任务中，是否也包含假词？

回应：感谢审稿专家的细致审查。在本研究中，相同-不同判断任务不包含假词，假词仅出现在真假词判断任务中。我们已在修改稿的实验方法部分对此予以明确说明(第 7 页 6 到 7 行)。

两种任务间刺激构成的差异根植于各自不同的认知加工需求。真假词判断任务要求被试对刺激进行词汇通达并作出词汇性判断，因此必须包含假词作为任务组成的必要部分，以确保真假词判断过程的真实性和有效性。相同-不同判断任务则要求被试比较两个依次呈现的刺激在正字法层面是否一致，其判断依据为视觉或正字法特征的匹配，不涉及词汇身份的识别。在此任务框架下，假词既不承担任务功能，也缺乏理论上的设置动机。相反，在相同-不同判断任务中引入假词可能诱发被试附加的词汇性加工，进而干扰任务类型变量对字形识别加工过程需求的操纵。因此，该任务中不设置假词是基于任务性质的合理设计选择。

值得强调的是，本研究的核心因变量，即重复启动量，均通过各任务内部重复条件与非重复条件的差异获得，不涉及两种任务之间的直接比较。因此，任务间刺激构成的差异不会混淆对重复启动效应的估计。

意见 12：建议报告各条件下的有效试次数。

回应：非常感谢审稿专家的建议。报告各条件下进入 ERP 平均的有效试次数是确保数据质量透明性的重要环节。有效试次数直接影响 ERP 波形的信噪比，进而关系到统计分析结果的可靠性(Luck, 2014)。此外，各条件间有效试次数的均衡程度也是排除因信噪比差异导致条件间 ERP 波幅系统性偏差的必要前提。遵照审稿专家的建议，修订稿已在方法部分补充各条件下进入 ERP 平均的有效试次数，以均值和标准差的形式报告(第 9 页第 5 到 11 行、第 15 页第 27 行到 16 页第 3 行)。

意见 13：请作者在附录中补上按电极位置呈现的各条件下的总平均图，以便读者更全面评估 ERP 成分的空间分布。

回应：非常感谢审稿专家的建议。按电极位置分别呈现总平均波形图是评估 ERP 成分空间分布特征的重要手段。正文中报告的组平均波形图基于感兴趣区域内多个电极的平均值绘制，这种做法虽能直观展示实验效应的整体趋势，但可能掩盖成分在不同头皮位置上的幅度差异与分布梯度。逐电极呈现总平均波形图有助于读者独立验证感兴趣区域选取的合理性，同时评估实验效应是否在该区域内呈现均匀分布或存在局部差异。遵照审稿专家的建议，修订稿已在附录中补充各条件下 F3、Fz、F4、FC3、FCz、FC4、C3、Cz、C4 共 9 个电极位

置的 ERP 总平均波形图，以便读者更全面地评估 N200 增强效应和 N400 效应的时空分布特征。

意见 14：建议作者进一步解释，既然干扰水平的设置让被试在相同-不同判断条件下都出现了 N200 增强，为何这种增强量在低干扰条件下反而略高于高干扰条件？

回应：感谢审稿专家的细致审阅。统计检验表明，相同-不同判断任务中高、低干扰条件的 N200 增强效应无显著差异($t(23) = 0.50, p = .620$)。我们理解审稿专家关注的核心问题在于，既然干扰水平的提升增加了对特异性感知证据的需求，为何 N200 增强效应的幅度并未随之进一步增大。

经过深入思考，我们认为这一结果恰好反映了刺激驱动的前馈加工与任务状态驱动的调谐性加工对 N200 增强效应的共同约束。一方面，干扰水平提升促使调谐性加工增加正字法加工深度以应对识别的不确定性。另一方面，前馈通路中的神经反应(如神经元适应或稀疏编码)对重复刺激的加工强度施加了衰减作用。这两种方向相反的力量相互抵消，使得 N200 增强效应在高、低干扰条件间未表现出显著差异。换言之，前馈加工构成字形识别的基础通路并设定了加工强度的基本约束，调谐性加工则在此基础上根据任务状态灵活调整加工深度。N200 增强效应的最终幅度由二者协同决定，而非单一机制的线性函数。

这一分析进一步澄清了本文的理论立场。本研究并非主张调谐性加工完全取代前馈加工，而是强调早期正字法加工在前馈通路的基础上展现出对任务状态的动态敏感性。我们已在修改稿讨论部分 4.1 节末尾补充了上述分析(第 20 页第 22 行到第 21 页第 3 行)。

.....

审稿人 2 意见：

本文通过两个脑电实验，系统考察了任务状态、重复概率、干扰水平对汉字 N200 增强效应的调节作用，验证了贝叶斯阅读器模型在汉字识别中的适用性。研究问题明确，但在理论衔接、研究方法等方面仍有改进空间。

意见 1："该公式包含三个核心参数。P(W)为目标字形的先验概率.....分母表征感知证据的边缘概率....."需对核心符号 $f(I|W)$ 进行中文命名解释，以利于读者理解；分母部分的表述未明确"干扰水平提升会使边缘概率增大"的核心逻辑，与后文研究假设的衔接关系不明确。

回应：审稿人的意见揭示了原稿在贝叶斯阅读器模型论述上的一个系统性不足。我们认为审稿人的意见包含三方面的关切。第一，原稿仅以数学形式 $f(I|W)$ 呈现公式核心符号，未提

供简明中文术语，增加了非贝叶斯背景读者的理解门槛。第二，原稿对分母项的描述较为简略，未充分阐明分母随干扰水平提升而增大的内在机制，使读者难以从公式本身推出实验 2 的操控逻辑。第三，由于上述两点的不足，公式与后文研究假设之间的衔接略显松散，研究假设无法被读者识别为公式的自然推论。我们认同审稿人的判断，并据此在修改稿中以三处定点修改予以系统改进。

第一，关于核心符号 $f(I|W)$ 的中文命名和解释。该项在 Norris (2006) 的模型中被定义为词汇 W 的似然函数(likelihood function of W)，其含义为在呈现单词 W 的条件下感知到感知证据 I 的概率。结合其在原文献的名称和功能含义，我们在修改稿中将其命名为“感知似然函数”，并补充说明该项反映视觉编码系统对感知证据的累积强度(第 2 页第 19 行到 20 行)。本修改即在术语层面与原始文献保持一致，也降低了非贝叶斯背景读者的理解负担。

第二，关于分母项随干扰水平变化的内在机制。原稿仅以“该值增大，识别所需的特异性证据随之增加”概括分母的变化结果，未阐明这一变化的机制，使读者难以从公式本身理解实验 2 引入正字法相似字以操控干扰水平的理论依据。根据 Norris (2006)对贝叶斯阅读者模型的阐述，分母项量化了视觉输入 I 由词汇库中任意字形生成的总体可能性。当某一目标字形的证据强度显著高于其他候选字形时，目标字形的后验概率趋近于 1。反之，若多个候选字形的证据强度彼此接近，则后验概率被稀释，识别系统需要积累更多感知证据才能达到判断阈值。基于这一机制，当任务情境中存在与目标字形共享主要正字法部件的竞争候选时，这些候选对视觉输入的证据强度 $f(I|w_i)$ 与目标字形的证据强度 $f(I|W)$ 处于相近水平，无法在求和过程中被忽略，从而有效纳入分母的求和过程。分母的扩张直接导致目标字形后验概率的下降，识别系统由此需要积累更多特异性感知证据才能达到判断阈值。据此，修改稿在公式之后的解释段补充了边缘概率的稀释机制，明确呈现“竞争候选的证据强度相近、纳入求和、分母扩张、后验概率下降”这一推理链条，使分母随干扰水平变化的内在逻辑更加清晰(第 2 页第 26 行到第 3 页第 1 行)。

第三，关于公式与后文研究假设之间的衔接。研究假设段的三项研究假设虽提及先验概率、任务复杂度与干扰水平三个因素，但未将每项预测明确锚定至公式中的特定参数，使研究假设与公式的对应关系不够明晰。修改稿对三项具体预测的措辞进行了重新组织，使每项预测均可追溯至公式中对应的参数变化。具体而言，先验概率的预测对应分子项 $P(W)$ 的变化，任务复杂度的预测对应决策阈值的变化，干扰水平的预测则对应分母项的扩张机制。

修改后的三项预测共同构成了由公式直接推导的可检验神经预测,公式由此从理论引用转化为假设推导的基础 (第 3 页第 21 行到 25 行)。

意见 2: "最新研究发现, N200 增强效应在双字条件下的波幅大于单字条件, 在真字条件下大于假字和非字条件, 这一现象表明汉语正字法重复不仅涉及字形特征的提取, 还参与了从亚词汇部件到整体词汇表征的整合过程(Hu et al., 2025)。"如作者这部分内容的叙述, 那本研究的两个实验分别选取了双字词和单字词进行研究, 是否会导致研究考察的加工机制不同, 或考察的加工过程不同, 应在前言中叙述这样进行操纵的目的和必要性, 并在讨论中进行解释。

回应: 非常感谢审稿专家的细致审阅。审稿专家精准地指出了 Hu 等 (2025)所揭示的双字词与单字在 N200 增强效应波幅上的差异, 以及这一差异对本研究两个实验比较性的潜在影响。我们就此问题进行了深入思考, 在区分加工机制的质与量差异的基础上, 对前言与讨论部分进行了相应修订。

我们认为, Hu 等(2025)所报告的波幅差异反映了正字法-词汇整合程度的连续性变化, 而非加工机制的本质转换。N200 增强效应在单字与双字词条件下均能稳定诱发, 且在两种刺激条件下均反映了早期正字法加工过程(Hu et al., 2025; Yin & Zhang, 2021; Zhang et al., 2012)。双字词条件下波幅更大, 源于该刺激类型同时激活了从亚词汇部件到整字、再到整词的多层级整合, 正字法证据需求量相应增加。单字条件下, N200 增强效应主要反映从部件到整字的整合过程。两者在加工层级的覆盖广度上存在差异, 但在加工性质上均属于早期正字法加工范畴。

更为重要的是, 本研究的研究问题为早期正字法加工是否对任务状态变量表现出系统敏感性。该命题的检验并不依赖于特定刺激类型, 而依赖于在该刺激类型下能否稳定观察到 N200 增强效应, 以及该效应能否随任务状态变量发生系统性调节。两种刺激类型均满足上述条件。因此, 实验 1 选用双字词以利用该刺激类型 N200 增强效应最为稳定的优势考察先验概率与任务复杂度的调节作用, 实验 2 改用单字以满足正字法相似性的方法学需求增加干扰水平。最终, 二者在不同的正字法-词汇整合层级上检验了同一理论命题。

跨刺激类型的会聚证据反而强化了本研究结论的普遍性。两个实验在不同正字法-词汇整合水平上均观察到一致的任务状态调节模式, 即真假词判断与语义判断任务诱发的 N200 增强效应均强于相同-不同判断任务。这一跨刺激类型的会聚模式表明, 任务状态驱动的调谐性加工并非局限于特定词汇单元层级, 而是早期正字法加工的普遍属性。

依据上述思考，我们对前言与讨论部分进行了如下修订。

在前言部分，我们在阐述实验 2 刺激类型调整的方法学考量之后,补充说明了两种刺激类型下 N200 增强效应在加工性质上的共通性，论证实验 1 与实验 2 虽采用不同刺激类型，仍能在各自对应的正字法-词汇整合层级上检验任务状态对早期正字法加工的调节作用，从而在前言阶段即明确两个实验设计的理论一致性(第 4 页第 26 行到 27 行)。

在讨论 4.4 节，我们在原有的基础上，新增了一段专门论述。该段首先提出审稿人所关注的核心理论问题，即不同词汇单元层级上的 N200 增强效应是否反映同一加工机制，以及该问题对本研究核心结论成立基础的潜在影响。继而引用 Hu 等(2025)的发现，从正字法-词汇整合广度的角度解释双字词与单字间的波幅差异，即双字词条件涵盖从部件到整字、再从整字到整词的多层级整合，单字条件则主要涉及从部件到整字的整合。在此基础上,我们从时间特性、地形分布及影响因素三个维度论证两种刺激类型下 N200 增强效应在加工性质上的共通性，并结合本研究两个实验在不同整合层级上观察到的一致波形特点与任务状态调节模式，论证刺激类型的差异并不削弱本研究对核心理论命题的检验效力。最后,我们承认本研究的设计无法直接量化刺激类型对调节效应强度的独立贡献，并提出未来研究可在同一实验中纳入多种刺激类型与任务形式以做进一步检验(第 23 页第 20 行到第 24 页 2 行)。

意见 3："相较而言，相同-不同判断任务可通过视觉特征的直接匹配完成，主要依赖正字法表层分析(Barron & Pittenger, 1974)。而语义判断任务要求被试通达词汇的语义表征....."。"真假词判断、相同-不同判断"、"语义判断任务"，概念需统一，并阐明不同任务的差异性，比如加工内容、加工阶段等。

回应：感谢审稿专家精准指出本文在任务概念使用上的不统一以及任务差异性论述不充分的问题。这一意见对厘清本研究的任务操纵逻辑、增强论述的理论一致性具有重要价值。经过深入反思，我们认为原稿存在两方面的不足。一方面，实验 1 的"真假词判断任务"与实验 2 的"语义判断任务"虽表层操作不同，但在加工要求上属于同一类任务，原稿未对此同质性进行明确界定。另一方面，三种任务在加工内容与加工阶段两个维度上的差异未得到系统阐述，削弱了论述的理论严密性。

为系统回应这一意见，我们以加工深度为核心维度，将本研究采用的三种任务整合为两大类。

第一类为词汇-语义通达任务，包括真假词判断任务与语义判断任务。两者在任务操作上确实存在差异，前者要求被试判断刺激是否为真实词汇，后者要求被试判断刺激的语义类

别归属。然而，两者在底层加工机制上具有同质性。二者均要求被试完成从正字法表征到语音表征再到语义表征的完整加工通路，必须通达词汇的语义层级方能做出有效反应。从加工内容看，该类任务的加工核心在于词汇身份与语义表征的激活。从加工阶段看，其加工过程延伸至词汇通达之后的语义提取阶段。因此，从贝叶斯阅读者模型框架审视，二者对正字法感知证据的累积要求处于同一水平，均需要较深层次的字形分析以支持后续的词汇-语义通达。

第二类为正字法匹配任务，即相同-不同判断任务。该任务可通过部分视觉特征的直接物理匹配完成(Barron & Pittenger, 1974)，被试无需足够正字法表征激活以通达语音与语义表征即可基于正字法表层分析做出判断。从加工内容看，该任务的加工核心在于视觉特征与正字法表层信息的比较。从加工阶段看，其加工过程主要局限于早期正字法表层分析阶段，无需进入词汇通达阶段。因此，该任务对正字法证据的累积要求显著低于词汇-语义通达任务。

这一统一框架不仅消除了原稿术语使用问题，更凸显了任务操纵的理论效力。两类任务的对比构成了贝叶斯阅读者模型中任务复杂度参数的核心实证操纵。词汇-语义通达任务对应高复杂度条件，要求积累更多正字法感知证据。正字法匹配任务对应低复杂度条件，所需感知证据相对较少。两个实验虽采用不同的词汇-语义通达任务实例(实验 1 为真假词判断，实验 2 为语义判断)，但其作为高复杂度条件的功能定位完全一致。这种实例间的可替换性恰恰反映了任务操纵效力跨越具体任务形式的稳健性。

基于上述分类框架，我们对修改稿进行了系统性修订，具体修改如下。

在引言部分系统引入两类任务的分类框架。该段从加工内容与加工阶段两个维度阐明词汇-语义通达任务与正字法匹配任务的本质差异，并明确其与贝叶斯阅读者模型任务复杂度参数的理论对应关系，为后续两个实验的任务操纵奠定统一的理论基础(第 4 页第 12 行到 17 行)。

第二，在两项实验的实验设计部分，将具体任务纳入统一的功能分类框架。实验 1 中的真假词判断任务与实验 2 中的语义判断任务均被界定为词汇-语义通达任务的实例，相同-不同判断任务则被界定为正字法匹配任务的实例(第 5 页第 26 行到 6 页第 3 行、第 13 页第 20 行到第 26 行)。这一界定明确凸显了两个实验在任务操纵层面的同质性，为跨实验整合解释提供了一致的概念基础。

第三，在摘要与总讨论 4.2 节同步使用上述分类术语展开论述。这些修订共同确保了从研究问题提出到结果整合的全文概念表述一致性。

需要进一步说明的是,"真假词判断""语义判断""相同-不同判断"等术语在视觉词汇识别领域已具有稳定的使用基础,直接更换可能损害本研究与既有文献的可对接性。基于这一考虑,我们采用"功能类别加具体实例"的层级化命名策略。"词汇-语义通达任务"与"正字法匹配任务"作为统摄性的功能类别名引入,原有任务名则作为具体实例保留。这一处理既维持了与既有文献的对接,也实现了任务操纵逻辑的概念统一。

通过上述修订,三种任务在统一的功能分类框架下获得清晰的层级定位,原稿术语使用不一致的问题得到系统性解决,任务操纵作为任务复杂度变量实证检验的理论效力亦得以凸显。

意见 4: 关于研究目的的创新性。据贝叶斯读者模型已明确可知,个体在阅读过程中会动态评估当前任务环境,并据此调节正字法加工的深度与强度。然而,"目前尚不清楚 N200 增强效应所反映的汉字正字法早期加工过程是否会随任务状态发生动态优化"。若如此,本研究的目的仅仅是为了考察这个 N200 指标是否适用于贝叶斯模型吗? 那该单独考察的意义和必要性还需要厘清。

回应: 非常感谢审稿专家就研究目的的创新性所提出的关键意见。我们认同,若仅将本研究的目的表述为"检验 N200 指标是否适用于贝叶斯读者模型",则不仅会削弱研究的理论意义,亦无法凸显其独立的学术价值。我们对审稿人意见的理解是,本研究需要在贝叶斯读者模型的一般性理论预测之外,明确界定其针对特定现象的独立实证贡献。经过梳理,我们将本研究的核心问题重新表述为一个具体而可检验的实证议题,即汉字重复对早期正字法加工的影响是否在刺激驱动的前馈加工外,还反映任务状态驱动的调谐性加工。

相较于原稿,这一重新表述在研究意义与必要性两方面实现了三重提升。其一,研究目的不再依附于既有模型的适用性检验,而是直接指向对汉字早期正字法加工机制本身的理论判定。其二,研究问题从笼统命题收紧为可被实证检验的具体判定,即重复对早期正字法加工的影响是否包含前馈加工之外的额外加工成分。其三,研究的实证贡献不再止于为某一模型补充新的 ERP 指标。研究直接检验任务状态驱动的调谐性加工是否在刺激驱动的前馈加工通路之上构成额外的加工成分,从而推进对汉字早期正字法加工层级架构的理解。

基于上述调整,我们在修改稿中的摘要部分(第 1 页第 5 行到第 7 行)、引言部分(第 2 页第 9 行到第 10 行)及结论部分(第 24 页第 4 行到第 6 行)均重新界定了本研究的目的,并相应调整了研究意义的论证逻辑与理论贡献的表述范围。

意见 5: "模块化加工理论则认为文字识别是一个模块化的前馈过程", 该理论与本研究的关系仅存在于实验 2 吗? 总讨论部分, 也未有关于该理论的讨论。另外, "前馈"的意义和内涵需要补充。

回应: 非常感谢审稿专家的建议。我们认同原稿对“模块化前馈加工理论”的引入和讨论仍不够充分, 容易给人一种该理论仅与实验 2 相关的印象。事实上, 该理论是本研究两个实验共同面对的替代性解释框架。经过梳理, 我们将该理论在本研究中的功能定位重新厘定为贯穿两个实验的对照性解释框架, 并以前馈加工与调谐性加工的层级架构整合了两个实验的核心发现。据此, 修改稿从三个层面进行了系统修订。

第一, 在引言部分明确"前馈"的理论内涵。我们从加工方向、决定因素及任务无关性三个维度对该术语加以限定, 并以神经元适应与稀疏编码作为其神经实现机制的具体例证, 进而将任务状态调节确立为区分前馈理论与贝叶斯读者模型的关键实证检验(第 3 页第 26 行至第 30 行)。

第二, 在讨论 4.1 节使前馈理论作为对照框架贯穿两个实验的论证。针对实验 1, 我们指出重复比率操纵导致视觉重复率与字词重复率同步升高, 前馈解释因此构成实验 1 结果的替代性解释。针对实验 2, 我们排除前馈解释, 并进一步论证前馈加工与调谐性加工的协同作用决定了 N200 增强效应的最终表现。前馈理论由此从针对单一实验的局部对照上升为贯穿两个实验的整体对照框架(第 20 页第 4 行至第 21 页第 6 行)。

第三, 在讨论 4.4 节将上述层级架构提升为本研究对汉字早期正字法加工机制的核心理论判定, 即前馈加工构成基础通路, 调谐性加工在该通路之上实现加工深度的弹性调整。该架构同时锚定两个实验的发现。实验 1 因无法完全分离两类机制的贡献而需要实验 2 的互补性检验, 实验 2 则在控制前馈因素的前提下进一步揭示了调谐性加工的独立作用。模块化前馈理论由此作为对照框架贯通两个实验, 而非仅构成实验 2 的替代性解释(第 23 页第 1 行到第 12 行)。

意见 6: 第三页 2.1.1 被试部分"由于技术原因及脑电图伪影过多, 我们最终使用了 24 名被试(8 男 16 女)的数据。"首先, 应补充数据量的计算方法; 其次, 应说明被试删除的原因(如电极接触不良、设备故障等); ERP 实验需明确伪影剔除的具体标准, 而不能仅提及"伪影过多"不符合学术报告规范。

回应: 非常感谢审稿人的宝贵建议。原稿对样本量来源、被试剔除原因以及伪影剔除标准的报告确实较为简略, 未能充分体现 ERP 研究的方法学规范要求。根据您的建议, 我们在修

订稿中对样本量计算、伪影剔除的具体标准和被试剔除原因进行了系统补充。

第一，样本量依据。原稿未报告样本量的统计学计算过程。修订稿在两个实验的被试部分前置补充了基于 G*Power 3.1 的先验功效分析，以重复测量方差分析中 2×2 交互效应作为关键检验，结合中等效应量与常规检验力参数推算出最小有效样本量 24 人，并据此确定实际招募规模。具体参数与计算结果已写入修订稿(第 5 页第 4 至 7 行，第 13 页第 7 行)。

第二，伪影剔除的标准和被试剔除原因。根据预处理流程，在试次水平上，凡任一通道电位振幅超过±100 μV 的 epoch 均予剔除。在被试水平上，若任一条件下有效试次少于 40 个，则该被试不进入 ERP 统计分析。该双重标准既保障单试次信噪比，又维持各条件间有效试次数的可比性(第 8 页第 10 行到第 9 页第 4 行)。

第三，被试剔除原因。原稿"技术原因及脑电图伪影过多"的并列表述未能精确反映实际剔除依据。经核查，两名被试的剔除均源于伪影剔除后有效试次低于前述预设阈值，并不涉及独立的设备或电极层面的技术故障。修订稿据此将相关表述调整为"由于 2 名被试数据的有效试次不足预先设定阈值"，使剔除决策与方法学标准形成明确对应(第 5 页第 11 至 12 行，第 13 页第 11 行)。

意见 7：原文第四页 2.1.2 实验设计部分"在高重复概率条件下，重复词占组块内所有刺激的比例为 41.67%；在低重复概率条件下，该比例为 20.83%"。建议补充比例选择的文献依据或预实验基础。此外，重复概率的设置是否等同于先验概率，这是本研究的核心变量，还需要说明和解释。

回应：感谢审稿人对实验设计细节的关注。我们将针对比例选择的依据和重复概率与先验概率的关系分别回复。

关于比例选择的依据，高、低重复概率条件的设定综合考虑了范式结构的内在约束、已有研究的实证基础以及数据采集的可行性。本实验采用视觉重复启动范式，严格限制每个汉字仅重复呈现一次以排除多次重复带来的混淆效应，因此单个组块内重复试次的理论上限为 50%。考虑到实验还需纳入反应刺激与填充刺激以维持被试的任务参与度，高重复概率条件最终确定为 41.67%。低重复概率条件设定为 20.83%，这一取值基于我们此前研究在约 20% 的重复概率下观察到的稳定且显著的 N200 增强效应(Yin et al., 2020; Yin & Zhang, 2020)。此外，两种概率条件之间约 2:1 的比例关系兼顾了实验效率的考量，在保证每个条件不少于 50 个试次的前提下，避免因总试次数过多而引发被试疲劳。修改稿已在方法部分补充了上述设计依据(第 5 页第 19 到 25 行)。

关于重复概率与先验概率的关系，二者并非概念上完全等同，而是理论变量与操作化定义之间的对应关系。贝叶斯阅读器模型中的先验概率 $P(W)$ 是一个理论构念，指识别系统在接收感知输入之前对目标字形出现可能性的内部估计。这一估计并非外部实验参数的直接映射，而是认知系统基于当前任务情境中积累的统计规律性所形成的内部表征。实验中的重复概率操纵正是通过改变这一统计规律性来影响先验概率的。具体而言，当组块内重复试次的比例从 20.83% 提升至 41.67% 时，被试在实验进程中持续接收到更高比例的重复事件，其认知系统据此更新对后续刺激重复出现的预期。这种基于经验的预期更新使得被试在高重复概率条件下形成更强的“当前刺激可能是重复刺激”的先验信念。根据贝叶斯阅读器模型的推导，当先验概率 $P(W)$ 升高时，后验概率在较少感知证据的积累下即可达到决策阈值，识别系统因此降低了对感知证据的需求。反映在神经层面，正字法加工深度的降低表现为 N200 增强效应幅度的减弱乃至消失。因此，重复概率的操纵能够通过改变被试对刺激重复性的统计学习，系统地调节其内部先验估计，从而实现对贝叶斯阅读器模型中先验概率这一理论变量的有效实验操纵。修改稿已在引言部分最后一段补充了上述说明(第 4 页第 7 到 12 行)。

意见 8：原文第 5 页 2.1.3 材料部分“我们选取 200 个高频双字词作为核心刺激，将其随机均分为四组”。仅说明选自《现代汉语常用字词频词典》，应明确“高频”的具体量化界定（如词频 $\geq$ 多少次/百万字），并说明为何选用高频词。

回应：感谢审稿专家指出实验材料细节的建议。我们认同原稿中对“高频双字词”的界定还不够具体。根据您的意见，我们已在修改稿中补充说明“高频”的量化标准，即实验 1 所选核心刺激均满足词频不低于 50 次/百万字。同时，我们进一步说明选用高频词的理由：一方面，高频词更符合熟练阅读中的常见加工对象，可减少因词汇熟悉度差异带来的额外噪音；另一方面，高频词在词汇通达上更稳定，有利于将实验效应更明确地归因于任务状态变量和重复加工，而非词汇可识别性本身的差异。我们也保留了四组材料在词频与笔画数上无显著差异的统计说明，以进一步保证材料可比性。我们已在修改稿材料部分补充了上述相关信息(第 6 页第 7 到 9 行)。

意见 9：原文第 5 页 2.1.3 材料部分中“这些假词由真实汉字随机组合而成，例如，‘爱可’和‘你船’。”仅举例说明假词构成，未说明假词的控制变量（如笔画数、视觉复杂度与真词的匹配性统计结果），建议补充假词的视觉复杂度与实验所用真词是否存在差异，排除其他视觉特征的干扰。

回应：感谢审稿专家提出的细致建议。我们认同，仅用例子说明假词构成不足以充分体现材料控制的严谨性。根据您的意见，我们在修改稿中进一步补充真词和假词的视觉复杂度(笔画数)之间是否存在差异说明。结果发现，真词的总笔画数($M = 18.75, SD = 5.10$)与假词的总笔画数($M = 18.40, SD = 3.94$)之间不存在显著差异， $F(1, 238) = 0.16, p = .688$ ，表明实验材料具有良好的可比性。我们已在修改稿材料部分补充了上述相关信息(第 6 页第 12 到 13 行)。

意见 10：前言和讨论主要关注了 N200，数据分析为何选用了 N200 和 N400，需要说明。

回应：感谢审稿专家的提醒。我们认同原稿在理论阐述上主要聚焦于 N200 增强效应，而未充分说明同时纳入 N400 增强效应分析的必要性。根据您的意见，我们在修改稿中对这一分析策略作了进一步澄清。

纳入 N400 效应分析的考虑主要基于以下三方面。其一，N200 增强效应主要反映汉字识别早期的正字法加工，而 N400 效应更多反映后期的语义与语音整合加工。同时分析两个成分，有助于在时间进程上区分任务状态对早期正字法加工与后期整合加工的不同调节模式。其二，纳入 N400 效应有助于评估 N200 增强效应的功能独立性。若两者在相同实验条件下表现出不同的调节模式，则可从时程和功能两个层面支持 N200 增强效应与 N400 效应具有相对独立的功能意义。其三，两个实验的结果均显示，N400 效应主要受任务复杂度驱动，并不明显受先验概率和干扰水平的制约，这与 N200 增强效应对上述三个任务状态参数均表现出系统性敏感性形成了清晰的分离模式。这一分离进一步凸显了 N200 增强效应作为早期正字法加工指标的理论独特性。

根据上述考虑，我们已在修改稿的 ERP 分析部分对此进行了补充说明(第 9 页第 15 到 19 行)，并在讨论部分进一步区分 N200 增强效应和 N400 效应的功能意义及其调节模式差异(第 22 页第 12 到 15 行)。

意见 11：原文使用"N200 增强效应""N400 效应""N400 效应"等，建议全文统一。

回应：非常感谢审稿专家的细致指正。我们认同原稿在 ERP 术语表述上存在不够统一之处。根据您的建议，我们已对全文相关术语进行了统一处理。考虑到此前文献的表达习惯，因此修订稿中统一使用"N200 增强效应"和"N400 效应"的表述。这样处理不仅提高了全文表述的一致性，也有助于更准确地呈现两种 ERP 成分在本研究中的功能意义。

意见 12：实验二无讨论。

回应: 感谢审稿专家的提醒。我们认同原稿在实验 2 结果之后缺少与实验 1 相对应的阶段性小讨论, 导致实验 2 的主要发现、理论含义未能得到及时而清晰的概括。根据您的意见, 我们在修订稿中补充了实验 2 结果之后的讨论(第 18 页第 20 到 29 行)。

意见 13: 原文第 18 页"我们推测, 高重复概率条件下 N200 增强效应的消失, 本质上反映了认知系统根据先验概率主动优化加工策略....."这里"主动优化加工策略"是一个描述性术语, 神经计算层面具体指什么? 表述过于模糊。建议与预测编码或贝叶斯推理的神经实现框架进行关联, 使推测更具理论深度和可检验性。

回应: 非常感谢审稿专家提出这一非常有价值的建议。我们认同原稿中“主动优化加工策略”的说法偏描述性, 尚未充分揭示其可能对应的神经计算含义。我们修改的核心思路是将解释锚定于贝叶斯读者模型的公式框架, 使原本模糊的"优化"获得参数层面的具体指向。因此, 我们在修改中进行了如下调整。删除"主动优化加工策略"这一缺乏计算约束的表述, 以"证据累积阈值调节机制"替代, 使解释直接对应贝叶斯读者模型中从先验概率到后验概率再到感知证据累积量的参数传递链条。经此修改, N200 增强效应幅度的变化被明确界定为证据累积深度变化的神经电生理表征, "优化"不再是一个隐喻, 而是指向可从公式参数出发推导出可观测 ERP 预测的具体计算过程。我们已在修改稿总讨论部分对相关表述进行了修正(第 19 页第 28 到第 20 页 5 行)。

意见 14: 原文第 19 页第 2 段"汉字作为语素文字, 正字法与语义的联结更为紧密(Perfetti et al., 2005), 因此任务依赖性在汉字加工中可能表现得更为明显。"该句仅引用了一篇文献, 且未具体说明 Perfetti 等人的理论如何支持"联结更紧密"以及为何会导致"任务依赖性更强"。建议补充解释, 使逻辑更完整。

回应: 非常感谢审稿专家的建议。我们认同原稿在该处的论述较为简略, 未充分展开“汉字正字法—语义联结更紧密”与“任务依赖性可能更强”之间的逻辑关系。根据意见, 修改稿从两方面补充了论证。第一, 明确"联结更紧密"的理论来源。汉字作为语素文字, 字形结构直接对应语素, 使正字法表征与语义表征之间存在比拼音文字更直接的映射关系。第二, 将这一文字系统特性与段落的核心论点相衔接。由于汉字正字法加工更直接地服务于语义通达, 正字法分析作为词汇-语义通达任务"刚性前提"的约束力在汉字中更强, 而正字法匹配任务仍可在表层分析即告完成。两类任务在正字法证据需求上的落差因此被拉大, 先验概率效应

的任务依赖性相应增强。文献引用也从单一文献扩展至已有参考文献列表中的相关文献。我们已在修改稿总讨论 4.2 节对相关表述进行了修正(第 21 页第 29 行到第 22 页 6 行)。

第二轮

审稿人 2 意见：

作者对第一次审稿中提出的所有问题均给予了认真、系统且具有理论深度的回应。修改稿在理论概念界定、实验逻辑表述、方法学报告完整性、结果讨论的严谨性等方面均有显著提升。补充的样本量依据、伪影剔除标准、材料控制数据以及逐电极波形图等，也符合 ERP 研究的方法学规范。

意见 1：首先，关于研究目的。作者关于研究目的进行了重新的解释和梳理，“然而，已有研究尚未明确重复对早期正字法加工的影响是否在刺激驱动的前馈加工外，反映任务状态驱动的调谐性加工。为回答该问题，本研究基于贝叶斯阅读者模型的理论视角，设计实验系统考察字形识别过程中的任务状态与 N200 增强效应间的关系。”作者将研究问题明确聚焦于“前馈加工之外是否存在调谐性加工”，并以此统摄两个实验的设计与解释，增强了文章的理论说服力，但是作者只解释了“前馈加工”和“贝叶斯阅读者模型的调谐性加工”的理论观点，但在前言中未列举任何相关实证研究的证据，所以不确定在该观点的冲突解释上，作者是否具有首创性，若是的话，还请明确说明；若不是的话，需补充相关的实证研究证据，并在此基础上阐述研究的创新点。另外需要注意，作者在修改过程中导致部分内容的改变，因此需要整合修改和前后顺序的调整，免得很重要的内容在较后面才出现，比如 P1 L9 出现的“前馈加工”，一直到 P2 L26 才进行了解释。

回应：非常感谢审稿专家指出前言中相关实证证据不足、理论创新边界不够明确，以及重要概念出现与解释顺序不当的问题。我们认同审稿专家的判断。原稿主要说明了贝叶斯阅读者模型和刺激驱动前馈加工的理论观点，但未充分梳理相关实证研究，因而没有清楚呈现本文与既有研究之间的关系。根据审稿专家的建议，我们重新检索并梳理了相关文献，对前言的实证依据、创新表述和内容顺序进行了调整。

首先，我们不主张前馈加工与任务状态驱动的调谐性加工之间的解释分歧为本文首创。修订稿补充了重复概率影响熟悉文字加工、重复概率对早期与晚期加工产生不同影响、类别预期调节早期正字法加工，以及视觉词启动效应具有任务依赖性等方面的研究证据

(Kinoshita & Norris, 2012; Li & Kovács, 2022; Song et al., 2024; Wang & Maurer, 2020)。这些研究表明，视觉文字加工能够受到概率、经验、预期和任务目标的影响，但相关证据仍较为分散，尚未在汉字 N200 增强效应中系统考察重复概率的作用是否依赖具体任务类型，以及其他任务状态因素是否进一步改变重复引起的早期正字法加工。修改详见第 4 页，第 3 行到第 19 行。

据此，我们重新界定了本文的创新点。本文的理论增量不在于首次提出早期正字法加工可能受到任务状态调节，而在于以贝叶斯阅读者模型为统一框架，系统考察先验概率、任务复杂度和干扰水平对汉字 N200 增强效应的影响。其中，实验 1 检验重复概率对 N200 增强效应的作用是否具有任务类型依赖性，实验 2 进一步从干扰水平这一不同参数维度进行考察。修改后，本文的创新表述建立在已有研究基础上，不再对理论问题的首创性作过度推断。

其次，我们对前言的重要概念及内容顺序进行了局部调整。原稿开头的“已有研究尚未明确重复对早期正字法加工的影响是否在刺激驱动的前馈加工外，反映任务状态驱动的调谐性加工”修改为“已有研究尚未明确，重复对汉字早期正字法加工的影响是否反映任务状态驱动的调谐性加工”，避免在尚未解释其含义时提前引入“前馈加工”。后文依次呈现贝叶斯阅读者模型及参数预测、相关实证证据、刺激驱动前馈加工的替代性解释和本研究的检验逻辑。修改详见第 2 页，第 8 行到第 9 行。

意见 2：其次，关于两个实验之间关系的论述。“基于前述贝叶斯模型的预测，实验 1 预期高先验概率和低任务复杂度条件下 N200 增强效应减弱。然而，通过重复比率操纵先验概率时，物理特征层面的视觉重复率与字词水平的重复率同步升高，使得实验 1 无法完全排除前馈加工机制的贡献。据此，实验 2 引入第三个参数，即干扰水平，以进一步分离前馈加工与调谐性加工。”诚如作者所述，实验 1 存在两个变量效应的混淆，即物理特征层面的视觉重复率与字词水平的出现概率同步升高，但是实验 2 并不是进一步在操纵先验概率的基础上，将两个混淆因素进行了区分，而是转而探讨了第三个参数，即干扰水平。所以两个实验探讨的内容是不同的，是否具有互补性的研究值得商榷。反而引发了一个疑问，若实验 2 即可解决前馈加工和调谐性加工之间的争论，那实验 1 存在的必要性是什么？

回应：非常感谢审稿专家进一步追问两个实验之间的关系及实验 1 的必要性。我们认同，原稿将实验 2 表述为在实验 1 基础上“进一步分离前馈加工与调谐性加工”，容易使读者误以为实验 2 延续了重复概率操纵，并解决了实验 1 中物理特征重复概率与字词水平重复概率同步变化的问题。实际上，实验 2 并未继续操纵重复概率，因而不能直接区分这两个层面概率的

独立贡献。原稿对两个实验互补关系的表述不够准确。根据审稿专家的意见，我们重新界定了两个实验各自的研究目的、证据功能及解释边界。

实验 1 旨在检验重复概率与任务类型是否共同影响 N200 增强效应。高、低重复概率的比例操纵在真假词判断和相同-不同判断任务中完全一致，但概率调节效应仅出现在相同-不同判断任务中。两类任务采用相同的概率设置却呈现不同结果，说明提高重复概率并不必然引起 N200 增强效应发生相同变化，其作用受到任务类型约束。因此，实验 1 的证据功能在于揭示重复概率效应的任务依赖性。与此同时，实验 1 无法进一步区分物理特征重复概率与字词水平重复概率的独立贡献，这构成了其证据边界。

实验 2 没有延续实验 1 的重复概率操纵，因而不能直接区分上述两个层面概率的贡献，也不能单独解决前馈加工与调谐性加工之间的解释问题。实验 2 以实验 1 揭示的任务依赖模式为参照，在较高物理特征重复背景下引入正字法相似字，从字词水平识别竞争这一不同维度继续比较两类任务。

实验 2 的结果显示，在语义判断和相同-不同判断任务中，高、低干扰条件下的 N200 增强效应均不存在显著差异。因此，实验 2 不能单独证明干扰水平对 N200 增强效应具有梯度调节作用。若脱离实验 1 独立解释，实验 2 主要表明任务类型影响早期正字法加工，而干扰水平的作用相对有限。结合实验 1 所揭示的任务依赖模式，实验 2 才能进一步表明，当字词水平的识别竞争状态发生改变时，N200 增强效应仍随任务要求而变化。

基于上述分析，我们仍保留两个实验具有“互补性”的定位，但重新明确了这一概念的含义。两个实验的互补性并不意味着实验 2 修复或取代实验 1，也不意味着实验 2 直接消除了实验 1 中的概率混淆。相反，两个实验承担了不同但相互关联的证据功能。实验 1 建立了重复概率与任务类型共同影响 N200 增强效应的结果模式；实验 2 则从字词水平识别竞争这一不同参数维度，检验该任务依赖模式的解释边界。结合两个实验的结果，可以较为完整地说明，汉字重复引起的早期正字法加工既与刺激的物理重复有关，也会随任务要求和字词水平识别状态而变化。

根据这一重新定位，我们对正文进行了以下修改。第一，在引言中删除了“实验 2 进一步分离前馈加工与调谐性加工”的表述，改为分别说明两个实验的研究问题、证据功能及互补关系。修改详见第 5 页，第 10 行到第 12 行，第 19 到第 21 行。第二，在实验 1 的讨论中，强化了两类任务采用相同概率操纵却呈现不同概率效应这一结果的理论意义，同时明确指出实验 1 无法区分两个层面概率的独立贡献。修改详见第 13 页，第 24 行到第 29 行。第三，在实验 2 的结果和讨论中，我们删除了“恢复消失的 N200 增强效应”及干扰水平产生梯度调

节的表述，并增加了以实验 1 结果模式为参照的解释。修订稿指出，高、低干扰条件下的 N200 增强效应无显著差异，跨实验模式的变化可能与词汇水平干扰的引入有关。修改详见第 17 页，第 28 行到第 30 行，第 18 页，第 10 行到第 13 行，第 20 页，第 2 行到第 25 行。第四，在总讨论中重新阐明两个实验的联合推理及各自的证据边界，不再将实验 2 表述为对实验 1 混淆因素的直接解决。修改详见第 20 页，第 11 行到第 16 行，第 21 页，第 17 行到第 21 行。

意见 3：第三，补充效应量是否具有检验力的依据。“实验 2 中重复类型与任务类型的交互作用效应量 $\eta_p^2 = 0.13$ 。上述效应量表明当前样本量具有充分的统计检验力。”

回应：非常感谢审稿专家指出第一轮回复中效应量与统计检验力表述不够严谨的问题。经重新核查，审稿专家所引用的“实验 2 中重复类型与任务类型的交互作用效应量 $\eta_p^2 = 0.13$ 。上述效应量表明当前样本量具有充分的统计检验力”出现在第一轮审稿意见回复中，而非论文正文。论文正文仅报告了数据采集前完成的先验功效分析。

第一轮回复中的原意，是说明样本量规划所依据的效应量假设与实际观察到的关键效应量之间具有一致性。具体而言，我们在数据采集前使用 G*Power 3.1 进行先验功效分析，以重复测量方差分析中的交互作用作为关键检验，采用中等效应量 Cohen's $f = 0.25$ 作为样本量估计依据，该效应量约对应 $\eta_p^2 = 0.06$ 。在显著性水平 $\alpha = 0.05$ 、统计检验力 $1 - \beta = 0.80$ 、重复测量间相关系数 $r = 0.50$ 的设定下，计算得到所需最小样本量为 24 人。实验 1 和实验 2 均以 24 名有效被试的数据进入最终分析。实验完成后，两个实验中观察到的关键效应量均不低于这一先验设定的效应量标准，其中实验 2 重复类型与任务类型交互作用的效应量为 $\eta_p^2 = 0.13$ 。

不过，我们认同审稿专家的意见，第一轮回复中“上述效应量表明当前样本量具有充分的统计检验力”这一表述不够准确，容易被理解为根据观察效应量对统计检验力进行事后推断。更准确地说，观察到的效应量只能作为与先验效应量假设相对照的描述性信息，说明实际结果的效应量水平没有低于样本量规划时预设的中等效应量标准，而不能作为统计检验力充分的独立证明。因此，在本轮回复中，我们对此作出澄清：本研究样本量的依据是数据采集前完成的先验功效分析，而非研究完成后观察到的效应量。

论文正文原本仅报告先验功效分析的参数与计算结果，并未依据观察效应量推断统计检验力。因此，我们未对正文作实质性修改。通过上述澄清，我们希望更准确地区分先验样本量规划与观察效应量描述的不同功能，避免将观察效应量误用为事后功效判断的依据。

意见 4: 第四,“由于汉字正字法加工更直接地服务于语义通达,正字法分析作为词汇-语义通达任务“刚性前提”的约束力在汉字中更强,而正字法匹配任务仍可在表层分析即告完成。两类任务在正字法证据需求上的落差因此被拉大,先验概率效应的任务依赖性相应增强。”由于汉字正字法加工更直接地服务于语义通达,所以两类任务在正字法证据需求上的落差因此被拉大,该表述是否正确,正字法和语义的密切性提高了,所以不同任务间在在正字法证据需求上的落差因此被拉大还是减少?建议核查相关文献补充证据。

回应: 非常感谢审稿专家指出该处逻辑表述仍需进一步核查。我们重新检查了第一轮回复及正文 4.2 节,发现原稿确实将“汉字正字法与语义联结更紧密”进一步推论为“两类任务在正字法证据需求上的落差被进一步拉大”,这一表述过于直接。我们认同审稿专家的意见,正字法与语义联系更紧密本身并不能必然推出不同任务间正字法证据需求差异增大,也可能意味着正字法信息能够更稳定地支持语义通达。因此,原表述中的因果推断需要收缩。

根据审稿专家的建议,我们对总讨论 4.2 节进行了修改。修改详见第 23 页,第 11 行到第 15 行。修订稿不再使用“两类任务在正字法证据需求上的落差被进一步拉大”这一表述,而是改为强调汉字正字法-语义联系对词汇-语义通达任务的约束作用。具体而言,汉字作为语素文字,字形结构与语素、语义之间存在较为直接的联系,形旁等正字法线索能够为语义通达提供亚词汇层面的信息支持。因此,在词汇-语义通达任务中,正字法分析不仅是早期视觉形式加工的一部分,也为后续语义表征的激活提供必要输入。相比之下,正字法匹配任务可以主要依据表层字形信息完成,不必以充分的语义通达为任务目标。由此,汉字正字法与语义之间的紧密联系可能使词汇-语义通达任务更稳定地维持对正字法证据的需求,从而降低高先验概率对 N200 增强效应的削弱作用。

通过这一修改,我们将原先较强的“任务差异被拉大”表述调整为更审慎的“词汇-语义通达任务中的正字法证据需求更稳定”解释,使相关论述更加符合现有文献依据和本研究结果。

意见 5: 第 10 页完整列出了后验概率计算公式,并分别解释了 $P(W)$ 、 $f(I|W)$ 、边缘概率等符号含义。但目前仅停留在定性描述层面,未结合本研究三个自变量(先验概率、任务复杂度、干扰水平)做变量的相关性逻辑推导。建议增强三个实验变量跟参数之间的关系,强化理论与实验的关系。

回应: 非常感谢审稿专家指出后验概率公式仅停留在定性描述、未将三个自变量与公式参数作相关性逻辑推导的问题。我们认同这一意见。原稿虽然列出了后验概率公式,并解释了 $P(W)$ 、 $f(I|W)$ 与边缘概率的含义,但没有充分说明实验操纵、理论构念和公式参数之间的对

应关系。根据审稿专家的建议，我们在引言中按照“实验操纵—理论构念—公式参数”的三层关系重新补充相关论述。

第一，在公式解释部分，我们先明确理论构念与公式参数之间的关系。先验概率对应公式中的 $P(W)$ 。任务复杂度对应感知证据累积量的阈值终止机制中的决策阈值。干扰水平对应公式分母所表征的边缘概率。在此基础上，修订稿进一步说明三者如何影响感知证据需求。先验概率越高，后验概率基线越高，达到决策阈值所需的感知证据越少。任务复杂度越高，决策阈值越高，所需累积的感知证据越多。干扰水平越高，候选字形之间的竞争越强，边缘概率越大，目标字形的后验概率越容易被稀释，系统因而需要积累更多特异性感知证据。修改详见第 2 页，第 22 行到第 3 页，第 7 行。

第二，在研究假设部分，我们将三项 N200 预测分别锚定到上述三个构念及参数。先验概率升高时，所需感知证据减少，N200 增强效应应减弱。任务复杂度增加时，决策阈值提高，所需感知证据增加，N200 增强效应应增强。干扰水平提升时，边缘概率增大，目标字形后验概率被稀释，系统需要更多特异性感知证据，N200 增强效应应增强。修改详见第 3 页，第 27 行到第 4 页，第 2 行。

第三，在研究设计部分，我们进一步明确实验操纵如何对应理论构念和公式参数。实验 1 以重复概率操作先验概率，并对应公式中的 $P(W)$ ；以任务类型操作任务复杂度，对应阈值终止机制中的决策阈值。实验 2 以正字法相似字操作干扰水平，对应公式分母所表征的边缘概率。通过这一调整，读者可以清楚看到重复概率、任务类型和正字法相似字这三个实验操纵如何分别对应先验概率、任务复杂度和干扰水平，并进一步对应 $P(W)$ 、决策阈值和边缘概率。修改详见第 4 页，第 29 行到第 5 页，第 9 行，第 5 页，第 13 行到 21 行。

通过上述修改，修订稿形成了“实验操纵—理论构念—公式参数—感知证据需求—N200 增强效应预测”的完整链条，使贝叶斯阅读器模型与本研究实验设计之间的关系更加清楚。

意见 6：引言、分实验讨论、总讨论中多次重复以下内容：两实验的互补关系、词汇-语义通达任务与正字法匹配任务的分类依据、单字与双字词加工差异不影响核心结论。建议删减交叉重复段落，使各板块功能分明：引言负责背景和设计逻辑，分实验讨论负责各自结果的直接解读，总讨论负责整合与升华。

回应：非常感谢审稿专家指出引言、分实验讨论与总讨论之间存在内容交叉重复的问题。我们认同，原稿对两个实验的关系、两类任务的分类依据以及单字与双字词的加工共通性进行了多次说明，虽然这些内容分别服务于实验设计和理论解释，但重复呈现削弱了各部分的功

能区分。根据审稿专家的建议，我们对相关内容进行了删减和重新组织。

首先，引言部分主要保留研究背景和实验设计逻辑。修订稿简要说明重复概率、任务类型和正字法相似字如何分别对应先验概率、任务复杂度和干扰水平，并交代实验 2 改用单字材料的设计原因。我们删除了引言中对两个实验联合证据及单字与双字词加工共通性的扩展论证，避免在介绍实验设计时提前展开结果整合。

其次，两个分实验讨论主要用于解释各自的结果。实验 1 讨论集中说明重复概率对 N200 增强效应的作用具有任务依赖性，并明确其无法区分物理特征重复概率与字词水平重复概率的独立贡献。实验 2 讨论则围绕识别竞争状态改变后 N200 增强效应的结果模式展开，仅以必要的跨实验对照说明其相对于实验 1 的变化，不再重复两类任务的分类依据、单字与双字词的加工关系及两个实验的整体理论贡献。

最后，总讨论集中承担跨实验整合与理论解释。4.1 节统一讨论两个实验的证据关系及其解释边界。4.2 节压缩了两类任务具体操作和加工流程的重复介绍，仅保留任务复杂度通过决策阈值影响感知证据需求的机制解释。修改详见第 22 页，第 23 行到第 27 行。4.4 节合并了两个实验设计差异及单字与双字词加工关系的相关内容，集中说明不同刺激类型下 N200 增强效应的共同性质、跨刺激类型的会聚证据及跨实验比较的限制。修改详见第 24 页，第 22 行到第 25 页，第 2 行。

通过上述调整，引言、分实验讨论和总讨论分别承担设计说明、结果解释和跨实验整合的功能，相关论述更加集中，也减少了不同部分之间的交叉重复。

意见 7：正文实验 1 2.2.2、实验 2 3.2.2 ERP 结果部分部分。作者已在方法部分完整报告各条件下有效试次数，并说明有效试次不足的被试剔除标准。建议在实验 1、实验 2 的 ERP 结果段落之前，补充各实验条件有效试次数的组间统计比较结果，明确说明各组有效试次数无显著差异。该补充可彻底排除因不同条件信噪比不均，进而造成 ERP 波幅出现系统性偏差的可能性，进一步强化数据质控。

回应：非常感谢审稿专家关于进一步比较各条件有效试次数的建议。根据该建议，我们使用逐被试、逐条件的有效试次数，对两个实验分别进行了重复测量方差分析。

实验 1 以有效试次数为因变量，进行了 2（重复类型：首次呈现、重复呈现） $\times$ 2（重复概率：高、低） $\times$ 2（任务类型：真假词判断、相同-不同判断）的重复测量方差分析。结果显示，各因素的主效应及交互作用均未达到显著水平， $p_s > .109$ 。修改详见第 10 页，第 28 行到第 11 页，第 2 行。

实验 2 同样以有效试次数为因变量,进行了 2 (重复类型:首次呈现、重复呈现) × 2 (干扰水平:高、低) × 2 (任务类型:语义判断、相同-不同判断) 的重复测量方差分析。结果显示,各因素的主效应及交互作用均未达到显著水平, $p_s > .149$ 。修改详见第 17 页,第 15 行到第 18 行。

上述结果表明,两个实验的各条件之间均未发现有效试次数的系统性差异,因而所观察到的 ERP 波幅差异不太可能由条件间有效试次数及信噪比不均衡造成。我们已将相关分析结果分别补充至实验 1 和实验 2 的 ERP 结果之前,以进一步完善数据质量控制的报告。修改详见第 17 页,第 15 行到第 18 行。

第三轮

审稿人 2 意见:

作者已较好得回答了前面提出的问题,建议接收。

编委意见:

论文所研究问题有理论意义,研究方法前沿。作者针对审稿意见和建议做了充分的修改,两位评审专家都表示满意。建议接受发表。

主编意见:

同意发表。