中国科学: 生命科学 2025 年 第 55 卷 第 7 期: 1395 ~ 1408
SCIENTIA SINICA Vitae lifecn.scichina.com
《中国科学》杂志社 SCIENCE CHINA PRESS
评 述

多感觉整合中因果推断的计算模型及神经机制

程羽慧[1*], 袁祥勇[2,3], 蒋毅[2,3], 曹一男[4,5*]

1. 南京师范大学心理学院, 南京 210024
2. 中国科学院心理研究所, 脑与认知科学国家重点实验室, 北京 100101
3. 中国科学院大学心理学系, 北京 100049
4. Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg 20246, Germany
5. Department of Cognitive Studies, Ecole Normale Superieure-PSL, Paris 75005, France
* 联系人, E-mail: chengyh@nnu.edu.cn; yinan.cao@ens.psl.eu

收稿日期: 2024-07-31; 接受日期: 2024-11-06; 网络版发表日期: 2025-05-27
国家自然科学基金(批准号: 32400864)、南京师范大学引进人才科研启动基金(批准号: 184080H201A45)和欧盟玛丽·居里学者人才计划(MSCA)独立研究基金(批准号: 101154160)资助

摘要 人类和其他动物的大脑时刻接收来自外界复杂环境中的多感官信息. 为快速有效地感知并做出决策和反应, 大脑需要整合共同来源的感官信号, 并分离独立来源的感官信号. 因此, 生物体需要解决多感觉信息处理中的“因果推断”问题. 本文综述多感觉信息处理领域中从早期强制整合模型到近期贝叶斯因果推断模型的发展历程, 总结多感觉因果推断中常用的实验范式及基本计算原理. 在神经层面, 大脑动态编码因果推断信息, 呈现层级递进加工, 具体体现为初级感觉皮层快速表征单模态信息, 随后在顶颞区强制整合, 最终达到额顶区进行因果推断. 最后, 未来研究可以整合多种实验范式, 探索特殊人群中因果推断的神经机制, 克服贝叶斯因果推断模型的局限性, 并关注神经网络在解释其生理机制方面的潜力.

关键词 多感觉整合, 贝叶斯因果推断, 层级加工, 计算模型, 神经网络

1 多感觉整合

在研究初期, 心理学家与神经科学家通常只关注单一感官模态, 如视觉研究者仅关注视觉加工, 而听觉研究者仅关注听觉加工. 但在现实生活中, 人类大脑不断接收来自不同感觉通道的信号, 包括视觉、听觉、触觉、嗅觉和味觉信息等. 只有当这些信息在大脑中被协调地整合与利用时, 人类才能获得丰富连贯的感知体验[1]. 因此, 也有许多研究者从多感觉整合的视角对感知进行研究[2]. 将多个分离的单感觉信息整合为连贯统一的多感觉信息的加工过程, 被称为多感觉整合(multisensory integration)[3,4]. 它不仅能弥补单一感官模态信息的不完整性, 还可以增强感觉信号的显著性, 从而增强人类对信息的感知与响应能力[5,6].

1.1 多感觉整合的强制整合模型

早期的经典研究表明, 大脑在整合多个感觉信号时遵循“可靠性加权原理”[7], 即大脑对信息进行强制整合, 在整合时更多依赖可靠性高的感觉信号, 而较少依赖可靠性低的信号[8]. 举例来说, 人类在过马路时

引用格式: 程羽慧, 袁祥勇, 蒋毅, 等. 多感觉整合中因果推断的计算模型及神经机制. 中国科学: 生命科学, 2025, 55: 1395–1408
Cheng Y H, Yuan X Y, Jiang Y, et al. Computational models and neural mechanisms of causal inference in multisensory integration (in Chinese). Sci Sin Vitae, 2025, 55: 1395–1408, doi: 10.1360/SSV-2024-0160

需要判断汽车的行驶位置. 在白天光照充足的条件下, 视觉系统通常更为可靠, 因此大脑更多依赖于视觉信号来判断汽车位置; 但在黑夜低光条件下, 视觉信号的可靠性降低, 于是大脑会赋予听觉更高的权重, 主要通过听到汽车引擎的嗡嗡声来判断其位置. 这个例子很好地说明大脑在处理感觉信息时的动态适应性, 以及如何在不同环境条件下灵活地整合多种感觉信号来改善感知和行为[9].

1.2 多感觉整合的因果推断模型

然而, 值得注意的是, 上述多感觉加工过程是一种强制整合, 即大脑对所有信号均进行整合利用. 但是大脑是否永远被动地遵循这单一原则呢? 以上述例子来进一步说明, 如果感知到的视、听信号来自同一辆汽车, 那么大脑应该将这些信号整合. 但如果听到的声音源于其他车辆, 那么大脑应该将这两个信号分离, 以避免错误地整合. 这表明多感觉整合存在特定的前提条件: 只有在判断多个感觉信号来自同一客体时, 整合才具有意义. 在复杂的现实情况中, 感觉信号可能有共同的来源, 也可能有不同的来源, 强制性地整合感觉信号可能会导致信息错误地分配[10]. 因此, 大脑所面临更为重要的挑战是正确分析不同感觉信号背后隐藏的因果结构(causal structure), 以确定哪些感觉信号来源于同一客体, 应被整合; 而哪些感觉信号来自不同的客体, 应被分离. 这种通过判定感觉信号的来源来决定整合还是分离的过程, 被称为“因果推断(causal inference)”[11,12]. 不难发现, 单感觉独立加工和多感觉强制整合实际上是因果推断中的两种特定情境. 因果推断在计算本质上实现从单感觉独立加工, 到强制整合, 再到“灵活整合”的连续统一体, 因此是当前多感觉研究领域一个核心议题.

因果推断自从被提出以来, 其计算原理和神经机制一直是多感觉研究中备受关注的热门问题. Körding等人[12]首次通过计算建模从行为层面提出多感觉整合中的贝叶斯因果推断模型. 随后, Rohe等人[13]突破性地为该模型提供神经层面的证据. 近年来, 随着神经成像和神经电生理证据的涌现, 认知神经科学研究者们进一步揭示因果推断在多感觉整合中动态层级加工的神经表征[14~17]. 本文旨在全面回顾因果推断在多感觉整合中的计算原则和神经机制, 首先总结典型的实验范式, 接着阐述因果推断的基本原理和影响因素, 进而深入探讨因果推断的神经机制, 最后提出对未来研究的展望. 目前, 随着脑成像技术和计算模型的不断发展, 当前这一研究领域逐渐成为学术界的一个新兴热点. 本文通过从因果推断的视角揭示多感觉整合的计算原则和神经机制, 不仅有助于深化对多感觉整合理论的理解, 还有望进一步促进人工智能技术的优化, 提高智能系统在复杂多感觉环境中的感知能力.

2 因果推断的实验范式

在人类实验中, 研究因果推断的实验范式主要集中于经典的视听任务. 由于大脑无法直接获取事件的因果结构, 人类必须依赖于信号在空间、时间或语音、语义上的一致性或者关联性来进行判断, 这些线索也常作为不同实验范式的操作变量. 因此, 研究者通常采用空间定位任务(腹语术效应, ventriloquist effect)[13], 闪光声音数量判断任务(声音诱发闪光效应, sound-induced flash)[18]或者速率判断任务[14], 以及语音判断任务(麦格克效应, McGurk effect)来探究视听整合中的因果推断[19]. 具体而言, 多感觉信号在空间位置中的一致性会影响因果判断[20,21]. 如果一个声音信号与一个视觉信号来源于相同或相近的位置, 大脑通常会将声音与视觉事件整合起来. 最经典的就是“腹语术效应”, 即腹语术师通过控制木偶的嘴部动作, 营造出“木偶在讲话”的假象. 观众通常会错误地将表演者自己的声音与木偶的嘴巴联系起来[8]. 此外, 大脑会依据不同感觉事件之间的时间同步性(temporal synchrony)来确定哪个事件可能是引起某个感知状态的原因[22,23]. 当两个事件在一个很小的时间窗内相继出现甚至同时出现时, 大脑会认为它们之间存在因果关系. 大脑在时间分辨上的能力是有限的, 因此会导致被试产生错觉, 例如“声音诱发闪光”效应[24], 即当单个视觉闪光与两个快速连续的声音信号同时出现时, 观察者倾向于感知到两次闪光[25]. 同样, 语音、语义信息的一致性也会影响因果推断[26,27]. 一个经典的例子是“麦格克效应”, 当人们看到某人的嘴型形成特定音节(如‘ba’)而听到的却是不同(但足够相近)音节(如‘ga’)时, 他们往往会错误地感知为介于两者之间的第三个音节(如‘da’)[28]. 在上述实例中, 这些信号均被整合; 然而, 当这些信号在时间、空间或者内容上变得足够不相似或者不相关时, 大脑就会停止信息整合, 并认

为这些刺激信号来源于不同事件或客体.

在实验室中, 研究者往往通过操纵空间、时间、语义等相关变量来影响因果判断. 以经典的腹语术效应任务为例, 被试同时接收短促的听觉信号(白噪声)和视觉信号(高斯分布的点状云团). 这些信号在四个可能的方位角(−10.5°, −3.5°, 3.5°或10.5°)随机组合呈现. 被试需要分别报告他们感知到的视觉或听觉信号的位置(图1A和B). 由于在空间定位任务中视觉可靠性高, 而听觉可靠性相对较低, 被试所判断的听觉位置往往会受到视觉位置的影响(即跨模态偏差). 有趣的是, 当视觉信号的可靠性被人为降低时(例如增大点状云团高斯分布的标准差), 听觉受视觉的影响也相应降低, 这恰好印证可靠性加权原理[13]. 而更关键的是, 对于较小空间差异的视、听信号, 被试会遵循可靠性加权的原理进行强制整合, 从而导致跨模态偏差效应; 而对于较大的跨感官冲突(即听觉信号与视觉信号间差距显著), 被试往往会倾向于将两个信号看作是独立的, 进而将其分离. 在这种情况下, 可靠性加权原理不再适用, 跨模态偏差效应减弱(图1C左)[12,13]. 此外, 研究者也经常直接让被试判断视听信号是否有共同的来源. 他们发现, 视听信号越接近, 他们报告为共同来源的概率更高, 随着视、听信号差异变大, 他们报告为共同来源的概率逐渐降低(图1C右)[29]. 有趣的是, Wallace等人[30]发现人类对信号进行空间判断时会受到其如何判断信号来源的影响. 当被试推断视听信号来自同一位置时, 他们感知到的听觉位置会偏向视觉, 即发生视听整合. 而当被试推断出两个信号来源于不同位置时, 他们感知到的听觉位置不受视觉的影响(甚至发生逆转). 可见, 不论是基于内隐的(implicit)位置判断任务还是外显的(explicit)同源判断任务, 这些实验结果都证明人类在进行多感觉整合时遵循灵活的因果推断原理[12,29]. 后面本文将详细阐述这种灵活的因果推断如何符合贝叶斯统计推断的原则.

除经典的视听任务外, 因果推断模型也适用于其他多感觉判断任务[31], 这一部分具体可参考章节5的内容. 特别是在猕猴动物实验中, 考虑到猕猴训练任务的复杂性, 研究者通常使用橡胶手错觉任务和航向判断等任务来探究因果推断问题[16,32,33]. 例如, Fang等人[32]的研究采用橡胶手错觉任务, 以探究猕猴整合本体感觉和视觉的因果推断机制. 实验中研究者要求猕猴利用本体感觉将被遮挡的手臂移动到视觉目标来获得奖励, 同时操控其看到的虚假手臂(视觉信号)和真实手臂(本体感觉信号)的空间距离. 实验因变量则为真实手臂和目标位置的角度差异. 他们发现, 当虚假手臂和真实手臂的空间差异较小时, 猕猴会产生身体幻觉, 即认为虚假手臂是自己的手臂, 因此被试在移动手臂时也会向假臂的方向偏移; 当虚假手臂和真实手臂的空间差异变大时, 移动手臂会更少地受到假臂的影响, 所产生的错觉效应也随之变小, 这一现象同样符合灵活的贝叶斯因果推断原理[32]. 总之, 不论采用何种实验范式, 人类和动物在进行多感觉整合任务时, 都似遵循着因果推断原理, 且表现出相似的行为模式.

3 因果推断的基本原理及影响因素

大脑时刻面临着外界纷杂环境中的多种感知输入, 而这些输入可能具有相同或不同的来源. 如上文所述, 大脑利用推断出的感知因果结构, 以便灵活地游走于多感觉信号的分离与整合之间. 认知心理学家们提出贝叶斯因果推断模型以解释这一现象.

3.1 因果推断的基本原理

贝叶斯因果推断模型的基本原理是, 在两种不同因果结构(同源或异源)的特定假设下, 对任务变量(例如刺激位置)进行相应的估计[12]. 这一原理遵循贝叶斯统计推断的基本定理, 即后验概率等于先验概率与似然概率的乘积[34]. 因此, 因果推断主要利用两种概率分布信息. 一种信息是似然概率(likelihood), 即在给定事件(存在特定来源)发生情况下, 受到噪声干扰的刺激被感知为某一观测值的条件概率. 知觉会受到噪声的干扰, 例如神经编码噪声、工作记忆中的噪声、信息判断过程中的噪声等, 因此任何感官刺激都不会被感知为真实情形, 而是在当前刺激条件下可能产生的感知分布. 一般而言, 研究者通常假设每种感官所产生噪音都是独立的, 且符合正态分布. 另一种信息是先验概率(prior), 即基于知识和经验, 判断两个信号具有同一来源的可能性. 在因果推断模型中, 先验是一个相对笼统的概念, 它一般被认为在当前实验时间框架中是相对稳定不变的. 这是一种合理的假设, 比如研究者可以预设在位置判断任务中的先验是实验开始前被试已有的一个预设: p (common), 即视听信号有多

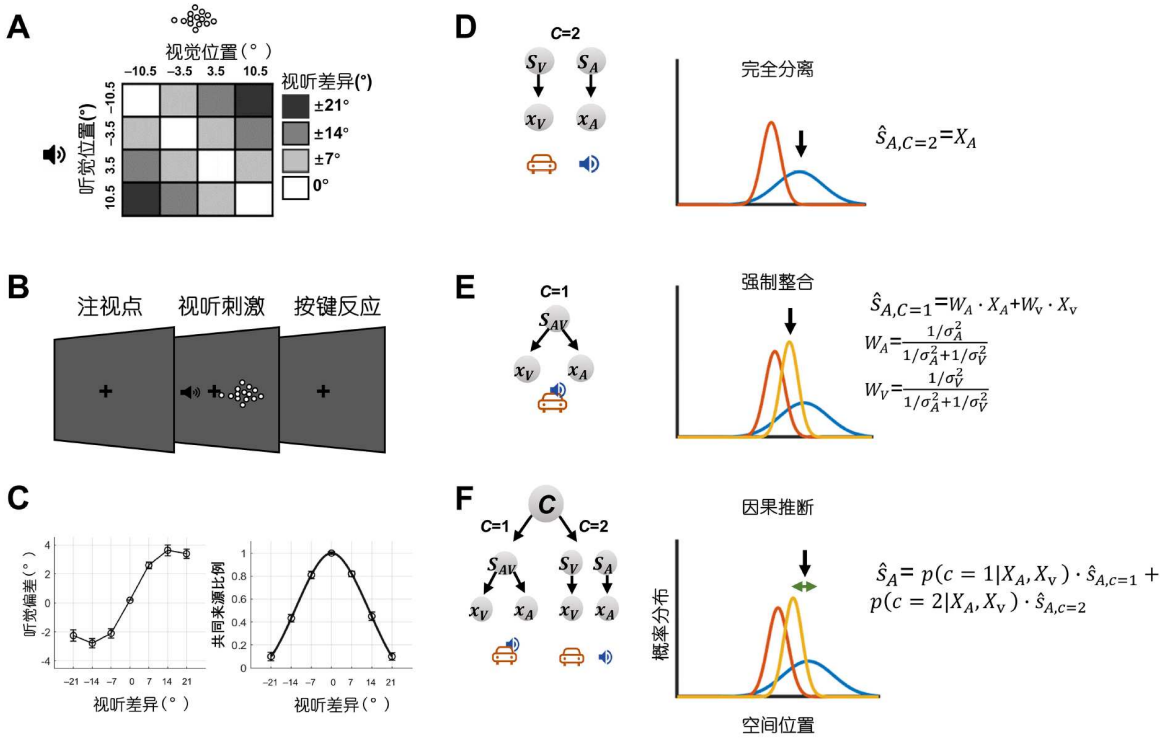

图 1　多感官整合的实验范式和贝叶斯因果推断模型. A~C图分别显示腹语术实验的刺激, 任务示例和预期实验结果. A图为实验中视觉(高斯分布的点状云团)和听觉(白噪声)信号和呈现不同位置的排列组合; B图为视听刺激空间位置判断任务中某试次的示例: 被试需要报告所感知到的声音信号或视觉刺激的位置; C图为作者未发表数据中某被试的实验结果, 反映在不同视听差异的条件下所观察到的被试听觉偏差(内隐任务)和共同来源判断比例(外显任务), 误差棒表示被试内不同试次间的变异. D~F图中左侧显示外界环境中视觉输入(看到一辆汽车)和听觉输入(听到鸣笛声)所产生不同因果结构的示意图. 图中感知输入可能来自同一辆汽车, 也可能来自不同的汽车. 其中S_A, S_V和S_{AV}分别表示客观的听觉、视觉或多感官物理刺激, X_A和X_V表示各自来源所引发的感官表征(即主观空间位置分布); D~F图中右侧显示不同假设下感觉表征的概率分布, 以及根据贝叶斯模型得出刺激位置的最佳估计(由向下的箭头表示)和数学公式. D图假设刺激来源是独立的(C=2), 因此最佳估计值即为每个单感官位置的最佳估计值; E图表示强制整合的情形, 即假设存在一个共同来源(C=1). 在这种情况下, 最佳估计值是视觉和听觉估计的加权平均, 其权重由它们的相对可靠性决定; F图表示在贝叶斯因果推断中, 大脑同时考虑两种不同假设(如共同来源或独立来源)的因果结构, 根据模型平均策略, 其最佳估计值是单感觉最佳估计值(C=1)和整合最佳估计值(C=2)的非线性加权

Figure 1　Experimental paradigm of multisensory integration and Bayesian causal inference model. Figure A–C show the stimuli, task, and expected experimental results of the ventriloquism experiment, respectively. Panel A shows the combination of visual (Gaussian cloud of dots) and auditory (white noise burst) signals and presents different positions in the experiment; panel B represents an example of the spatial location judgement task of audiovisual stimuli: Subjects are required to report the location of a perceived sound signal or visual stimulus; panel C shows unpublished data from one participant, reflecting the auditory bias (implicit task) and the proportion of common-source judgments (explicit task) observed under conditions with different audiovisual differences. The error bars represent the variability across trials for each participant. Panels D–F on the left illustrate the causal structure resulting from different sensory inputs in the external environment, such as seeing a car and hearing a horn. These sensory inputs may come from the same car or from different cars; panels D–F on the right illustrate the probability distributions of sensory representations under different assumptions, along with the optimal estimate of the stimulus location (indicated by the downward arrows) and mathematical formulas derived from the Bayesian model. S_A S_V and S_{AV} represent objective auditory, visual, or multisensory physical stimuli, respectively, and X_A and X_V represent sensory representations triggered by their respective sources (i.e., subjective spatial location distribution). Panel D assumes the stimulus sources are independent (C=2), so the optimal estimate is the best estimate for each individual sensory modality; panel E shows the case of forced integration, assuming a common source (C=1). In this case, the optimal estimate is the weighted average of the visual and auditory estimates, with the weights determined by their relative reliabilities; panel F illustrates that, in Bayesian causal inference, the brain simultaneously considers two different hypotheses (e.g., common or independent sources). Using a model averaging strategy, the optimal estimate is a nonlinear weighting of the best estimate for the individual sensory modality (C=1) and the integrated best estimate (C=2)

大的概率来自同源. 然而, 当前实验的统计数据也可能会改变这个先验. 在动态贝叶斯模型中, 当前时间点的先验在经过贝叶斯证据积累后变成下一时间点的后验[35]. 因此, 未来的研究对于多感觉整合中先验的解读需要谨慎, 并且需要进一步研究多感觉整合中的先验在多大程度上与其他感知领域中的先验一样具有长

期的稳健性(例如慢速先验)[36].

贝叶斯因果推断模型将上述两种信息结合起来, 继而判断多模态信号存在共同原因的可能性并估计线索的位置. 以视听任务中空间位置判断为例, 可分为三种不同的情况. 首先, 在已知完全独立来源(C=2, 即有两个独立的源"Causes")的情况下, 听觉和视觉刺激是绝对分离的. 因此, 声音刺激的方位是可以被独立估算的: 其最佳估计值即为单感官位置的最大似然估计, 即为$\hat{s}_{A,c=2}=X_A$. 同理, 视觉刺激的方位被估算为视觉单感官位置的最大似然估计. 这也被称为完全分离模型(图1D). 其次, 在已知"共同来源"的情况下(C=1), 视听位置的最佳估计值是基于视觉和听觉感知的可靠性加权的平均值, 即为$\hat{s}_{A,c=1}=W_A\cdot X_A+W_v\cdot X_v$, 其中$W_A$和$W_v$觉和视觉所占的权重, 与其分布的方差成反比. 这也被称为强制整合模型(图1E). 以上为两种"极端"情形, 而第三种情况为大脑不知道其潜在的因果结构(C=1还是C=2?), 因此需要从感觉信号中推断出来. 在这种情况下, 大脑需要同时考虑上述两种因果结构发生的可能性, 并利用决策策略将两种因果结构中任务变量的估计值结合起来. 常见的决策策略包括模型平均(model averaging)、模型选择(model selection)和概率匹配(probability matching)等[37,38]. 根据"模型平均"策略, 大脑将强制整合的空间估计值与分离的单感官(即听觉或视觉)的空间估计值相结合, 并根据各自后验概率按比例加权运算得出变量的最终估计值. 以听觉估计为例(下同), $\hat{s}_A=p\left(c=1\mid X_A,X_v\right)\cdot\hat{s}_{A,c=1}+p\left(c=2\mid X_A,X_v\right)\cdot\hat{s}_{A,c=2}$, 其中$p\left(c=1\mid X_A,X_v\right)$为在当前刺激下共同来源的概率, $p\left(c=2\mid X_A,X_v\right)$为在当前刺激下共同来源的概率(图1F). 根据"模型选择"策略, 大脑会从更有可能的因果结构中选择性地得出空间位置的最终估计值, 即

$$\hat{s}_A=\begin{cases}\hat{s}_{A,c=1}\text{如果}p\left(c=1\mid X_A,X_v\right)>0.5,\\ \hat{s}_{A,c=2}\text{如果}p\left(c=1\mid X_A,X_v\right)\leqslant 0.5.\end{cases}$$

根据"概率匹配"策略, 大脑会随机报告选择的因果结构的空间位置估计值, 其比例与该因果结构的后验概率成正比, 即

$$\hat{s}_A=\begin{cases}\hat{s}_{A,c=1}\ \text{if}\ p\left(c=1\mid X_A,X_v\right)>\alpha,\alpha\sim U(0,1),\\ \hat{s}_{A,c=2}\ \text{if}\ p\left(c=1\mid X_A,X_v\right)\leqslant\alpha,\alpha\sim U(0,1).\end{cases}$$

"模型平均"作为一种符合标准贝叶斯概率过程的策略, 常被验证为解释被试多感觉因果推断行为的最佳模型[13,14]. 但有研究表明, 被试也会采用概率匹配策略, 即利用启发式捷思算法(heuristics)来处理多感官决策问题[37,38]. 上述详细的数学推断过程和公式可以参见Körding等人[12]和Cao等人[14]的文章.

值得注意的是, 以上模型推导结论是建立在默认成本函数(cost或loss function)为二次成本函数(quadratic)的前提下得出的. 在贝叶斯模型中, 最优决策的目标是最大化期望效用. 在连续估计任务中, 例如本文提到的视听位置判断任务(即腹语术任务), 被试需要精确判断信号源的物理位置. 在这种情况下, 二次成本函数是一种合理的选择, 定义为$\text{cost}(s,\hat{s})=(s-\hat{s})^2$, 其中$s$为客观物理值, $\hat{s}$为估计值, 其他可能的成本函数及其特征可以参考展望部分.

3.2 因果推断的影响因素

已有研究发现因果推断的影响因素众多. 然而, 其本质上取决于因果推断所依赖的两种关键信息, 似然概率和先验概率. 首先是似然概率, 即感知刺激的条件分布. 其中最关键的就是刺激的可靠性(reliability), 其被定义为刺激分布方差的倒数[8], 即方差越大(不确定性越高), 可靠性越低. 研究者发现, 大脑的感官权重分配(sensory weighting)主要依赖于刺激的可靠性. 特别是在共同信号来源的假设下(C=1), 感觉可靠性完全决定不同信号的权重分配[39]. 在实验中, 研究者通过调整符合高斯分布的点状云团的横轴方差来调节视觉信号的可靠性, 例如将其设定为2° (高可靠性)或14° (低可靠性)[13]. 除刺激本身的物理特性, 其他高阶认知变量也能间接调节大脑对刺激的加工. 一项近期研究揭示了注意作为重要的认知变量, 可以影响多感觉因果推断. 具体而言, 在刺激出现之前, 注意的分配能够改变早期(单通道)感觉表征的可靠性, 从而影响其在后期多感觉整合过程中的权重[40]. 例如, Odegaard等人[41]发现在空间任务中, 选择性注意会提高早期视觉表征的精确度, 而在时间任务中, 选择性注意会提高视听觉表征的精确度, 但都不会改变被试整合信号的先验概率. 同样, 实验任务也可以通过调节注意来影响多感觉加工中的权重分配. 研究发现如果当前任务让被试报告视觉(听觉), 视觉(听觉)信息被赋予更多的权重[40]. 更进一步地研究发现, 刺激出现前对视觉(听觉)的注意("前注意")增强视觉(听觉)信号的可靠性和视

觉(听觉)皮层的感觉表征; 而在刺激呈现后报告的感觉模态("后注意")则会影响更高级的顶叶皮层, 恰与因果推断有关[42]. 此外, 反复暴露于具有空间差异的听觉和视觉刺激后, 会导致听觉似然函数均值发生变化, 被试定位听觉刺激会向先前视觉方向偏移[43], 这被称为腹语术后效(ventriloquist aftereffect, VAE), 也可以被因果推断模型很好地解释[44].

另外, 先验概率也是影响因果推断的重要因素. 先验可分为不同层级, 它既包括涉及更长时间进程, 如知识、语义等因一方面素所塑造的先验, 也包括涉及短期的统计学习或适应所产生的改变[45]. 研究发现长期知识经验在塑造先验的过程中发挥着重要的作用. 例如, 在上述橡胶手错觉实验中, 如果虚假手臂被一个长方形木头(大小相同)所取代, 那么共同来源的先验概率就会下降, 从而影响因果推断[32]. 此外, 研究者发现通过短期实验操作也可以改变先验. 例如, 如果被试反复暴露于不匹配的视、听觉信号中, 他们对共同来源的先验预期会被改变, 并影响因果推断[46]. 类似现象也发生于其他感官通道, 比如视、触觉. 一段时间的视、触觉信号关联学习使得共同来源的先验预期升高, 从而增大强制整合的可能性[47]. 可见, 不同层级的先验对被试共同来源的先验概率产生不同的影响, 这一现象值得未来进一步探索.

4 多感觉整合中贝叶斯因果推断的神经机制

心理物理和计算建模的工作已经证明, 人类和动物在进行多感觉加工时遵循贝叶斯因果推断的原则[12,48~50]. 近年来, 研究者深入探究多感觉整合中贝叶斯因果推断的神经表征, 通过人类神经影像实验和猕猴电生理实验, 尝试揭示大脑在不同因果结构下如何编码和整合多感觉信号. 此外, 他们还关注哪些脑区参与贝叶斯因果推断, 以及其时间进程[13~16,32,51].

4.1 人类影像学证据

Rohe等人[13]最早揭示多感觉中因果推断的神经表征. 他们采用功能性核磁共振成像(functional magnetic resonance imaging, fMRI)技术, 记录人类被试在完成空间定位任务时的大脑活动. 具体而言, 基于上述贝叶斯因果推断模型预测四种空间位置的估计值: 独立听觉位置、独立视觉位置、强制整合的空间位置以及基于因果推断的空间位置. 研究者采用多变量分析, 根据fMRI信号对四种空间估计进行解码, 随后计算上述四个模型获得的空间估计与fMRI信号解码的空间估计之间的相关性. 此外, 他们运用超越概率(exceedance probability; 即该模型比其他模型与脑数据更吻合的概率, 用于模型比较)以此来确定特定大脑区域中主导的空间编码方式. 实验结果表明, 大脑表征多感官因果推断呈现层级加工的特征. 具体而言, 初级感觉皮层对独立来源(即完全分离)的单一感觉信息进行编码. 例如, 初级听觉皮层(primary auditory cortex, A1)相对单一地编码听觉空间位置, 而视觉区域(V1~V3)相对单一地编码视觉空间位置. 后顶内沟区域(intraparietal sulcus, IPS)基于视听信号为共同来源(即强制整合)的情况下对信号进行编码. 更高层级的脑区, 如顶内沟前部(anterior IPS), 则会编码那些遵循贝叶斯因果推断模型的感觉信号, 并充分表征因果结构的不确定性[13].

上述fMRI实验结果首次揭示多感觉加工中可能参与因果推断的脑区, 并证明大脑表征因果推断的层级加工特性. 为深入了解大脑进行因果推断的神经计算过程如何随时间动态展开, 研究者采用高时间分辨率的EEG技术和相似的实验范式对此问题进行探索. 与脑成像的结果类似, 因果推断在时间层面上也呈现出层级加工的特性. 具体而言, 多变量解码的结果显示, 在视听刺激发生后的早期阶段(约100 ms), 感觉信号表征是独立估计的, 主要受单感觉信号位置的影响. 随后, 从100到约200 ms, 脑电信号主要反映基于强制整合模型所估计的空间位置, 表明这个阶段受到视听信号的共同影响. 从200 ms开始, 大脑才会考虑外界信号因果结构的不确定性, 并在感觉整合与分离之间进行判断[15]. 重要的是, 以上发现在其他实验范式中也得到验证[14,18]. 例如, 在闪光融合任务中[24], 被试需要报告闪光次数或蜂鸣声次数. 结果同样发现在相对早期的阶段(120 ms), 大脑独立表征闪烁光环和蜂鸣声的数量. 之后, 在刺激呈现后450 ms, 大脑会依据贝叶斯因果推断对刺激进行表征[18].

最近, Cao等人[14]利用高时空分辨率的脑磁图技术(magnetoencephalography, MEG), 深入而全面地探究多感官因果推断在大脑中的时空动态表征. MEG具有与EEG相同的毫秒级时间分辨率, 但空间分辨率高

于EEG. 颅骨和脑周围的组织不会影响MEG测量的磁场, 但会强烈影响EEG测量的电位. 当来自大脑的电信号通过颅骨和头皮时, 它们会被扭曲并严重减弱. 而这些组织对大脑产生的磁场是相对无扰的. 因此, MEG能比EEG提供更准确的大脑活动空间估计, 并能更可靠地定位大脑功能. 研究者首先计算基于刺激出现引发的脑磁活动, 以此来探索大脑区域在何时采用何种计算模型来表征感官信息. 结果发现, 视听信息最初在各自的感觉皮层进行独立编码, 例如双侧距状裂皮层(bilateral calcarine cortex, 即初级视觉皮层, V1)从100 ms开始对视觉信息进行加工, 听觉皮层从140 ms开始对听觉信息进行加工. 随后, 在180至260 ms左右, 左侧颞上回、楔前叶和顶叶小叶、后扣带回腹侧和后颞上回等大脑区域主要依据可靠性加权整合模型进行感觉表征. 最后, 在刺激呈现后620 ms左右, 背外侧和腹外侧前额叶皮层(特别是左侧额叶下回)等更高级的脑区依据因果推断进行感觉表征. 此外, 他们还分析了基于被试逐次按键反应(选择)前的脑磁活动, 即反应锁定分析(response-locked analysis). 结果同样发现, 在被试反应前140到220 ms, 顶颞叶和右侧楔前叶依据强制整合模型进行感觉编码; 而在被试反应前80到100 ms, 双侧额上回依据贝叶斯因果推断模型进行感觉判断. 这些结果表明早期单感觉表征始于初级感觉皮层, 随后经过颞顶区进行强制整合, 最终到达额叶进行因果推断(图2A). 在这项研究中, 研究者发现前额叶皮层主要编码因果推断信息. 这不同于Rohe等人[13]研究结果, 即因果推断仅由顶内沟编码, 原因可能是他们并未检查前额叶区域, 或者是因为实验任务并未要求被试明确进行因果推断. 在另一研究中, 多变量fMRI模式分析显示侧前额叶皮层是唯一主要编码被试因果推断信息(即, 共同原因与独立原因)的脑区[52]. 在Cao等人[14]的研究中, 作者使用MEG定位波束成形技术, 以高空间分辨率定位神经活动, 成功地在顶内沟(IPS)发现从单感官分离到强制融合(posterior IPS), 再到灵活因果推断(anterior IPS)转变的层级结构(图2B). 这一结果有助于协调他们的发现与先前研究之间的差异. 此外, Cao等人[14]扩展Rohe等人[13]的研究结果, 将涉及的皮层网络扩展到顶内沟区域之外. 以上神经影像学的证据共同表明, 涉及单感官编码的神经计算发生在相对较早的阶段并依赖于较低级别的大脑区域, 而涉及贝叶斯因果推断则在相对较晚的阶段出现, 并在较高级别的大脑区域中发挥作用.

4.2 猕猴电生理证据

因果推断的电生理证据最早来自Fang等人[32]在猕猴中进行的橡皮手错觉实验. 研究人员训练猕猴移动手臂去触摸一个视觉目标(“reaching”任务), 并探究视觉假臂(可见)的位置如何影响猕猴真实手臂(不可见)偏向视觉目标的运动轨迹(即错觉效应). 他们发现, 当本体感觉(proprioceptive)信号(真实手臂)和视觉信号(虚假手臂)一致时, 猕猴倾向于整合信息; 当差异逐渐增大时, 猕猴则倾向于将信号分离. 这一行为数据可以被贝叶斯因果推断模型所预测并量化. 更重要的是, 通过微电极技术记录猕猴大脑的放电活动, 研究者发现前运动皮层(premotor cortex)中神经元群体的活动与观察到的错觉效应相一致. 这表明, 对因果推断的编码主要与前运动皮层神经元的活动有关[32]. 随后, 他们采用相似的实验范式进一步揭示额顶环路中前运动皮层和顶叶皮层(parietal area 5 and area 7)在因果推断中如何动态更新先验知识和感觉表征. 他们发现额顶环路通过层级加工的方式实现贝叶斯因果推断整合, 其中前运动皮层可以对共同来源进行自上而下的反馈加工, 并且该区域监控顶叶皮层进行感觉整合的权重分配. 前运动皮层和顶叶皮层在整合感觉信息的同时, 可以动态更新潜在的因果结构, 这样大脑能够更精确地估计信号来源[53]. 此外, 在现实生活中, 刺激的可靠性也是动态变化的, 那么大脑如何在多感觉决策任务中动态编码这些信息呢? Hou等人[33]通过改变刺激的动态特性(如加速度或速度)模拟自然环境中信息可靠性实时变化的过程. 他们训练猕猴依据前庭(vestibular)与视觉两种模态的感觉信息来判断其自身的运动方向. 结果发现在其顶内沟外侧壁(lateral intraparietal area, LIP)的神经元累积不同模态信息的证据. 研究者进一步通过神经网络的方法证实这些神经元遵循“线性不变概率性群体编码”(invariant linear probabilistic population code, ilPPC)的原则, 即神经元群体的实时放电活动可以直接表征信息输入的可靠性. 神经元群体只需要对感觉输入进行突触权重不变的简单线性叠加, 就可以实现信息的贝叶斯最优整合[33]. 这些结果表明, 猕猴的多感官决策行为符合贝叶斯因果推断原理, 并且随着越来越多的电生理证据和神经元群体编码算法的出现, 相应的神经机制也

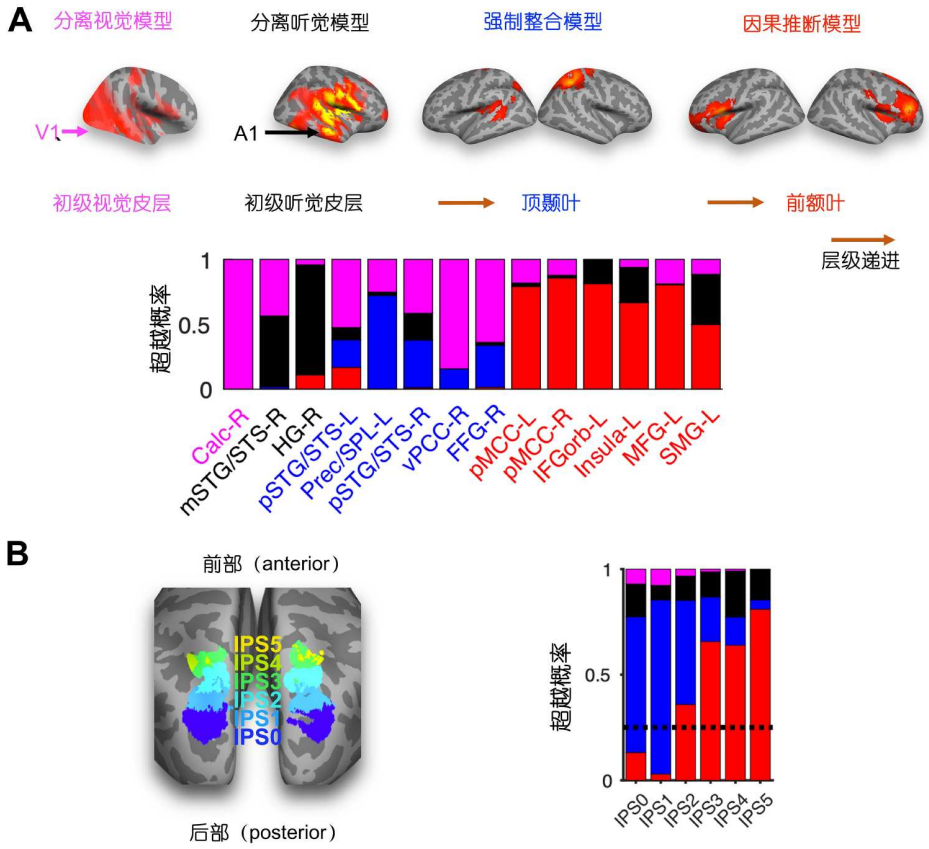

图 2 大脑皮层中的贝叶斯因果推断的层级结构. A图示意从单感官分离到强制融合再到因果推断的层级加工过程; B图示意在顶内沟(IPS)内自后向前的因果推断的层级表征[14]. Calc: calcarine (即V1)初级视觉皮层/纹状区; mSTG/STS: middle superior temporal gyrus/sulcus 中部颞上回/沟; HG(即A1): 赫施尔回; pSTG/STS 后颞上回/沟; Prec: 前楔叶; vPCC: 腹侧后扣带回; FFG: 侧颞回; pMCC: 中后扣带回及沟; IFGorb: 额下回(眶部); Insular: 岛叶; MFG: 额中回; SMG: 顶缘回; IPS: 顶内沟; L/R, 左/右半球. 超越概率是用来量化一种可能性的指标, 即某一特定脑区编码一个给定模型所预测的表征相似性模式(representational similarity pattern)相较于编码其他备选模型所预测的表征相似性模式更符合数据情况的可能性

Figure 2 Hierarchical structure of Bayesian causal inference in the cerebral cortex. Panel A illustrates the hierarchical processing from unisensory separation to forced fusion and then to causal inference; panel B shows the hierarchical representation of causal inference from posterior to anterior in the intraparietal sulcus (IPS)[14]. Calc: calcarine cortex (V1), primary visual cortex; mSTG/STS: middle superior temporal gyrus/sulcus; HG: Heschl's gyrus (A1); pSTG/STS: posterior superior temporal gyrus/sulcus; Prec: precuneus; vPCC: posterior-ventral cingulate gyrus; FFG: fusiform gyrus; pMCC: middle-posterior cingulate gyrus and sulcus; IFGorb: inferior frontal gyrus (pars orbitalis); MFG: middle frontal gyrus; SMG: supramarginal gyrus; IPS: intraparietal sulcus; L/R: left/right hemisphere. Exceedance probability quantifies the likelihood that a given model's representational similarity pattern, as encoded by a specific brain region, aligns more closely with the data than the patterns predicted by alternative models

逐渐显现.

另外, 在背内侧颞上区(dorsal medial superior temporal area, MSTd)和顶内沟腹侧区(ventral intraparietal area, VIP), 存在一类特殊的神经元, 它们对同一方向的前庭和视觉信息具有调谐特性(tuning properties), 即这些神经元存在特定的方向"偏好", 且在前庭与视觉信号指向相同的特定方向时激活最强, 因此被称为同向神经元("congruent" neurons); 而另一类神经元则对相反方向的前庭和视觉信息具有调谐特性, 被称为反向神经元("opposite" neurons)[17,54]. 换句话说, 同向神经元对视觉和前庭线索的调谐曲线相似, 而反向神经元对视觉线索的调谐曲线比对前庭线索的调谐曲线偏移180°. 研究发现同向神经元基于可靠性线索加权的原则, 将两种感觉信息进行整合[17,39,55]; 而反向神经元的作用目前尚不清楚. 但研究发现当视觉线索和前庭线索之间的差异变大时, 反向神经元的反应也会增强, 因此有观点认为它们作用于因果推断[55]. 最近, 研究者通过拟合基于生物学原理的分布式网络模型(decentralized network model), 不仅成功地再现反向神经元的调谐特性, 还证明同向和反向神经元发挥着互补功

能. 具体来说, 同向神经元负责线索整合, 而反向神经元负责计算分离的线索差异. 神经系统可以据此评估整合的有效性, 而两组神经元的相互作用可以实现高效的多感官信息处理[56]. 此外, Rideaux等人[16]利用人工神经网络进一步揭示同向和反向神经元如何协同完成多感觉整合中的因果推断. 在他们的实验中, 猕猴被训练利用外界视觉线索和本体前庭感觉进行位置判断, 从而区分当前是自我运动还是场景运动. 然而, 这种判断对猕猴进行因果推断也是一个挑战, 因为视觉运动可能来自自身运动或场景运动, 而前庭运动则完全由自身运动引起[39]. 研究者通过训练一个无约束(unconstrained)的前馈神经网络来执行因果推断以及对场景位置的估计. 该网络最初接受视觉和前庭输入. 视觉输入是在平移(x, y)、径向(z)和旋转(r)(即绕z轴旋转)方向上以不同速度移动的自然图像序列, 由此产生的活动传递到卷积-池化层(convolutional-pooling layer; 对应猕猴middle temporal “MT层”). 前庭输入由四个噪声高斯分布组成, 代表耳石器官和半规管在前庭核内产生的运动信号. 该输入由一个全连接层(fully-connected layer; 对应顶叶内侧前庭皮层, parieto-insular vestibular cortex, “PIVC层”)提取、读出. 随后, 视觉信息和前庭信息进一步共同汇聚到一个通用的全连接层(对应上颞叶背内侧子区, dorsal subdivision of the medial superior temporal area, “MSTd层”). 该连接层的神经活动经过训练可以解码出自我和场景运动的估计值, 并执行因果推断. 该神经网络模型不仅很好地拟合在人类和猕猴身上观察到的方向判断任务的行为数据, 还展现出近似于在猕猴MSTd脑区真实记录到的同向和反向神经元. 同向神经元单元主要影响对自身运动的估计, 而反向神经元单元则主要影响对场景运动的估计. 此外, 同向和反向神经元之间的动态平衡可以用来解决因果推断决策. 也就是说, 当同向神经元单元更活跃时, 信号被归因于同一事件, 而当反向神经元单元更活跃时, 信号被归因于不同事件[16].

综上所述, 随着神经成像和电生理技术的发展, 已有大量研究探究人类和猕猴进行因果推断时所表现出层级加工特性的神经表征. 然而, 探究多感觉因果推断的神经生物学算法与机理仍处于起步阶段, 许多问题仍有待进一步研究. 例如, 虽然目前研究揭示因果推断所涉及的脑区, 但是脑区之间的功能连接还不清楚, 未来可采用动态因果模型探索此问题; 此外, 虽然目前的研究表明额顶叶负责因果推断, 但依然缺乏因果性的证据. 建立和验证机制模型需要证明神经活动与行为之间的因果联系. 精细的神经通路操控工具, 例如光遗传, 使研究者能够以更高的特异性控制神经通路和特定细胞类型的神经活动; 在人类实验中经颅磁刺激(TMS)的使用也可以更明确地揭示参与多感觉因果推断的神经网络, 已建立完整、具体的神经机制, 并有效弥补贝叶斯概率模型在解释算法和生理层面的神经过程方面的不足. 此外, 目前研究者对皮层(以及可能的皮层下)回路如何实现贝叶斯推断的动态过程理解仍然有限[57,58]. 虽然Rideaux等人[16]的模型从算法层面为因果推断提供生物学上更加可行的机制, 但是这些新模型思想可能局限于特定的感知模态, 例如在视听整合和听嗅整合中是否成立还不得而知. 因为这些模型所基于的神经元仅在猕猴MSTd和VIP脑区发现, 这类神经元或具有类似特性的神经元是否必然存在于表征其他感觉通道(如听觉、触觉)的大脑皮层需要进一步验证. 然而, 这些模型侧面说明实现因果推断的“多重实现”(multiple realization)问题需要得到重视: 实现因果推断不需要复杂的概率群体编码和贝叶斯推断, 其他机制也能实现. 这是否意味着多感觉因果推断是多种机制共同作用的结果还尚不可知, 未来研究需要利用神经数据和证据来约束理论, 并更好地规划模型.

5 总结与展望

生物体存在于复杂多变的环境中, 正确进行因果推断以整合感觉信息并优化决策对其生存和发展至关重要. 对这一问题的探讨不仅有助于理解生物体如何进行多感觉决策, 还能为开发人工智能系统提供新的启示. 在过去的10至15年中, 多感觉整合的因果推断机制研究取得了显著进展, 涵盖从行为研究到计算模型及神经机制等多个层面. 尤其是在对人类和猕猴的研究中, 研究者已初步揭示因果推断的神经机制, 为未来相关研究奠定基础. 尽管取得一些进展, 但当前因果推断的研究仍处于初级阶段, 相关研究领域仍然较为分散, 许多问题亟待深入研究.

前文在介绍多感觉整合的因果推断模型时, 重点探讨人类视-听觉整合(空间感知任务)和动物视-前庭觉整合(身体感知和朝向判断任务), 但因果推断思想

和核心计算方法同样适用于其他感觉通道和任务[31], 例如大小重量错觉[59]、自身运动方向感知[60], 以及垂直度感知等[61]. 此外, 研究还发现人类的主动控制感(sense of agency), 涉及自我行动对产生结果的因果关系判断, 也可以通过贝叶斯因果推断模型来解释[62]. 可见, 因果推断作为一种典型的计算过程, 广泛指导着各种适应行为[31]. 总的来说, 涉及多种线索的整合加工, 其本质都遵循因果推断的基本原理. 然而, 目前尚不清楚不同任务和不同模态所涉及的因果推断的神经机制是否一致, 未来研究可以从整合的视角探讨这些问题. 例如, 研究者在考察人类和猕猴的因果推断机制时采用不同的实验范式, 这为比较人类和动物的研究结果带来困难. 未来可以采用相同的实验范式及跨物种比较的方法来探讨这一问题, 以进一步揭示因果推断能力在进化过程中的演化轨迹[63,64].

再者, 目前关于多感觉整合的因果推断模型多集中于探究正常成年群体, 少数研究对比幼儿(6~8岁)、大龄儿童(9.5~12.5岁)和成人在触觉任务中的表现, 发现从幼儿时期开始, 个体就能够推断多感觉信号的因果关系[65]. 此外, 老年人在进行视听整合因果推断任务中表现出与年轻人相似的能力, 但是模型拟合的结果分析发现, 老年被试采取的决策阈值更高, 因此需要牺牲反应速度来完成判断[66]. 这些研究初步探究因果推断在人类身上的发展轨迹, 但由于因果推断任务难度较大, 目前探讨因果推断在婴儿阶段的研究相对较少. 未来的研究可以考虑采用更简单的实验任务和记录指标, 来探究婴儿因果推断的机制. 探讨这个问题还有助于回答人类因果推断的能力是天生遗传的还是后天习得的[67]. 如果该能力是后天习得的, 那么关于其何时出现、学习速度、具体机制等问题都值得进一步研究. 此外, 计算精神病学家也把目光关注到特殊群体的因果推断能力, 例如孤独症患者如何完成因果推断. 研究者发现, 虽然孤独症群体在整合多感觉刺激时表现出与正常人类相似的位置判断能力, 但当视听信号差异增大时, 他们仍倾向于整合线索, 这与因果推断模型的预测相反, 表明他们在判断感觉信息是应当整合还是分离时存在能力障碍. 矛盾的是, 在需要明确报告信号是否来源于共同原因的任务中, 他们报告共同原因的可能性更小. 这说明孤独症群体可能存在一种补偿机制, 使他们倾向于在外显报告中弥补其整合偏差[68]. 虽然该研究揭示人类与特殊群体在进行因果推断时表现不同, 但目前关于特殊人群进行因果推断的研究仍相对较少, 相关机制的讨论仍然不足, 仍需进一步探究.

贝叶斯因果推断模型不仅在许多研究中与行为数据高度吻合, 表现优于其他模型[12,50], 而且也为实验结果提供更全面的解释[69]. 具体而言, 因果推断模型可以定量检验似然函数和先验在多感觉整合中所发挥的作用[70]. 例如, 研究者发现守门员相比于外场球员表现出更低的多感觉错觉效应, 原作者将这种多感觉错觉程度的降低完全归因于被试视听整合先验倾向的降低, Zhu等人[69]通过拟合因果推断模型, 发现它们所观察到的结果可能缘于守门员具有更好的视觉精度(似然函数发生改变). 但同时, 因果推断模型本身的发展也面临一些挑战和局限. 呼应前面章节对成本函数的介绍, 虽然二次成本函数是一种合理的选择, 但这并不意味着其他成本函数在某些多感觉整合任务中不适用. 每个效用函数在最大化期望效用时都会导致不同的决策规则. 其他成本函数, 例如定义为$\text{cost}(s,\hat{s}) = |s-\hat{s}|$的绝对误差成本函数, 也是合理的. 在最大化期望效用时, 绝对误差成本函数促使决策者报告后验分布的中位数而不是均值. 此外, 在离散的选项任务中, 0/1成本函数可能更为合理. 但已有研究表明, 人类在感觉运动学习中真正使用的成本函数, 其成本在小误差情况下近似二次成本函数, 而在大误差情况下则显著低于二次成本[71]. 目前, 研究者普遍持有生物大脑在感知过程中存在优化成本函数的观点. 大脑中特定区域的神经元可以改变其属性, 例如突触属性, 以便更好地完成成本函数定义的任务. 不同脑区的成本函数是多种多样的, 并且可能随着生物体发展而变化. 一些在生物学上可行的算法已经被提出, 包括广义再循环[72]、对比赫布学习[73]、随机反馈权重结合突触稳态[74]. 尽管这些机制在细节上有所不同, 但它们都涉及传递误差信号的反馈连接. 对感知任务的学习通过对成本函数的优化而实现, 即将预测与目标对比进而利用误差信号优化成本函数[75]. 综上, 在大脑中优化成本函数是人工智能和神经科学交叉领域的一个具有挑战性的话题, 值得未来进一步的研究.

此外, 贝叶斯模型在数值估计时直接依赖概率计算, 缺乏对受生物学启发的神经过程的考虑. 换言之, 贝叶斯模型虽然能够很好地解释复杂的多感觉决策行为, 但并非详细的具有生物学机理的“过程模型”(pro-

cess model). 近年来, 神经网络的迅速发展也为揭示多感觉因果推断机制提供新的思路. 首先, 有些研究者尝试使用基于生物学启发的神经网络来模拟多感觉整合的过程. 他们受启发于神经群体编码理论[76,77], 构建生物学上可行的神经网络. 其中包含两条通过跨模态突触连接的单感觉神经元链(一条是听觉神经元, 一条是视觉神经元), 他们发现多感觉错觉效应在很大程度上取决于跨模态突触权重[78]. 进一步, 研究者在此模型基础上, 加入因果推断机制. 新的模型由三层拓扑结构组成: 最初两层分别编码听觉和视觉刺激, 它们之间通过兴奋性突触相互连接, 并向下游第三层发送兴奋性连接. 该模型通过元素之间的侧向突触连接来实现因果推断, 即对外界相近位置敏感的元素会共同兴奋, 而对不同位置敏感的元素会相互抑制, 这种突触结构可以使神经网络将这两个活动分开, 并确定它们是否由不同事件产生. 在近距离两个刺激的情况下, 通常会认为存在一个共同来源, 两个刺激会相互吸引, 从而产生典型的多感觉错觉现象(如腹语术效应). 在远距离两个刺激的情况下, 通常会认为这些刺激来自不同的输入源, 这时区域内抑制作用则成为主导机制[79]. 此外, 如前所述, 研究者也在尝试使用人工神经网络探究多感觉因果推断机制[16]. 最近的一项研究显示, 通过简单的基于错误的学习规则训练的通用神经网络, 在九种常见心理物理任务中表现出接近最优的概率推断能力, 包括因果推断任务[80]. 目前, 神经网络推动该领域的进一步发展, 但仍有以下问题等待回答. 考虑到高级决策往往会自上而下影响感觉表征, 未来研究也值得探讨反馈通路在因果关系决策中所发挥的作用; 由此引发的问题是, 循环神经网络是否能更好地捕捉因果推断. 此外, 这些网络模型能否迁移到其他任务以及更复杂的自然环境, 也需要更深入的研究[81]. 最后, 随着人工智能的发展, 智能机器人利用神经网络来处理多感觉信息并进行因果推断具有广阔的应用前景[82,83].

参考文献

1 Stein B E. The New Handbook Of Multisensory Processing. Cambridge: Mit Press. 2012
2 De Gelder B, Bertelson P. Multisensory integration, perception and ecological validity. Trends Cogn Sci, 2003, 7: 460–467
3 Liu R, Wang L, Jiang Y. Recent progress in the study of consciousness and multisensory integration (in Chinese). Chin Sci Bull, 2016, 61: 2–11 [刘睿, 王莉, 蒋毅. 意识与多感觉信息整合的最新研究进展. 科学通报, 2016, 61: 2–11]
4 Wen X H, Liu Q, Sun H J, et al. Theoretical models of multisensory cues integration (in Chinese). Adv Psychol Sci, 2009, 4: 659–666 [文小辉, 刘强, 孙弘进, 等. 多感官线索整合的理论模型. 心理科学进展, 2009, 4: 659–666]
5 Van der Burg E, Olivers C N L, Bronkhorst A W, et al. Pip and pop: nonspatial auditory signals improve spatial visual search. J Exp Psychol-Hum Percept Perform, 2008, 34: 1053–1065
6 Noesselt T, Bergmann D, Hake M, et al. Sound increases the saliency of visual events. Brain Res, 2008, 1220: 157–163
7 Ernst M O, Bülthoff H H. Merging the senses into a robust percept. Trends Cogn Sci, 2004, 8: 162–169
8 Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. Curr Biol, 2004, 14: 257–262
9 Seilheimer R L, Rosenberg A, Angelaki D E. Models and processes of multisensory cue combination. Curr Opin Neurobiol, 2014, 25: 38–46
10 Roach N W, Heron J, McGraw P V. Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. Proc R Soc B, 2006, 273: 2159–2168
11 Shams L, Beierholm U R. Causal inference in perception. Trends Cogn Sci, 2010, 14: 425–432
12 Körding K P, Beierholm U, Ma W J, et al. Causal inference in multisensory perception. PLoS one, 2007, 2: e943
13 Rohe T, Noppeney U, Kayser C. Cortical hierarchies perform bayesian causal inference in multisensory perception. PLoS Biol, 2015, 13: e1002073
14 Cao Y, Summerfield C, Park H, et al. Causal inference in the multisensory brain. Neuron, 2019, 102: 1076–1087.e8
15 Aller M, Noppeney U, Petkov C. To integrate or not to integrate: temporal dynamics of hierarchical Bayesian causal inference. PLoS Biol, 2019, 17: e3000210
16 Rideaux R, Storrs K R, Maiello G, et al. How multisensory neurons solve causal inference. Proc Natl Acad Sci USA, 2021, 118: e2106235118
17 Gu Y, Angelaki D E, DeAngelis G C. Neural correlates of multisensory cue integration in macaque MSTd. Nat Neurosci, 2008, 11: 1201–1210

18 Rohe T, Ehlis A C, Noppeney U. The neural dynamics of hierarchical Bayesian causal inference in multisensory perception. Nat Commun, 2019, 10: 1907
19 Magnotti J F, Beauchamp M S, Gershman S J. A causal inference model explains perception of the McGurk effect and other incongruent audiovisual speech. PLoS Comput Biol, 2017, 13: e1005229
20 Lewald J, Guski R. Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. Cogn Brain Res, 2003, 16: 468–478
21 Spence C. Just how important is spatial coincidence to multisensory integration? Evaluating the spatial rule. Ann New York Acad Sci, 2013, 1296: 31–49
22 Stevenson R A, Wallace M T. Multisensory temporal integration: task and stimulus dependencies. Exp Brain Res, 2013, 227: 249–261
23 Stevenson R A, Zemtsov R K, Wallace M T. Individual differences in the multisensory temporal binding window predict susceptibility to audiovisual illusions. J Exp Psychol Hum Percept Perform, 2012, 38: 1517–1529
24 Shams L, Kamitani Y, Shimojo S. What you see is what you hear. Nature, 2000, 408: 788
25 Hirst R J, McGovern D P, Setti A, et al. What you see is what you hear: twenty years of research using the Sound-Induced Flash Illusion. Neurosci Biobehav Rev, 2020, 118: 759–774
26 Adam R, Noppeney U. Prior auditory information shapes visual category-selectivity in ventral occipito-temporal cortex. Neuroimage, 2010, 52: 1592–1602
27 Lee H L, Noppeney U. Physical and perceptual factors shape the neural mechanisms that integrate audiovisual signals in speech comprehension. J Neurosci, 2011, 31: 11338–11350
28 Boston D. Hearing lips and seeing voices. Br J Audiol, 1977, 11: 86–87
29 Rohe T, Noppeney U. Sensory reliability shapes perceptual inference via two mechanisms. J Vision, 2015, 15: 22
30 Wallace M T, Roberson G E, Hairston W D, et al. Unifying multisensory signals across time and space. Exp Brain Res, 2004, 158: 252
31 Shams L, Beierholm U. Bayesian causal inference: a unifying neuroscience theory. Neurosci Biobehav Rev, 2022, 137: 104619
32 Fang W, Li J, Qi G, et al. Statistical inference of body representation in the macaque brain. Proc Natl Acad Sci USA, 2019, 116: 20151–20157
33 Hou H, Zheng Q, Zhao Y, et al. Neural correlates of optimal multisensory decision making under time-varying reliabilities with an invariant linear probabilistic population code. Neuron, 2019, 104: 1010–1021.e10
34 Ma W J. Bayesian decision models: a primer. Neuron, 2019, 104: 164–175
35 Yu A J, Dayan P, Cohen J D. Dynamics of attentional selection under conflict: toward a rational Bayesian account. J Exp Psychol Hum Percept Perform, 2009, 35: 700–717
36 Stocker A A, Simoncelli E P. Noise characteristics and prior expectations in human visual speed perception. Nat Neurosci, 2006, 9: 578–585
37 Meijer D, Noppeney U. Computational models of multisensory integration. In: Sathian K. ed. Multisensory Perception. Amsterdam: Elsevier, 2020. 113–133
38 Wozny D R, Beierholm U R, Shams L, et al. Probability matching as a computational strategy used in perception. PLoS Comput Biol, 2010, 6: e1000871
39 Fetsch C R, Pouget A, DeAngelis G C, et al. Neural correlates of reliability-based cue weighting during multisensory integration. Nat Neurosci, 2012, 15: 146–154
40 Rohe T, Noppeney U. Reliability-weighted integration of audiovisual signals can be modulated by top-down attention. Eneuro, 2018, 5: ENEURO.0315-17.2018
41 Odegaard B, Wozny D R, Shams L. The effects of selective and divided attention on sensory precision and integration. Neurosci Lett, 2016, 614: 24–28
42 Ferrari A, Noppeney U, Summerfield C. Attention controls multisensory perception via two distinct mechanisms at different levels of the cortical hierarchy. PLoS Biol, 2021, 19: e3001465
43 Wozny D R, Shams L. Computational characterization of visually induced auditory spatial adaptation. Front Integr Neurosci, 2011, 5
44 Hong F, Badde S, Landy M S, et al. Causal inference regulates audiovisual spatial recalibration via its influence on audiovisual perception. PLoS Comput Biol, 2021, 17: e1008877
45 de Lange F P, Heilbron M, Kok P. How do expectations shape perception? Trends Cogn Sci, 2018, 22: 764–779
46 Odegaard B, Wozny D R, Shams L. A simple and efficient method to enhance audiovisual binding tendencies. PeerJ, 2017, 5: e3143

47 Ernst M O. Learning to integrate arbitrary signals from vision and touch. J Vision, 2007, 7: 7

48 Kayser C, Shams L. Multisensory causal inference in the brain. PLoS Biol, 2015, 13: e1002075

49 Mohl J T, Pearson J M, Groh J M. Monkeys and humans implement causal inference to simultaneously localize auditory and visual stimuli. J Neurophysiol, 2020, 124: 715–727

50 Acerbi L, Dokka K, Angelaki D E, et al. Bayesian comparison of explicit and implicit causal inference strategies in multisensory heading perception. PLoS Comput Biol, 2018, 14: e1006110

51 Rohe T, Noppeney U. Distinct computational principles govern multisensory integration in primary sensory and association cortices. Curr Biol, 2016, 26: 509–514

52 Mihalik A, Noppeney U. Causal inference in audiovisual perception. J Neurosci, 2020, 40: 6600–6612

53 Qi G, Fang W, Li S, et al. Neural dynamics of causal inference in the macaque frontoparietal circuit. Elife, 2022, 11, e76145

54 Chen A, DeAngelis G C, Angelaki D E. Functional specializations of the ventral intraparietal area for multisensory heading discrimination. J Neurosci, 2013, 33: 3567–3581

55 Morgan M L, DeAngelis G C, Angelaki D E. Multisensory integration in macaque visual cortex depends on cue reliability. Neuron, 2008, 59: 662–673

56 Zhang W H, Wang H, Chen A, et al. Complementary congruent and opposite neurons achieve concurrent multisensory integration and segregation. eLife, 2019, 8: e43753

57 Fiser J, Berkes P, Orbán G, et al. Statistically optimal perception and learning: from behavior to neural representations. Trends Cogn Sci, 2010, 14: 119–130

58 Pouget A, Beck J M, Ma W J, et al. Probabilistic brains: knowns and unknowns. Nat Neurosci, 2013, 16: 1170–1178

59 Buckingham G, Michelakakis E E, Cole J. Perceiving and acting upon weight illusions in the absence of somatosensory information. J Neurophysiol, 2016, 115: 1946–1953

60 Dokka K, Park H, Jansen M, et al. Causal inference accounts for heading perception in the presence of object motion. Proc Natl Acad Sci USA, 2019, 116: 9060–9065

61 de Winkel K N, Katliar M, Diers D, et al. Causal inference in the perception of verticality. Sci Rep, 2018, 8: 5483

62 Legaspi R, Toyoizumi T. A Bayesian psychophysics model of sense of agency. Nat Commun, 2019, 10: 4250

63 Friedrich P, Forkel S J, Amiez C, et al. Imaging evolution of the primate brain: the next frontier? Neuroimage, 2021, 228: 117685

64 Tsao D Y, Moeller S, Freiwald W A. Comparing face patch systems in macaques and humans. Proc Natl Acad Sci USA, 2008, 105: 19514–19519

65 Verhaar E, Medendorp W P, Hunnius S, et al. Bayesian causal inference in visuotactile integration in children and adults. Dev Sci, 2022, 25: e13184

66 Jones S A, Beierholm U, Meijer D, et al. Older adults sacrifice response speed to preserve multisensory integration performance. Neurobiol Aging, 2019, 84: 148–157

67 Wang Y, Wang L, Xu Q, et al. Heritable aspects of biological motion perception and its covariation with autistic traits. Proc Natl Acad Sci USA, 2018, 115: 1937–1942

68 Noel J P, Shivkumar S, Dokka K, et al. Aberrant causal inference and presence of a compensatory mechanism in autism spectrum disorder. eLife, 2022, 11: e71866

69 Zhu H, Beierholm U, Shams L. The overlooked role of unisensory precision in multisensory research. Curr Biol, 2024, 34: R229–R231

70 Quinn M, Hirst R J, McGovern D P. Distinct profiles of multisensory processing between professional goalkeepers and outfield football players. Curr Biol, 2023, 33: R994–R995

71 Körding K P, Wolpert D M. The loss function of sensorimotor learning. Proc Natl Acad Sci USA, 2004, 101: 9839–9842

72 O'Reilly R C. Biologically plausible error-driven learning using local activation differences: the generalized recirculation algorithm. Neural Comput, 1996, 8: 895–938

73 Xie X, Seung H S. Equivalence of backpropagation and contrastive hebbian learning in a layered network. Neural Comput, 2003, 15: 441–454

74 Lillicrap T P, Cownden D, Tweed D B, et al. Random synaptic feedback weights support error backpropagation for deep learning. Nat Commun, 2016, 7: 13276

75 Marblestone A H, Wayne G, Kording K P. Toward an integration of deep learning and neuroscience. Front Comput Neurosci, 2016, 10: 94

76 Ma W J, Beck J M, Latham P E, et al. Bayesian inference with probabilistic population codes. Nat Neurosci, 2006, 9: 1432–1438

77 Pouget A, Dayan P, Zemel R S. Inference and computation with population codes. Annu Rev Neurosci, 2003, 26: 381–410
78 Cuppini C, Magosso E, Bolognini N, et al. A neurocomputational analysis of the sound-induced flash illusion. Neuroimage, 2014, 92: 248–266
79 Cuppini C, Shams L, Magosso E, et al. A biologically inspired neurocomputational model for audiovisual integration and causal inference. Eur J Neurosci, 2017, 46: 2481–2498
80 Orhan A E, Ma W J. Efficient probabilistic inference in generic neural networks trained with non-probabilistic feedback. Nat Commun, 2017, 8: 138
81 Noppeney U. Solving the causal inference problem. Trends Cogn Sci, 2021, 25: 1013–1014
82 Zeng T, Tang F, Ji D, et al. NeuroBayesSLAM: neurobiologically inspired Bayesian integration of multisensory information for robot navigation. Neural Netws, 2020, 126: 21–35
83 Tan H, Zhou Y, Tao Q, et al. Bioinspired multisensory neural network with crossmodal integration and recognition. Nat Commun, 2021, 12: 1120

Computational models and neural mechanisms of causal inference in multisensory integration

CHENG YuHui[1*], YUAN XiangYong[2,3], JIANG Yi[2,3] & CAO YiNan[4,5*]

[1] *School of Psychology, Nanjing Normal University, Nanjing 210024, China*
[2] *State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China*
[3] *Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China*
[4] *Department of Neurophysiology and Pathophysiology, University Medical Center Hamburg-Eppendorf, Hamburg 20246, Germany*
[5] *Department of Cognitive Studies, Ecole Normale Superieure-PSL, Paris 75005, France*
* *Corresponding authors, E-mail: chengyh@nnu.edu.cn; yinan.cao@ens.psl.eu*

Humans and other animals continuously receive multisensory information from complex environments. To efficiently perceive, make decisions, and respond, the brain must integrate sensory signals from common sources while segregating those from independent sources. This process requires organisms to solve the "causal inference" problem in multisensory information processing. This review traces the evolution of multisensory models, from the early forced integration model to the recent Bayesian causal inference model, and summarizes commonly used experimental paradigms and foundational computational principles in multisensory causal inference. At the neural level, the brain encodes causal inference information dynamically and hierarchically. Specifically, this process begins with the rapid representation of unimodal information in the primary sensory cortex, progresses to forced integration in parieto-temporal areas, and ultimately culminates in causal inference within frontoparietal regions. Future research should integrate multiple experimental paradigms to further explore the neural mechanisms underpinning the causal inference model, especially in individuals with cognitive disorders. Additionally, it should address the limitations of Bayesian causal inference models and examine the potential of network models to reveal the neural basis of causal inference.

multisensory integration, Bayesian causal inference, hierarchical processing, computational model, neural network

doi: 10.1360/SSV-2024-0160