

第 1 轮修改审稿意见及回应

(意见已标记为蓝色)

审稿人 1:

给作者的意见:

本文综述了高效编码约束的贝叶斯推理模型在视知觉研究中的应用，模型在解释视知觉偏差、先验解构及特殊群体研究中的应用具有重要价值。然而，文章在某些方面仍需进一步改进和完善。若作者能针对上述意见逐一修改，文章将更具影响力和可读性。

回复: 感谢审稿人在百忙之中对稿件的详细审阅，并提供了颇具建设性的建议。本文已根据您和其他两位审稿人的建议对文章进行了修订。但《心理学》有严格的版面限制，很多细节无法保留，请见谅。具体回复请见下文。再次感谢您的时间与宝贵建议！

具体意见

1. 文章开篇对“视知觉偏差”的普遍性和研究挑战缺乏系统阐述，导致研究动机不够突出。对“高效编码假说”和“贝叶斯推理”的背景介绍较为简略，未充分说明二者为何需要结合，以及这种结合在视知觉研究中的独特价值。非专业读者可能难以理解该模型的必要性。

回复: An excellent suggestion! 我们完全同意您的看法，清晰阐述视知觉偏差的普遍性、现有理论的局限性以及新模型结合的必要性，对于突出研究动机至关重要。我们已根据您的建议，对前言部分进行了重点修改，具体如下：

- 强化了视知觉偏差的普遍性与理论挑战：开篇即点明视知觉偏差并非个别现象，而是贯穿于多种视觉属性的普遍规律，并指出构建一个统一计算模型来解释这些多样化偏差是该领域长期面临的核心挑战。
- 深化了“高效编码假说”与“贝叶斯推理”的背景介绍：我们分别补充说明了这两个理论框架的优势与各自局限性。明确指出经典贝叶斯模型侧重于“解码”阶段的先验知识利用，但忽略了编码阶段的生物学约束；而高效编码假说虽解释了神经资源的优化分配，却未直接关联到最终的知觉内容。
- 突显了二者结合的必要性与独特性：我们明确指出，将二者结合恰恰是为了弥补单一理论的不足。这种结合构建了一个从“神经编码”到“知觉解码”的完整因果链条，为多种知觉偏差提供了统一的、具有生物学合理性的计算解释，这是其最核心的理论价值。

具体内容请见引言前三段。我们相信，经过上述修改，文章的开篇已经能够清晰地勾勒出研究领域的问题全景、现有理论的空白，以及本文所述模型的创新性与必要性，从而为后续内容奠定了更坚实的逻辑基础。再次感谢您的深刻见解，帮助我们提升了论文的质量。同时我们也静候您进一步指导！

2. 文章提出贝叶斯理论框架发展经历三个重要阶段，三个阶段划分的依据是什么呢？

回复：我们在本次修改中，取消了“三阶段”的表述。严格按照模型间的递进关系进行讲述。Cheers！

3. 关于模型部分：

1) 文章中如何定义高效编码约束的贝叶斯模型？反应映射是否包涵在这个模型当中？

回复：根据您的第三个建议，通过绘图的形式，在第二章节开篇就展示了所论述的模型结构，很好地帮助读者理解我们的模型。感谢您的建议。

是否可以增加不同模型之间的对比？以及高效编码约束的贝叶斯模型相比其他模型有哪些具体的提升或者优势。

回复：完全理解您的建议，也完全同意。如果能够用单独的章节或段落点明模型优势更有利于读者理解和使用模型。我们现在主要是在论述过程中讲解不同模型或模型阶段的作用，比如：

在第二章中，2.1 节中，我们点明高效编码部分可以解释“可以发现，物理量的估计值会偏离真实值，并且远离先验分布的峰值，这一现象正是排斥性偏差的数学表征（如图 1b 中红色和蓝色实线描述的趋势，Wei & Stocker, 2015, 2016, 2017）。”

在 2.2 节中，“基于贝叶斯解码的理论模型能够很好地解释人类对多种视觉属性的感知偏向先验峰值的现象（如图 1d 中黑色实线描述的趋势），如颜色（Olkkonen, McCarthy, & Allred, 2014）、时间（Jazayeri & Shadlen, 2010）、深度（Burge & Geisler, 2011）、速度（Stocker & Simoncelli, 2006）、基于光流的自身运动方向（Xu et al., 2022）等。”

在 2.3 节中，“针对这一问题，Sun 等人（2025b）通过系统性实验揭示了反应范围的关键调控作用：当反应范围较小时，被试表现出向心偏差（贝叶斯偏差）；而反应范围较大时，则转为离心偏差（反贝叶斯偏差）。基于这一发现，研究者在经典高效编码贝叶斯框架（Wei & Stocker, 2015）基础上，引入了“知觉-反应映

射” 计算模块, 构建了三级联的 “反应-映射调控的高效编码约束贝叶斯推理模型”。”

虽有涉及模型比较, 但不够明显, 因此根据您的建议, 我们对 2.4 小结进行了提升, 突出模型比较和价值。请见: “这一整合模型在解释范围与机制深度上显著超越了早期理论。相较于经典贝叶斯模型 (Knill & Richards, 1996), 本框架不仅涵盖其对吸引力贝叶斯偏差 (如中心趋势) 的解释, 更进一步通过高效编码约束, 为排斥性反贝叶斯偏差 (如倾斜效应) 提供了源自神经资源分配的机制性成因, 无需依赖额外假设。与早期高效编码理论 (Atteneave, 1954; Barlow, 1961) 相比, 本模型突破了其仅描述神经调谐特性的局限, 借助贝叶斯解码将编码优化与具体知觉偏差定量关联, 实现了从表征到知觉的跨越。此外, 通过纳入知觉-反应映射机制, 模型能够解释相同刺激在不同任务情境 (如反应范围变化) 下诱发不同偏差模式的现象, 这是以往单一理论所无法涵盖的。”

综上所述, 该模型体系的价值在于将原本分散的计算原则整合为一条从编码约束、推断权衡到行为输出的完整因果链, 为理解知觉提供了统一且生物合理的解释框架。”

希望上述修改可以解答您的问题。

2) 核心理论部分直接介绍是高效编码, 贝叶斯解码和知觉反应映射三个步骤。是否应该先介绍贝叶斯理论框架? 在此基础上添加高效编码约束等模块。

回复: 感谢您提出的宝贵且富有洞察力的建议。我们完全认同您的看法, 即首先确立贝叶斯理论框架, 确实更有助于读者理解模型如何在经典基础上实现创新。

在充分吸收您意见的基础上, 我们结合原文所遵循的“信息流”层级——即从“编码”到“解码”再到“反应映射”的知觉加工顺序——在“核心理论”章节的开头单独增补了一个引导段落。该段落旨在提前阐明全章的结构思路, 在尊重知觉处理层次的同时, 也兼顾理论叙述的清晰性与连贯性。

我们希望这一调整能够在保持原有分析框架的基础上, 更好地引导读者把握整体逻辑。

具体内容请见“本章节中, 对“本章围绕“高效编码约束的贝叶斯推理模型”(图 1a - 1c), 按“编码 - 解码 - 行为映射”的级联逻辑展开。2.1 节

“高效编码”确立了知觉起点, 阐明在资源限制下感官信息如何被优化表征, 奠定了解码阶段输入的基本特性。2.2 节“贝叶斯解码”则探讨大脑如何基于该表征进行推断, 重点说明高效编码约束如何将理想推断转化为资源限制下的实际权衡, 从而统一解释吸引性与排斥性偏差。2.3 节“知觉-反应映射”进

一步揭示任务情境如何调控内部知觉至外显行为的转换，完善感知到行为的完整通路。全章表明，贝叶斯推理构成模型数学核心，高效编码与反应映射则分别从输入与输出端予以拓展，共同构建更具生物合理性的理论体系。”

3) 建议补充一张示意图，直观展示“高效编码-贝叶斯解码-知觉反应映射”三级模型的整体框架，增强可读性。

回复：再此，我们绘制了一张示意图，方便读者理解这个模型。请见：

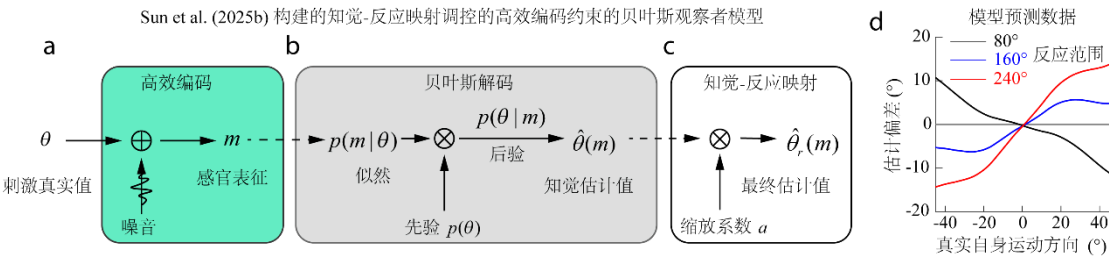


图 1. Sun et al. (2025b) 构建的知觉-反应映射调控的高效编码约束的贝叶斯观察者模型。(a-c) 其包含了“编码-解码-反应映射调控”三个层级，反应了物理刺激到外部反应的信息流向。(d) Sun et al. (2025) 揭示的反应范围对自身运动方向估计偏差的影响。其中,80°反应范围下，表现出中心偏差，其是贝叶斯偏差的典型示例；而当反应范围为 160°和 240°时，表现出离心偏差，其实反贝叶斯偏差的典型示例。

4) 部分公式的意义或直观解释可以更详细，以帮助心理学背景的读者理解。

回复：您的这个建议很重要，但请原谅我们没有做出更进一步的修改。数学语言，特别是相关符号和信息论的相关体系很庞大，很难再进一步通俗化。现在的读者可以借助大语言模型可以容易读懂。感谢您的理解。

4. 应用方面：

回复：以下 5 个建议非常全面且有深度。我们根据您的建议依次对文章章节做了修改。

1) 4.1 节提到跨通道研究，但未具体说明如何解决不同感官系统编码效率差异的挑战，建议补充可能的解决方案（如多模态信息整合模型）。

回复：对第 4.1 节做了如下修改：

1. 我们明确指出了“量化并整合不同感官模态在编码效率上的固有差异”是模型拓展面临的核心挑战。
2. 我们提出了一个具体的解决方案：借鉴并整合经典的“多模态信息整合的贝叶斯模型”。我们详细说明，该模型通过基于各感官通道编码效率（费舍尔信息 $J_m(\theta)$ 或感知不确定性 σ_m^2 ）的权重 w_m ，来线性加权各通道的感官

证据，从而形成最优的跨通道后验估计。这一机制得到了大量实证研究的支持（如 Ernst & Banks, 2002）。

3. 我们进一步将本框架的“知觉-反应映射”机制与跨通道整合联系起来，提出它可能负责协调各加权重后的通道估计，形成统一的多模态知觉。这为未来的研究提供了一个清晰且可检验的理论假设。

请您审定。

2) 在 4.2 节作者提出“需要揭示社会认知先验对行为的塑造是否遵循高效编码原则”，但社会认知和知觉加工过程有很大不同，需要具体解释社会认知中的高效编码理论原理及其假设。目前描述过于笼统。

回复：对第 4.2 节做了如下修改：

1. 明确定义了社会认知语境下的“高效编码原则”：我们将其具体阐述为“大脑会调整其社会特征编码神经元的群体反应特性，使得有限的神经资源优先用于区分在社会环境中最常出现或行为上最关键的社会刺激”。
2. 提出了三个可检验的核心理论假设：为了将这一原则操作化，我们推导并明确列出了三个具体的、可供未来研究验证的假设：
 - （1）神经资源分配假设：关于编码资源如何依赖社会先验进行分配。
 - （2）表征扭曲与偏差假设：关于资源非均匀分配如何通过贝叶斯解码导致可预测的社会判断偏差（如信任度评估）。
 - （3）跨领域一致性假设：关于社会认知与知觉是否共享相似的优化数学规律。
3. 指明了关键研究方法：我们强调了构建此类模型的关键在于为抽象社会特征定义其特征空间、量化其先验分布，并测量其行为辨别力，从而将理论探讨与可行的实证路径联系起来。

3) 4.3 节中并未体现高效编码约束的贝叶斯模型的应用，更多是关注不同障碍的特异性编码缺陷，可否更详细阐述以体现高效编码约束的贝叶斯模型价值。

回复：对第 4.3 节做了如下修改：

1. 突出了模型的统一解释框架价值：开篇即强调该模型为建立“跨诊断计算精神病理学框架”奠定了基础，将其定位从一个应用实例提升为理论工具。
2. 具体阐释了模型在不同障碍中的应用机制：我们不再仅仅提及 ADHD、阅

读障碍和焦虑障碍，而是深入且具体地阐述了该模型如何为每一种障碍提供独特的、可量化的计算假设：

将 ADHD 的注意分散与感觉噪声升高联系起来，并关联其与先验依赖减弱的理论预测。

将阅读障碍的核心困难与特定维度编码效率降低联系起来，并关联其与辨别阈值的数学关系。

将焦虑障碍的敏感性解释为对威胁刺激编码资源分配的异常或畸变的先验。

3. 明确了模型的方法论价值：结尾处指出，该模型的核心应用在于通过“逆向工程”和“计算表型”分析，量化不同群体的特异性参数配置，从而为实现精准的计算诊断与干预提供路径。

4) 4.4 节关于神经机制的描述较笼统，可增加具体的技术路径（如 fMRI 与单细胞记录的结合）或潜在的研究设计。

回复：对第 4.4 节做了如下修改：

1. 明确了多模态技术的协同路径：我们具体提出，可以结合 fMRI 的高空间定位能力与 MEG/EEG/单细胞记录的高时间分辨率，以此来分别刻画模型不同计算组件（如费舍尔信息、后验分布）的神经表征及其动态过程。
2. 设计了计算模型驱动的实验逻辑：我们阐述了如何通过系统性地操纵先验、噪声和反应范围等核心计算变量，在记录神经活动的同时，将“编码-解码-映射”三个阶段的神经基础进行有效分离和鉴定。
3. 引入了从相关到因果的验证环节：我们特别强调了如何利用 TMS 或光遗传学等神经扰动技术，来检验计算模型做出的特定预测，从而为神经机制提供因果性证据，而不仅仅是相关关系。

5. 文章中并未讨论该模型是否有局限性，比如该模型如何解释一些常见现象如知觉适应（adaption），重复抑制（repetition suppression）和期望抑制（expectation suppression）。

回复：很有深度建议！引发了我们的深入思考。的确很难穷尽，未来值得做的事情很多，当前篇幅已经严重超出了期刊要求，因此，对于该建议，我们在 4.5 小结只是简单提及“同时，本文主要论述的是静态刺激知觉，该模型对与动态知觉现象，如重复抑制（repetition suppression）和期望抑制（expectation suppression）(Aukstulewicz & Friston, 2016) 等是否具有同样效力，亦需要认真

思考。”未能进行详细、深入讨论，还望见谅。

6. 一些小意见：

1) 提升文章可读性，部分长句逻辑稍显复杂，建议拆分以提升可读性。

回复：本次修改中，我们也对语言进行了适当调整，让语言更加流畅，易读。

2) 提升术语使用基本规范，如“反贝叶斯偏差”等术语首次出现时可加简要说明。为何要将排斥效应称作反贝叶斯偏差，是直接翻译自 Xuexin Wei 文章么？

回复：Good suggestion. 我们在必要的术语方面，增加了英文术语。

3) 公式 7 中的 χ

回复：此处符合数学规范，不过考虑到心理背景的读者，我们增加了额外解释。

请见“其中， χ 为积分哑变量，在特征空间上积分，其物理意义与 θ 一致。”

4) Sun 等人(2025)有两个文章，需要进行区分，可以使用 Sun 和 Gong 等(2025)与 Sun 和 Xu 等(2025)作为区分

回复：我们用 a 和 b 进行了区分。Cheers。

Sun, Q., Gong, X. M., & Sun, Q. (2025a). Impact of conflicts between long-and short-term priors on the weighted prior integration in visual perception. *Cognition*, 254, 106006.

Sun, Q., Xu, L. H., & Stocker, A. A. (2025b). A linear perception-action mapping accounts for response range-dependent biases in heading estimation from optic flow. *PLoS Computational Biology*, 21(6), e1013147.

再次感谢您的审稿建议!!! 您的建议对提升文章整体深度和完整性提供了巨大帮助！但受制于期刊要求，很多内容无法进行深入细致论述，再次表达歉意。

（意见已标记为蓝色）

审稿人 2:

给作者的意见:

本综述主要聚焦视知觉研究中，目前前沿且极具未来探索意义的“efficient coding”和“贝叶斯推理模型”两方面，并将这两个有内部联系的建模方法进行链接，最后探索和展望了他们的应用前景。文章撰写流畅清晰，主要观点介绍得当，论文引用合理，内容深度和可读性平衡较好。应用的拓展得当，很方便其他研究者进一步在本领域内进行应用研究。我个人认为撰写出色。若要提意见，希望作者可以在以下几个小方面思考一下：

回复：感谢审稿人在百忙之中对稿件的详细审阅，并提供了颇具建设性的建议。本文已根据您和其他两位审稿人的建议对文章进行了修订。具体回复请见下文。再次感谢您的时间与宝贵建议！

1. 本综述涉及的内容比较深，有一些数学要求。我觉得介于篇幅和内容，肯定不方便列出更多更细节的推导公式，但是作者可以考虑专门拿出一段介绍一下如何用软件和特定工具包（开源的最佳）实现上述分析和推导，方便应用。

回复：我们提供了一份开源代码，详见：

https://osf.io/t47v6/overview?view_only=b8467965b39947b88342352c8c124b2b。

感谢您宝贵建议。

2. 部分过程可能介绍过于简单，比如上述模型在序列依赖为代表的多个视觉认知过程的论文，可以适当更多一些，更丰富一些。比较一下他们采用的不同的模型和分析。

回复：很好的建议。已根据您的建议增加了相应的讨论：“该框架的解释力进一步延伸至高阶社会知觉领域。近期关于序列面孔加工的研究显示，视觉系统在处理面部社会特质（如吸引力、信任度与支配力）时，其序列依赖性同样呈现出复杂的系统特性，并可借助类似的计算思路（隐藏马尔科夫链）进行建模。例如，Yu 和 Ying (2025) 发现，当被试在不同面部特质（如吸引力与支配力）之间切换判断时，序列依赖性会受到破坏，并违反马尔可夫假设，表明不同特质判断可能依赖于不同来源或权重的先验信息，与前述的分层整合机制相契合。更为关键的是，Yu, Yang 和 Ying (2023) 的研究直接探索了时空信息的整合机制，他们运用类似模型，将群体面孔感知中的空间对比效应（即“朋友效应”）建模为驱动序列判断的“隐藏状态”，从而在一个统一模型中揭示了面孔知觉中时空统计规律的整合方式，为理解视觉系统如何动态协调多尺度环境信息提供了重要范例。”

3. 部分内容来自于其他数理学科。可能有些知名学者翻译可以参考通用名称，比如香农。

回复：已校对相应的翻译及文章出现的错别字。

（意见已标记为蓝色）

审稿人 3:

给作者的意见:

本研究系统梳理了视知觉研究领域基于高效编码约束的贝叶斯推理模型，全面阐述了该模型的研究背景、理论基础、计算逻辑及其主要变式。通过与其他视知觉模型比较分析，重点探讨了该模型在解释视知觉现象方面的独特理论优势。作为当前视知觉研究领域的前沿理论之一，该模型在国内学术界尚未得到充分关注，确需通过系统的理论引介推动其在国内研究中的应用与发展。

本研究具有以下贡献：首先，对模型计算逻辑进行了系统而深入的梳理，这一基础性工作为读者理解和掌握该模型提供了重要支撑，具有理论价值；其次，基于模型对视知觉完整计算过程（编码-解码-映射）的系统建模特征及生理学基础，探讨了该模型在特殊群体研究中的应用前景；最后，通过对模型未来发展方向的系统展望，为后续研究提供了理论指引。

回复: 感谢审稿人在百忙之中对稿件的详细审阅，并提供了颇具建设性的建议。本文已根据您和其他两位审稿人的建议对文章进行了修订。具体回复请见下文。再次感谢您的时间与宝贵建议！

本文的不足，主要体现在以下几个方面：

（1）语言稍显晦涩，特别是计算理论部分，需要有较好的数学训练。建议作者增加一个开源代码，辅助读者理解并掌握计算逻辑。

回复: 我们提供了一份开源代码，详见：

https://osf.io/t47v6/overview?view_only=b8467965b39947b88342352c8c124b2b。

感谢您宝贵建议。

（2）注意校对文章中的错别字；

回复: 我们对文章的错别字进行了校对。

（3）对语言流畅性再进一步思考，比如“在朝向编码方面，自然图像中水平和垂直朝向的高频出现 (Girshick, Landy, & Simoncelli, 2011) 与 V1 区对应朝向神经元数量优势及更窄的调谐曲线 (Mansfield, 1974) 相匹配”这一句理解起来并不顺畅；

回复: 很好地建议，我们首先对本自然段进行了修改，详见“这一理论框架为理

解神经调谐特性与环境统计的关联提供了数学基础。多项实证研究为该理论提供了有力支持。在朝向编码方面，自然图像中水平和垂直朝向出现频率较高 (Girshick, Landy, & Simoncelli, 2011)，这与初级视觉皮层 (V1) 中对应朝向神经元数量更多、调谐曲线更窄的生理发现 (Mansfield, 1974) 相一致。在空间频率编码方面，自然场景中低频成分占主导的统计特性 (Van Hateren & Van der Schaaf, 1998)，也与 V1 区低频神经元数量较多、辨别能力更强的生理证据 (Cavanaugh, Bair, & Movshon, 2002) 相吻合。进一步地，Ganguli 和 Simoncelli (2011) 通过建立费舍尔信息优化模型，进一步验证了环境统计特征与神经元调谐特性之间的定量关系，从而为高效编码假说提供了坚实的理论与实验证据。”

同时，我们也对文章的其他部分进行了仔细修订，希望对文章的语言质量有进一步提升。

(4) 当前参考文献中全部为英文文章，虽然都是高水平期刊，但有没有可能适当增加相关中文研究的讨论，进而推动该方向或研究方法在国内的传播。

回复：很好的建议！但不得不承认，当前该模型在国内期刊中尚未有文章发表，这也是作者撰写本文的目的之一，推动相关研究在国内的传播与模型应用。如您或后续有其他推荐研究，我们会在后续修订中增加。感谢您的建议。

