多感官整合的规范模型与神经网络模型

Marcus Ghosh[1, *] **和 John S. Butler**[2]

[1]伦敦帝国理工学院生命科学系，伦敦，英国[2]都柏林理工大学数学与统计学院，格兰奇戈曼下区，都柏林，爱尔兰[*]电子邮件：m.ghosh@imperial.ac.uk

arXiv:2611.32829v1 [q-bio.NC] 24 Nov 2026

摘要

动物持续感知气味、视觉信息和声音，并由大脑将其整合成多种多感官信息流。多感官整合能够加快学习、提高反应速度，并有助于消除感觉歧义。为研究这一过程，该领域发展出一系列模型：从仅包含少数变量的简洁数学方程，到拥有数十亿个参数的人工神经网络。在这篇综述中，我们将这些模型划分为两类：规范性模型以关于多感官信号应如何整合的假设为出发点，而神经网络模型则旨在模拟神经回路可能如何整合这些信号。这些模型共同表明，观察者能够以统计最优方式整合线索，并有助于解释实验发现，包括多感官神经元响应特征和网络架构的多样性。在这篇综述中，我们将以通俗易懂的方程和 Python 代码，对多感官模型进行教学式概述，并比较它们的假设和解释力。我们详细介绍从这些模型中获得的认识。最后，我们将讨论如何协调经典理论模型与人工神经网络所提出的观点和见解，以及这种融合应如何塑造未来研究。

引言

一切都是多感官的。在环境层面，自然栖息地中充满了气味、视觉信息和声音；即使在受控实验室环境中，执行“单感官”任务的动物仍会感知到其他感觉信号（或感知到这些信号不自然地缺失）。在神经层面，感觉表征似乎不像经典观点所认为的那样彼此分隔。初级皮层区域中可以观察到多感官过程[1–3]。而且，初级感觉皮层之间存在解剖学连接[4,5]，高阶多感官区域也会通过反馈连接返回“单感官”区域[2]。这些功能和解剖学结果支持这样一种观点：多感官整合是一种普遍且基础的神经计算[6]。

然而，多感官整合并不是某一种单一的计算过程。相反，感觉如何整合，很可能取决于正在结合的是哪些感觉模态的成对组合或其他组合，以及提取了哪些特征。例如，在自运动知觉中，视觉主要提供有关位置和速度的信息，而前庭系统则提供有关速度、加速度和加加速度的信息[7]。此外，听觉-触觉实验数据表明，频率方面存在多感官交互作用[8]，但在响度[9]或持续时间[10]方面不存在。

探索多感官整合十分重要，因为这将增进我们对感觉加工、感觉运动转换和多区域计算的认识，使我们能够开发更好的人工多感官系统[11,12]，并开始应对神经发育障碍[13]和神经精神障碍[14]；人们认为，在这些障碍中，多感官整合发生了改变，甚至可能促成相关症状。

理解意味着构建模型。也就是说，如果我们想弄清一个复杂过程，就需要将其拆分为更简单的部分，并描述这些部分之间的关系[15,16]。对于实验研究者而言，模型通常采用“方框和箭头”图的形式，并由实验数据提供支持：脑区 A 的活动 → 行为。理论研究者旨在刻画这些联系（→），通过开发能够解释神经回路如何产生驱动适当行为的神经动力学的模型。对于理论研究者而言，模型有多种形式，从仅包含少量变量的简洁数学方程，到包含数十亿个参数的人工神经网络。

在研究多感官整合时，研究人员沿着这一连续谱的不同位置开发了各种模型。在本文中，我们将这一范围划分为两类：规范性模型——从关于多感官信号应如何整合的理论出发；以及人工神经网络，旨在模拟神经回路可能如何整合这些信号。这一区分之所以有意义且有帮助，有三个原因。首先，两者的动机往往存在根本差异。规范性模型强调“自上而下”的科学方法——从某项原理出发，探究该原理是否能够以及如何由神经回路实现。相反，神经网络模型支持“自下而上”的方法——从神经回路出发，逐步理解它们所执行的计算。其次，规范性模型和神经网络模型在可解释性方面可能有所不同。大多数规范性模型只使用少量参数，因此具有很高的可解释性——例如，在一个包含两个参数的情形下，可以绘制模型拟合效果随这两个参数变化的函数图。相比之下，许多神经网络模型使用数千至数百万个参数，这使其更难以解释。第三，两者通常依据不同的标准进行验证。总体而言，规范性模型是根据其拟合实验数据的能力来验证的，而神经网络模型则根据其执行多感官任务的能力进行评判，这些任务类似于用于测试实验受试者的任务。尽管如此，个别模型可能会模糊这一界限。例如，一些机制模型利用神经网络仿真检验特定计算（如除法归一化）的作用；同时，也出现了一些将规范性要素与神经网络要素相结合的新方法。

除了帮助解释实证数据、增进我们对多感官整合的认识之外，模型还有助于：在实验应用之前检验某种分析方法的效果[17]，指导实验设计[18]，并提出新的、可通过实验检验的假设。

在这篇综述中，我们将以教学性导论的形式介绍多感官整合的规范模型与神经网络模型。最后，我们将论证，该领域最有前景的发展方向是聚焦于整合这两类模型的研究，以开发出可解释模型，能够捕捉自然情境下的多感官行为和多脑区神经活动的复杂性。

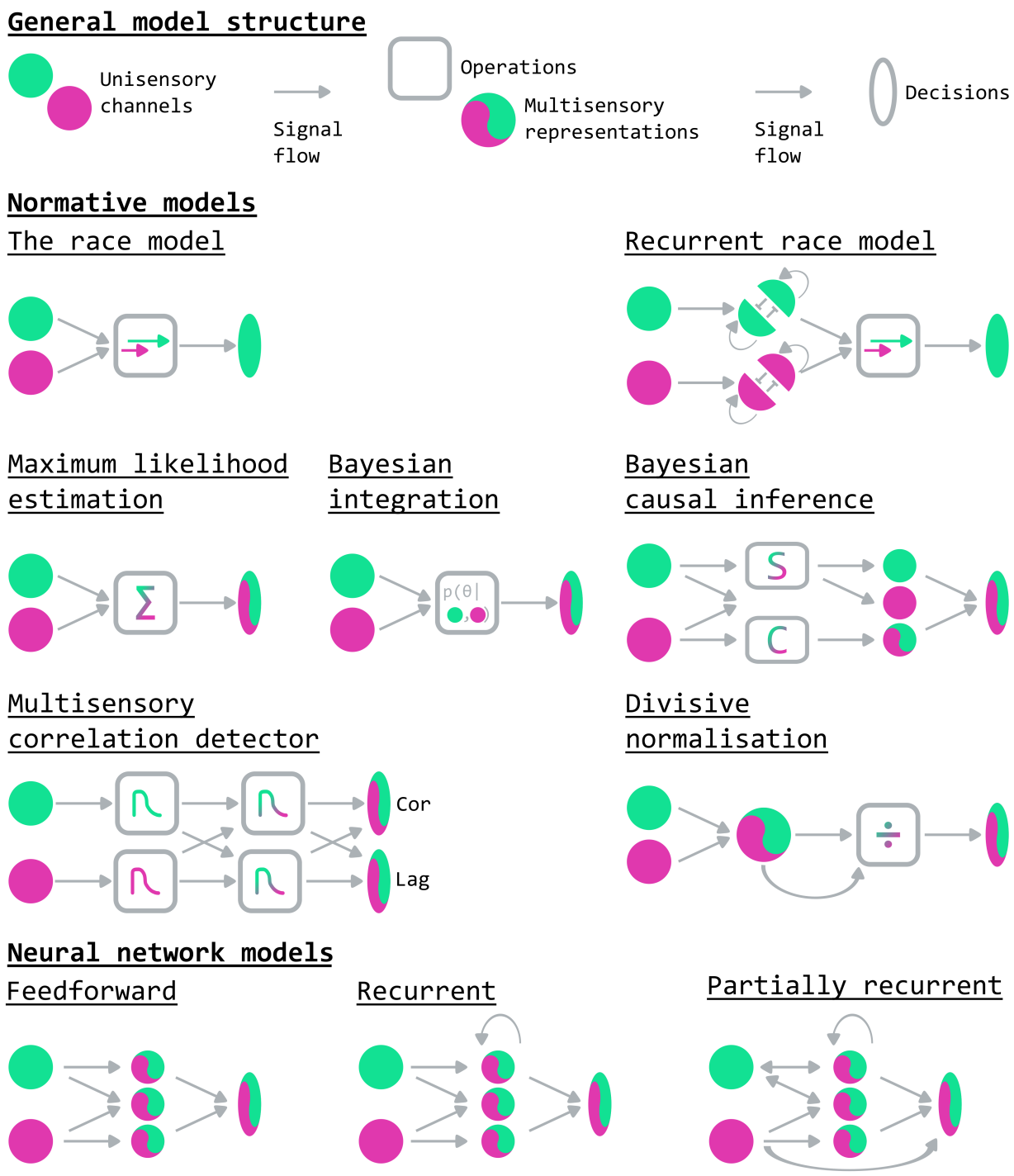

图1。 多感官整合的规范模型与神经网络模型。所有模型均采用相同的总体结构；单感官信号（绿色和粉色圆圈）输入一个或多个规范性运算（方形）或多感官表征（绿色和粉色圆圈），其输出为模型的决策提供依据（垂直椭圆）。此处列出所有模型，以便进行模型间比较。后文将再次展示每个模型，并对其进行更详细的解释。

规范性模型

加性模型

规范性模型从关于多感官信号应如何整合的假设出发，并通过考察其解释实验数据的能力来评估。例如，Meredith 和 Stein（1986年）使用加性模型（方程 1）描述猫上丘中神经元的放电率[19]：

$$FR_{Add} = FR_A + FR_V \tag{1}$$

$$Percent = \left(\frac{FR_{AV} - FR_{Add}}{FR_{Add}} \right) * 100 \tag{2}$$

其中FR_A、FR_V和FR_{AV}分别表示神经元对听觉、视觉和视听刺激的放电率，而FR_{Add}是加性模型预测的放电率。百分比指标（方程 2量化了观测到的视听反应与加性预测之间的差异程度。

虽然加性模型准确预测的神经元放电率反应寥寥无几，但它提供了一个框架，可将多感觉神经元的放电率分为超加性（当多感觉反应大于单感觉反应之和时，见图2 左图）或次加性（当多感觉反应小于单感觉放电率之和时，见图 2 右图）。

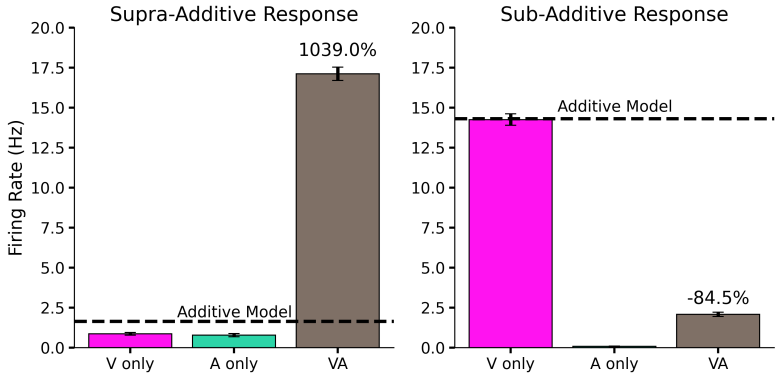

图2。 根据以下公式，对猫上丘神经元仅视觉（粉色）、仅听觉（绿色）和视听（棕色）刺激下的模拟放电率，以及加性模型进行模拟1 （虚线）。结果基于[19]。左：超加性。右：次加性

在 Python 中，我们可以编写 这是神经元的单感官和多感官放电率，使用以下函数（代码清单1).

```
导入numpy 作为np定义放电_率_add(FR_A,FR_V,FR_AV):""" 比较神经元的视听放电率与其单感觉反应之和。参数:
```

```
FR_添加 = FR _一个 + FR _V

百分比=(np.mean(FR_AV)-np.mean(FR_Add))/np.mean(FR_Add)*100
返回 百分比
```

代码清单1. 用于加性模型的 Python 代码。

识别具有超加和或次加和反应的神经元，意味着感觉通道之间存在交互作用（即多感官过程），但并未提出其潜在机制。尽管如此，最近一项针对小鼠额叶皮层的研究表明，该区域以加和的方式整合听觉和视觉证据，这与动物行为相吻合[20]。此外，一项模拟听觉和视觉反应时的建模研究发现，加性模型能在多感官试次中产生最优的提速效果[21]。

竞赛模型

竞赛模型已被用于研究反应时数据中的多感官整合[22–25]。从概念上讲，每种感觉都会“竞赛”以产生独立决策，速度最快者赢得竞赛（图3）。从技术上讲，它根据两个或更多单感官反应时分布定义了概率求和的界限[26]；这考虑到，即使没有感觉整合，从多个感觉通道中选择最快的决策，也会比从单一模态中选择产生更快的反应。

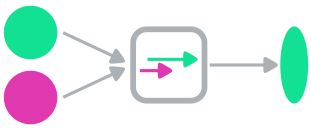

图 3. 在竞赛模型中，每种模态都独立进行加工（绿色和粉色圆圈及箭头），观察者根据最快的感觉（此处为绿色感觉）作出决策（竖直椭圆）。

在标准表述中，竞赛模型界限表示为[23]:

$$F_{AV}(t) \leq F_A(t) + F_V(t), \tag{3}$$

其中 $F_A(t)$ 和 $F_V(t)$ 表示在时间 t 之前完成的试次比例（单模态累积反应时分布），而 F_{AV} 表示累积多感官反应时分布。因此，观察到反应时快于该界限（公式3），就意味着多感觉反应无法仅用概率求和来解释，而是反映了感觉通道之间的共同激活[23]。

在假设感觉通道之间具有随机独立性的前提下，可以推导出一个更严格的界限。这意味着，听觉和视觉加工在不同试次间的波动彼此独立。实际操作中，这意味着在方程3中加入一个额外项:

$$F_{AV}(t) \leq F_A(t) + F_V(t) - F_A(t)F_V(t), \tag{4}$$

w这对应于两个独立过程的最小值的累积分布 .

图 4 展示了计算竞赛模型界限的步骤。单感觉听觉（粉色）和视觉（绿色）的反应时分布（图 4 左）被转换为累积分布函数（图 4 中），表示试次比例随反应时的变化。随后利用单感觉累积反应时分布，通过公式 3 或 4生成界限。然后从多感觉数据中减去该界限，以计算竞赛模型违背（图 4右）；正值表示多感觉反应快于统计促进所预期的时刻。在 Python 中，我们可以编写清单 2。

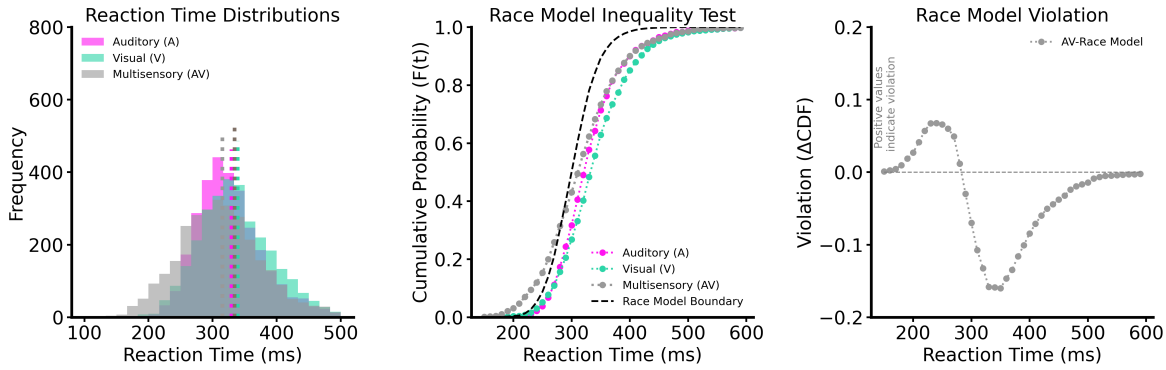

图4. 听觉（A）、视觉（V）和视听（AV）条件下的反应时分析。左：直方图展示了听觉、视觉和视听条件下反应时的分布。中：累积分布函数（CDF）展示了竞赛模型不等式检验。虚线表示竞赛模型边界，多感官累积分布函数在较短潜伏期处超过该边界。右：竞赛模型违背图（ΔCDF），显示多感官条件相对于竞赛模型预测的正偏差（违背），以及相应的线性比较。正值表示多感觉反应快于统计促进所允许的时刻。模拟结果基于[27]。

```
导入 numpy 作为 np

定义 获取_cdf(data, bins):
    # 计算累积分布函数（CDF）。
    返回 [np.mean(data <= t) 针对 t 在 bins]

定义竞速_模型_边界(RT_a,RT_v,RT_av):
    """
    参数:
        RT_a：听觉刺激的反应时（毫秒）数组。
        RT_v：视觉刺激的反应时（毫秒）数组。
        RT_av：视听刺激的反应时（毫秒）数组。
    返回值:
        竞赛_模型_边界：将是一个随时间变化的竞赛模型边界数组。
        违反量：将是竞赛模型边界与
            随时间变化的视听累积分布。
    """
    # 以 10 毫秒为步长，定义从 150 毫秒到 590 毫秒的反应时间分箱。
    # 这些分箱用于评估累积分布函数（CDF）。

    时间_分箱 = np.arange(150, 600, 10)

    # 计算听觉、视觉和视听条件的 CDF。
    F_a = np.array(get_cdf(RT_a, 时间_bins))
    F_v = np.array(get_cdf(RT_v, 时间_bins))
    F_av = np.array(get_cdf(RT_av, 时间_bins))

    比赛_模型_边界 =F_a +F_v - F_a * F_v                    # 竞赛模型界限（方程4）

    违反 = F_av - 竞赛_模型_边界

    返回 竞赛_模型_边界，违反
```

代码清单2。 用于计算竞赛模型边界及其违反情况的 Python 代码。

Otto 等人[24,28,29] 扩展了竞赛模型，以考虑刺激的一致性以及试次的重复/切换，这些因素可能会扭曲对多感官整合的解释[30]。此外，Colonius 与 Diederich（2017）提出了一个

用于刻画神经元放电的竞赛模型版本[31]。在这一背景下，该不等式在系统层面的神经生理学与行为之间架起了一座桥梁：超过竞赛模型界限意味着，组合后的感觉输入对决策过程产生了超出预期的影响，这与整合性神经机制相符。

概率模型

最大似然估计（MLE）的引入为多感官建模提供了一个预测框架[32–34]。Ernst and Banks（2002）证明，成人会以统计上最优的方式[35] 整合多感官线索，这与MLE一致。考虑视觉估计 $\hat{S}_{\mathrm{Vis}}$ 和前庭估计 $\hat{S}_{\mathrm{Vest}}$[36,37]。MLE（图5 左）预测，最优的多感官估计 $\hat{S}_{\mathrm{VisVest}}$ 是这些单感官估计的加权和（公式 5）：

$$\hat{S}_{VisVest} = w_{Vis}\hat{S}_{Vis} + w_{Vest}\hat{S}_{Vest}, \tag{5}$$

其中，每个权重都与其线索的可靠性（方差倒数）成正比，$w_i = {}^{1/\sigma^2}{}_i \Sigma_j 1/\sigma^2_j$，并假设各来源条件独立。直观地说，方差较低的感官线索会获得更大的权重，因为它们能提供对外部世界更可靠的估计。

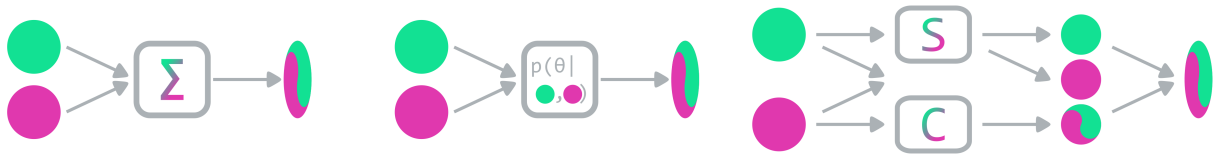

图 5. 概率模型。在最大似然估计（左图）中，每种感觉模态（绿色和粉色圆圈）都经过加权并线性组合（Σ），形成多感官估计（多色垂直椭圆）。在贝叶斯推断（中图）中，感觉输入还会与先验概率（θ）相结合。因果贝叶斯推断（右图）加入了一个先验，用于判断感觉信号是否具有共同来源、应当合并（C），还是来自相互独立的来源、应当分开（S）。

在实践中，通过将心理测量函数拟合到单感官和多感官反应数据（图 6 左图）来估计这些参数。每种条件下都会提取均值（也称为 主观相等点 ）和方差， σ^2其中方差与 最小可觉差 成正比（图 6 中图）（列表 3）。拟合值用于观察和预测单感官权重，以及多感官条件下方差的最优降低幅度（图 6 右图），该幅度由方程 6给出：

$$\sigma^2_{VisVest} = \frac{\sigma^2_{Vis}\sigma^2_{Vest}}{\sigma^2_{Vis} + \sigma^2_{Vest}}. \tag{6}$$

值得注意的是，这种由最大似然估计预测的加权方式也已在神经元层面得到观察[33,38],，为该模型提供了生物学支持。一种相关方法将最大似然估计框架扩展为预测多感官整合过程中随时间推移进行的最优证据累积[39]。最大似然估计基于可获得的感官证据进行估计，而贝叶斯整合还会纳入观察者在获取这些证据之前的预期（图5 中）。这些预期由以往经验塑造，并使知觉偏向于更常被观察到的刺激值。

尽管最大似然估计和贝叶斯整合能够成功预测许多多感官现象，但二者都假设线索具有共同来源，因此无法解释空间和时间差异如何影响整合。这些差异会影响人们将线索知觉为来自单一来源还是来自两个不同来源，并伴随不同的行为[40] 和神经生理学[38] 特征。因果推断模型（图5 右）通过在贝叶斯框架中加入信号是否源自共同来源的先验概率来解决这一问题（$C = 1$），或源自独立来源（$C = 2$）[41–43]。这些模型同时处理线索组合和线索冲突，其中共同原因的后验概率由贝叶斯定理给出（公式7）：

$$\Pr(C|x_V, x_A) = \frac{\Pr(x_V, x_A|C)\Pr(C)}{\Pr(x_V, x_A)}. \tag{7}$$

最近，这些因果推断过程也已使用人工神经网络进行建模[44,45]。

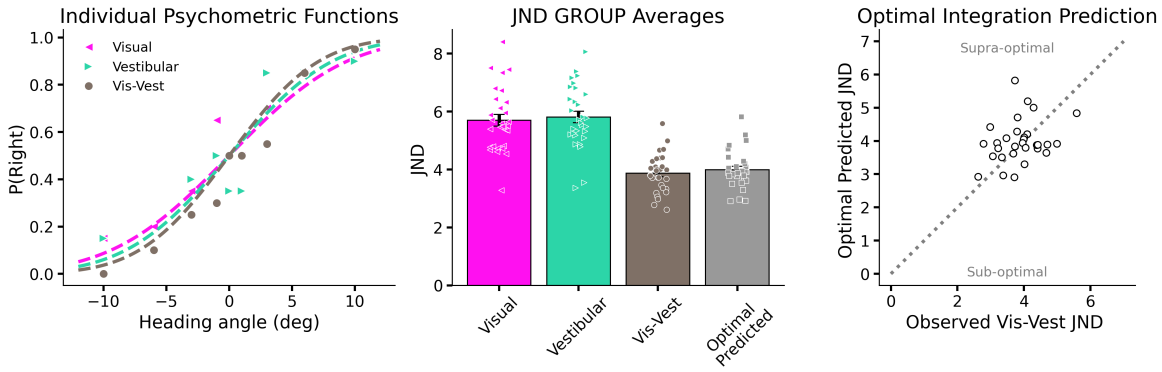

图 6. 拟合并检验多感官整合的最大似然估计模型。 **(左)** 示例参与者的个体心理测量函数，显示了在单感觉视觉（粉色）和前庭（绿色）线索，以及视觉-前庭（Vis-Vest）组合条件（棕色）下，“向右”反应的比例随航向角的变化。虚线表示累积高斯拟合，据此可提取各条件下的主观相等点（均值）和最小可觉差（JND；与 σ有关）。**(中)** 视觉、前庭和 Vis-Vest 条件下的组平均最小可觉差，以及根据单感觉方差由最优线索整合预测的最小可觉差（公式 6）。误差线表示均值标准误。实测的 Vis-Vest 最小可觉差与最优预测高度吻合，这与统计最优线索组合一致。**(右)** 个体参与者预测与实测的最小可觉差（空心圆），一致线表示模型与数据完全吻合。靠近该线的点表明，最大似然估计能很好地预测个体层面的多感觉优势。基于[37]的模拟。

```
定义 预测_最优_组合_最小可觉差（航向_角，视觉_本体感觉_右侧，前庭_本体感觉_右侧，
                              视觉前庭_本体感觉_右侧）：
    """
    计算视觉-前庭航向
    航向任务。

    参数:
        航向_角：测试的航向角（度）
        visual_prop_right：视觉条件下各航向角的向右反应比例
            视觉条件
        vest_prop_right：前庭条件下各航向角的向右反应比例
            前庭条件
        visvest_prop_right：视觉-
            前庭条件
`
        拟合_心理测量_函数，使其符合累积高斯函数，并返回
        主观相等点 (PSE) 和最小可觉差 (JND)。


    返回:
        预测的_jnd：预测的最佳多感觉最小可觉差
        visvest_jnd：观测到的视觉-前庭最小可觉差
    """
    vis_pse, vis_jnd=拟合_心理测量_函数(
        航向_角度，视觉_本体感觉_右)
    前庭_主观相等点，前庭_最小可觉差 = 拟合_心理测量_函数(
        航向_角度，前庭_本体感觉_右)
    视觉前庭_主观相等点，视觉前庭_最小可觉差 = 拟合_心理测量_函数(
        航向_角度，视觉前庭_本体感觉_右)

    变量_视觉，变量_前庭 = 视觉_最小可觉差**2，前庭_最小可觉差**2
    预测的_最小可觉差 = np.sqrt(var_vis * var_vest / (var_vis + var_vest))

    返回 预测的_最小可觉差，视觉-前庭_最小可觉差
```

列表 3。 用于计算视觉-前庭航向任务的最大似然最小可觉差的 Python 代码。

机制模型

机制模型旨在通过提出共同的神经元层面的计算来解释实验观察结果[46]。

除法归一化被提出作为一种机制，用于解释多感觉神经元的反应特征[47,48]。根据该模型，每个多感觉神经元的活动都会除以多感觉神经元群体的综合活动（图7）。Ohshiro等人（2011年）表明，除法归一化可以解释多感官整合的经典经验原理，包括逆有效性以及空间增强和抑制[48]。他们的双层模型再现了在上丘、腹侧顶内区和背侧内侧上颞区神经元中观察到的反应模式。

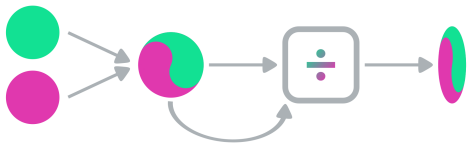

图7。在除法归一化中，单感官信号（绿色和粉色圆圈）由多感觉神经元群体（多色圆圈）进行整合，其活动在彼此之间进行归一化（÷），随后参与决策（多色垂直椭圆）。

另一类机制模型将感觉信号之间的时间相关性作为判断是否应进行整合的依据[49–51]。受 Hassenstein–Reichardt运动检测器[52], 的启发，Parise和Ernst（2016）提出了一种多感觉相关性检测器[49], ，通过模态特异性的时间滤波和乘性交互来检测跨模态相关性、同步性和时间延迟（图8）。该模型再现了广泛的人类心理物理学发现，例如贝叶斯最优线索整合，同时也解释了这种整合在时间冲突下为何会失效。

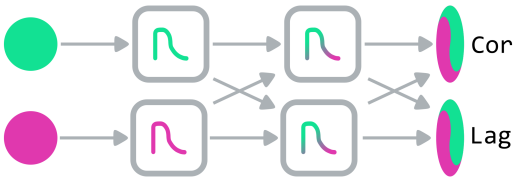

图 8。在多感觉相关性检测器中，单感官信号（绿色和粉色圆圈）由一组时间滤波器（方框）处理。随后，这些输出可用于计算不同感官之间的相关性和时间延迟（多色竖向椭圆）。

最后一类机制模型采用神经网络仿真，其中参数（例如神经元间连接的强度）由人工设定；设定依据可以是实验数据，也可以是系统地探索一系列取值。这些模型能够检验在面对模拟刺激时，不同连接强度或整合规则所产生的影响。

Magosso等人，2008年和Ursino等人，2009年使用一种神经网络架构对上丘中的多感官整合进行建模，其中两个单模态区域和一个多模态区域通过前馈连接（单模态到多感官）和反馈连接（多感官到单模态）相互连接[53,54]。在给定一组固定参数的情况下，这些模型表现出超加和性和次加和性、逆有效性、跨模态增强和跨模态抑制；这表明，这些经典的多感官现象都可以源自一个相对简单的网络模型。

同样，Brady与Butler（2025年）采用了递归平均场模型，其中微分方程描述群体放电率的时间演变（图 9）[21]。他们表明，其他架构——包括胜者全得、线性整合和依赖历史的机制——都可以再现多感官反应优势，同时基于不同的机制假设。与其他机制模型一样，这种方法将电路层级架构与行为特征联系起来；此外，它还提供了一个框架，用于考察神经元群体的时间动力学和其他整合机制如何产生相似的行为结果。

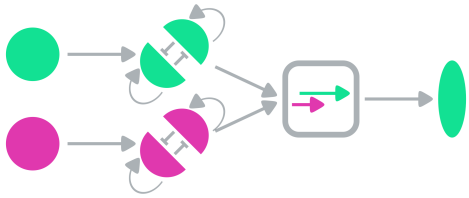

图 9。 在循环竞赛模型中，每个单感觉通道（绿色和粉色圆圈）均由多个神经元群体（半圆）进行处理，并具有群体内兴奋性连接（灰色箭头）和群体间抑制性连接（灰色条）。最终决策（垂直椭圆）取决于哪个群体首先达到决策阈值（此处为绿色群体中的一个）。

总体而言，多感官整合的规范性模型有助于解释许多实验观察到的现象。这些模型的主要优势在于可解释性很强，因为它们明确提出假设，且使用的参数较少。然而，这些模型的主要缺点是，它们拟合实验数据的能力有限，而且能够处理的任务复杂度也有限。

神经网络模型

大脑通过在相互连接的神经元群体之间传递信息来转换感觉输入。类似地，人工神经网络通过在相互连接的单元层之间传递信号来进行计算。利用机器学习技术[55]，人工网络可以接受训练，以执行复杂的多感官任务，这些任务类似于用于测试实验（人类或动物）受试者的任务。由于这些模型是完全可观测的，因此可以进一步研究其学习到的连接性、神经活动和行为，观察它们是否能够再现实验中观察到的现象。此外，由于这些网络是完全可塑的，因此可以用于开展计算机模拟实验，探索超出体内实验可能范围的内容。通过这种方式，神经网络模型可以用于检验现有实验工具是否能够提供有意义的洞见[56]，，为新方法提供基准测试[57] 并提出新的假设。

在本节中，我们回顾具有不同架构（即神经元之间连接模式）的网络如何以这种方式得到应用，并重点介绍这些研究带来的见解。

前馈网络

在最简单的架构之一中，信号从感觉输入向前传递至隐藏神经元，再传递至决策输出（图10）。

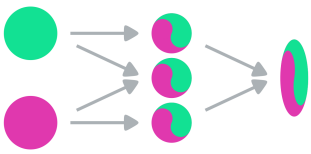

图10。在前馈神经网络中，单感官信号（绿色和粉色圆圈）向前传递（灰色箭头）至多感觉神经元（多色圆圈），然后传递至决策输出（多色竖椭圆）。

我们可以用数学方式对一个包含三个神经元群体（输入 - i、隐藏层 - h 和输出 - o）的前馈网络进行建模，如下所示（公式8）：

$$\hat{y} = g(W_{ho} f(W_{ih} x)) \tag{8}$$

其中 x 是表示网络多感官输入的向量，W_{ih} 和 W_{ho} 是权重矩阵，用于指定神经元之间连接的强度（$W_{from-to}$），f 和 g 表示激活函数，它们决定每个神经元的行为，而 $\hat{y}$ 是表示网络输出的向量。请注意，为清晰起见，我们省略了偏置项。在 PyTorch[58]，（一个用于使用 Python 构建神经网络的软件包）中，我们可以这样编写（清单 4）：

```
1 import torch 2 import torch.nn.functional as F 3 4 def feedforward(x, W_ih, W_ho, f=F.relu, g=F.soft
```

ma
数:
隐藏
hid

列表 4。 用于前馈神经网络的 PyTorch 代码

通过调整神经网络的参数（在本例中为 W_{ih} 和W_{ho}中的连接强度），我们可以训练网络，将其输入（x）转换为特定的输出（$\hat{}y$）[55]。因此，我们可以训练网络执行多感官任务。例如，针对随机点运动和侧化听觉信号等方向性刺激，选择左侧或右侧。

通过在多感官任务上训练前馈网络，研究人员观察到了多种多样的多感官现象。在行为层面，这些网络能够最优地整合来自多个感觉通道的信号[59,60] 并进行因果推断[45]。在神经元层面，研究人员发现了一致神经元和相反神经元[45]，具有亚加和、加和反应和超加和反应的神经元[59,60]，以及表现出逆有效性和跨模态抑制的神经元[60]。总体而言，这些结果表明，无需规范性模型所提出的某些机制，多感官整合的若干标志性特性也能在经过任务训练的前馈网络中涌现。例如，除法归一化通常被认为涉及侧向连接或反馈连接，以及一个明确的归一化池[47,48]，而预测性多感官加工可能需要自上而下的反馈连接[61]。这并不是说多感官回路中不存在这些连接类型，而是说明，对于这些特定的多感官行为和神经元特性，前馈网络已经足够。

超越观察层面，研究人员利用虚拟消融来检验具有不同特性的多感觉神经元对于特定任务的必要性。具体而言，Rideaux 等，2021证明了相容神经元和相反神经元在因果推断中的作用[45]，，而 Ghosh 等，2024证明了在多感官信号密集或稀疏的情境下，加性/超加性神经元对于任务表现的相对重要性[59]。

总体而言，这些结果表明，数量惊人的多感官现象都可以由纯前馈网络再现，并突显了使用这些可观测且易于调整的模型的优势。然而，由于这些模型缺乏记忆或状态，因此并不适合处理具有自然时间变化的多感官信号的任务，也不适合建模神经动力学的复杂性[62–64]。

循环网络

标准循环神经网络（RNN）[65] 在隐藏层中的神经元之间增加了一组额外的连接（W_{hh}），图 11。这使这些神经元能够将传入的感觉信号（x_t）与网络当前的内部状态（h_{t-1}）相结合；后者可以充当网络过去输入和隐藏状态的记忆。

从数学上来说，我们可以在前馈方程（8）中引入一个单独的项，以对循环神经网络（方程 9）进行建模 ：

$$\hat{y_t} = g(W_{ho} f(W_{ih} x_t + W_{hh} h_{t-1})) \tag{9}$$

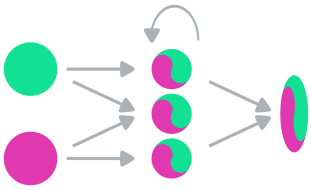

图11。 在循环神经网络中，多感觉神经元（多色圆）将传入的感觉信号（绿色和粉色圆，通过灰色从左向右的箭头表示）与网络的当前状态（灰色从右向左的箭头表示）相结合。这些多感觉激活塑造了网络的输出（多色垂直椭圆）。

在 Python 中，我们可以如下修改前馈代码（清单 5）：

```
1 import torch 2 import torch.nn.functional as F 3 4 def recurrent(x, h_tm1, W_ih, W_hh, W_ho, f=F.relu, g=
```

F.sc
需要
在
11 V
输出
藏单
@ h

Li第 5 部分。循环神经 网络

在行为层面，Yamashita 等，2013 证明，具有固定连接的循环神经网络能够适当地分离或整合多感官刺激：相邻多感觉神经元之间为兴奋性连接，而相距较远的神经元之间为抑制性连接；具体取决于这些刺激是否来自相同或不同的来源[66]。然而，循环神经网络模型的一个关键优势在于，它能够学习使用复杂的时变信号执行多感官任务。例如，Anil 等，2025 开发了一项多感觉捕食者—猎物任务，其中观察者（扮演捕食者）必须估计猎物运动方向的总体方向，但猎物只会发出稀疏信号脉冲，仿佛它们正在藏身处之间快速穿梭[67]。值得注意的是，这项任务对规范性贝叶斯模型而言已经过于复杂！也就是说，人们可以写出解决方案，但在计算上无法实际求解[67,68]。相比之下，前馈网络可以在每个时间步估计猎物的运动，并且随着时间累积这些估计结果，可以取得合理的表现。但循环神经网络能够利用其内在记忆来检测猎物发出的信号脉冲，因此表现显著更好[67]。这凸显出，研究自然情境任务可能很快就需要神经网络，而非规范性模型。

在神经层面上，Wierda 等人（2025年）表明，多感官单元[69]在循环神经网络中会学习到异质的调谐模式（例如，对感觉变量或输出变量的调谐）。Terada 与 Toyoizumi（2024年）则证明，通过对隐藏层到隐藏层权重矩阵（W_{hh}）施加强增益参数来诱导混沌神经动力学的多感官循环神经网络，能够学习输出估计序列，使其分布与贝叶斯先验和后验分布相匹配[44]。

基于标准循环神经网络架构以及真实多感官回路的复杂性，多项研究构建了具有额外连接通路的模型。Ursino 等人（2009年）在其模型的听觉（W_{aa}）和视觉区域（W_{vv}）神经元之间加入了递归连接，并探究了模型的单模态与多模态递归之间的相对强度平衡如何影响模态内抑制和跨模态抑制[54]。同样，Magosso 等人（2008年）和 Tugsbayer 等人（2024年）都探究了具有多个递归区域的模型中反馈连接（从多感官区域到单感官区域，W_{hi}）的作用[54,70]。在此基础上，Ghosh 和 Goodman（2026年）系统地研究了具有不同递归连接、反馈连接和跳跃连接组合的128种多感官回路（称为部分递归神经网络）[71]。他们发现，不同架构的表现存在显著差异（包括任务表现、学习速度以及对各种扰动的鲁棒性），并通过证明不同架构会学习具有不同刺激敏感性和记忆动态的解决方案，解释了这些差异[71]。

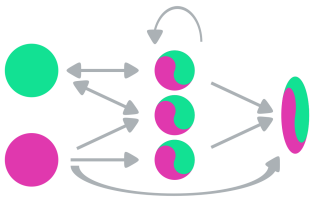

图12. 在部分递归神经网络中[71]，信号会通过不同组合的通路，在多个区域之间传播。例如，信号从单感觉输入（绿色和粉色圆圈）向前传递至一组多感觉神经元（多色圆圈），再传至决策输出（多色椭圆）；同时，信号也会从隐藏神经元反向传至绿色的“单感觉”神经元，在多感觉神经元之间横向传递，并直接从紫色单感觉神经元传至决策输出，绕过隐藏层。其他部分递归神经网络也会采用不同组合的通路，因此形成不同的信号流模式。

替代架构

多模态机器学习的进展为解决尚未解决的神经科学问题[12]提供了新的架构、概念和工具。例如，要探究感觉模态应在早期还是晚期融合[3]，，可以构建、训练并比较具有不同多感官汇聚点的网络。在这方面，研究表明，早期融合能够提升多模态卷积神经网络的性能[72] 和抗噪鲁棒性[73] 。此外，Nagrani 等人（2021）使用多模态Transformer[74]，证明，早期但受限的跨模态通信既能提升性能，又能提高能效[75]。

整合规范性模型与神经网络模型

上文介绍了许多多感官整合模型，并强调了该领域从中获得的洞见。现在，展望未来，我们将思考我们究竟可以期望从多感官整合的规范模型与神经网络模型中了解到什么？

通过将**规范性模型**拟合到实验数据，可以：了解给定模型对数据的拟合程度，比较不同模型之间的拟合效果以确定哪个模型对数据的解释最佳，并解读拟合参数，从而深入了解整合过程。这种方法的一个杰出例子来自 Coen 等人，2023 年，他们训练小鼠执行视听定位任务[20]。训练后，一个相对简单的加性模型便能很好地解释小鼠的行为和神经活动，但几乎没有证据表明，更复杂的多感官交互项能够提高解释力。

规范性模型的关键优势在于，它们通常只使用少量参数，因此具有很高的可解释性。然而，它们的主要局限是所谓的模型退化[76,77]现象。也就是说，我们可以设想出大量关于多感官整合的规范性模型，而在实践中，许多模型都可能同样出色地解释相同的数据。遵循奥卡姆剃刀原则，解决这一问题的一种方法是优先选择更简单的模型。例如，具有纯前馈输入和 S 形[53,60] 或泄漏积分发放[59] 输入—输出函数的基础神经元模型，能够表现出逆有效性并产生从次加和到超加和的状态，因此，对于这些多感官现象而言，除法归一化[48] 或调节性“反馈”[78] 等机制可能并非必要。另一种解决方案是设计能够明确区分候选模型的实验。比如，Ghosh 等，2024 引入了一组非线性多感官任务，有助于直接比较加和整合与非加和整合[59]；在最极端的情境下，任何表现显著高于随机水平的观察者都必须

采用非加性策略。展望未来，该领域甚至可以发展出生成式对抗协作——让不同理论接受正面交锋式的明确检验[79,80]。

通过开发多感官整合的**神经网络模型**，我们可以探讨 Marr的三个层次[81–83]上的问题。在计算层次，我们可以研究哪些类型的网络能够实现哪些行为，甚至创建模拟环境，探索多感官整合在哪些条件下更具优势。在算法层次，我们可以研究这些模型所使用的行为算法和神经算法——借助神经科学和机器学习可解释性的方法，例如解码[84] 或神经元层面消融[57]。在实现层次，我们可以考量某种算法由神经元网络实现的可行性。例如，大多数多感官研究都聚焦于两种感官的组合，但针对感官数量建立标度律，或许有助于我们区分不同模型的合理程度。

神经网络模型的主要优势在于能够学习复杂的心理物理学任务或自然情境任务，并生成神经元或神经回路层面预测。其一项常被提及的局限是缺乏可解释性，或具有“黑箱”性质。不过，情况并非必然如此。最近，Ji-An等人，2025年证明，小型循环神经网络（包含1至4个隐藏神经元）可以经过训练来模拟逐试次行为数据，其表现优于经典认知模型；同时，它们仍足够小，可以借助动力系统理论的方法进行解释[85]。类似地，Ghosh 等，2024年表明，包含两个隐藏神经元的网络可以经过训练来执行多感官心理物理学任务[59]；而Brady与Butler，2025年则仅用少量参数模拟了循环多感官网络[21]。即使必须使用更复杂的模型，它们也为开发和测试理解生物网络所需的工具奠定了基础[56]。更深层的挑战在于，使用不同训练数据、架构和超参数（甚至仅仅是不同随机种子）的网络，可能会学习到具有不同表现和神经表征的解决方案。虽然这为研究这些组成部分如何塑造网络功能[86,87], 提供了机会，但也凸显出报告不同随机种子和超参数组合下多个神经网络模型结果的重要性[88,89]。我们不应只报告我们的神经网络模型重现了某种行为，而应力求报告我们的模型在多大程度上、以及在什么条件下做到了这一点。

展望未来，正在出现一些整合规范性模型与神经网络模型的方法，以克服二者各自的局限。Eckstein等人， 2026[90]将神经网络与规范性组件相结合，并利用这些混合模型对人类奖励学习进行建模。Castro等人，2025年则采用了一种新颖的搜索算法——在该算法中，大语言模型监督一个进化过程[91], ，以自动发现能够在心理物理任务中准确描述人类和动物行为的规范性模型[92]。这些令人振奋的方法表明，该领域最有前景的未来方向可能是将二者结合起来，开发能够捕捉自然行为复杂性、同时保持可解释性的模型。

结论

归根结底，规范性模型与神经网络模型为理解多感官整合提供了互补的途径。规范性模型阐明了在关于不确定性和因果结构的特定假设下，观察者应执行哪些计算。神经网络模型则展示了这些计算如何从具有特定架构、目标函数和学习规则的网络中涌现。未来的挑战在于构建不仅准确、还能提出可检验假设的模型：其失效之处能够区分不同理论，其内部表征能够生成可检验的神经预测，其假设也能够通过实验加以操控。有鉴于此，最有前景的方向或许不是在规范性模型与神经网络模型之间做出选择，而是发挥二者的优势并将其整合起来。

术语表

- **激活函数**——一种用于计算人工神经网络中神经元输出的变换。
- **一致性神经元** — 调谐偏好在不同感觉模态间一致的多感觉神经元。例如，偏好向右运动的视觉刺激和前庭刺激的神经元。
- **跨模态增强** — 多感觉反应超过最佳单模态反应。
- **跨模态抑制** — 多感觉反应低于最佳单模态反应。
- **隐藏神经元**— 人工神经网络中既非输入单元也非输出单元的神经元。
- **逆有效性** — 较弱的单模态刺激往往会产生较大的多感觉反应，这通常与两个较弱的单感觉反应有关。

- **最小可觉差**——观察者能够可靠辨别的两个刺激之间的最小差异（例如强度、方向或速度上的差异），通常用作知觉（或神经）敏感度／精确度的衡量指标。

- **多感觉因果推断**——大脑（隐式或显式）推断两个或多个感觉信号是否源自共同原因，因此应当整合；还是源自不同原因，因此应分别处理（分离）。

- **多感官整合**——神经系统将来自两种或多种感觉模态（或同一感觉模态中的多个线索）的信息整合为单一知觉估计的过程；该估计通常更精确、更可靠。

- **多感觉最优整合**——通过按照各单感觉线索自身方差的倒数进行加权，获得方差（或最小可觉差）尽可能低的多感官整合估计；这符合最大似然／贝叶斯线索组合的原理。

- **相反神经元**——其跨感觉模态的调谐偏好方向相反的多感觉神经元。例如，在一种感觉模态中偏好向右运动，而在另一种模态中偏好向左运动。
- **主观相等点**——观察者认为两个刺激（或一个刺激与一个参照刺激）相等时的刺激值，通常是心理测量函数上的 50% 点，可用作衡量知觉偏差的指标。
- **冗余信号效应（RSE）**——与单独呈现任一信号相比，同时呈现两个或更多冗余的任务相关感觉信号时，观察到的行为反应加速（反应时缩短）和／或准确率提升。

次加性——对神经元多感官放电率的分类；对多感官刺激的反应低于其各单一感觉模态反应之和。

- **超加性**— 对神经元多感官放电率的分类：对多感官刺激的反应大于其单模态反应之和。

参考文献

1. Ghazanfar, A. A. & Schroeder, C. E. 新皮层本质上是多感官的吗？《认知科学趋势》10, 278–285（2006）。2. Driver, J. & Noesselt, T. 多感官交互揭示了对“感觉特异性”脑区、神经反应和判断的跨模态影响。《神经元》57, 11–23（2008）。3. Bizley, J. K., Jones, G. P. & Town, S. M. 知觉决策中的多感官信号在哪里进行整合？《神经生物学观点》40, 31–37（2016）。4. Falchier, A., Clavagnier, S., Barone, P. & Kennedy, H. 灵长类纹状皮层多模态整合的解剖学证据。《神经科学杂志》22, 5749–5759（2002）。5. Cappe, C. & Barone, P. 支持猴低级皮层加工中多感官整合的异模态连接。《欧洲神经科学杂志》22, 2886–2902（2005）。6. Stein, B. E. & Stanford, T. R. 多感官整合：从单个神经元视角看当前问题。《自然综述：神经科学》9, 255–266（2008）。7. Butler, J. S., Campos, J. L. & Bülthoff, H. H. 在感觉间运动轮廓相互冲突的条件下实现最优视觉—前庭整合。《实验脑研究》233, 587–597（2015）。8. Yau, J. M., Olenczak, J. B., Dammann, J. F. & Bensmaia, S. J. 听觉与触觉之间存在关联的时间频率通道。《当代生物学》19, 561–566（2009）。9. Yau, J. M., Weber, A. I. & Bensmaia, S. J. 听觉—触觉音高交互与响度交互的机制彼此独立。《心理学前沿》1, 160（2010）。10. Butler, J. S. et al. 听觉与触觉时长加工中共同或冗余的神经回路。《神经科学杂志》31, 3400–3406（2011）。11. Baltrušaitis, T., Ahuja, C. & Morency, L.-P. 多模态机器学习：综述与分类体系。《IEEE模式分析与机器智能汇刊》41, 423–443（2018）。12. Liang, P. P., Zadeh, A. & Morency, L.-P. 多模态机器学习的基础与最新趋势：原理、挑战和开放问题。arXiv 预印本 arXiv:2209.03430 4（2022）。13. Dionne-Dostie, E., Paquette, N., Lassonde, M. & Gallagher, A. 多感官整合与儿童神经发育。《脑科学》5, 32–57（2015）。14. Hornix, B. E., Havekes, R. & Kas, M. J. 神经精神疾病谱中的多感觉皮层加工及其功能障碍。《神经科学与生物行为评论》97, 138–151（2019）。15. Beer, R. D., Barwich, A.-S. & Severino, G. J. 挤奶球形奶牛：神经科学中的玩具模型。《欧洲神经科学杂志》60, 6359–6374（2024）。16. Ghosh, M. 不要儿戏：神经科学为何需要玩具模型。《The Transm.》（2025）。17. Zhu, H., Beierholm, U. & Shams, L. 多感官整合分析：该做与不该做之事。bioRxiv 2026–04（2026）。18. Jones, P. R. 线索组合与信号检测理论教程：利用敏感度变化评估观察者如何整合感觉信息。《数学心理学杂志》73, 117–139（2016）。19. Meredith, M. A. & Stein, B. E. 上丘细胞汇聚视觉、听觉和躯体感觉输入，从而产生多感官整合。《神经生理学杂志》56, 640–662（1986）。20. Coen, P., Sit, T. P., Wells, M. J., Carandini, M. & Harris, K. D. 小鼠额叶皮层介导加性多感官决策。《神经元》111, 2432–2447（2023）。21. Brady, R. M. & Butler, J. S. 利用循环平均场网络建模视听反应时间。bioRxiv DOI: 10.1101/2025.05.29.656149（2025）。https://www.biorxiv.org/content/early/2025/06/01/2025.05.29.656149.full.pdf。22. Raab, D. H. 简单反应时的统计促进。《纽约科学院汇刊》（1962）。23. Miller, J. 分配性注意：冗余信号共同激活的证据。《认知心理学》14, 247–279（1982）。24. Otto, T. U. & Mamassian, P. 多感官决策：竞赛模型的检验、逻辑与效力。《多感觉研究》30, 1–24（2017）。25. Roberts, K. & Otto, T. U. 一场竞赛并不足够时：接力模型对多感觉反应时的解释。bioRxiv 2025–07（2025）。

26. Diederich, A. & Colonius, H. 双模态和三模态多感官增强：刺激起始时间和强度对反应时的影响。知觉与心理物理学 66, 1388–1404 (2004). 27. Brandwein, A. B. 等。高功能自闭症中多感官整合的发展：高密度电生理映射和心理物理测量揭示了视听输入处理方面的缺陷。大脑皮层 23, 1329–1341 (2013). 28. Otto, T. U., Dassy, B. & Mamassian, P. 多感觉行为的原理。《神经科学杂志》33, 7463–7474 (2013). 29. Roberts, K., Jentzsch, I. & Otto, T. U. 语义一致性调节多感觉反应的加速效应。科学报告 14, 567 (2024). 30. Shaw, L. H. 等。在多感觉情境中运作：评估多感觉反应时促进效应与感觉间任务切换效应之间的相互作用。Neuroscience 436, 122–135, DOI: https://doi.org/10.1016/j.neuroscience. 2020.04.013 (2020). 31. Colonius, H. & Diederich, A. 测量多感官整合：从反应时到脉冲计数。科学报告 7, 3023 (2017). 32. Yuille, A. L. & Bülthoff, H. H. 贝叶斯决策理论与心理物理学。载于《作为贝叶斯推断的知觉》（剑桥大学出版社，1993）。33. Anastasio, T. J., Patton, P. E. & Belkacem-Boussaid, K. 使用贝叶斯定理对上丘中的多感官增强进行建模。神经计算 12, 1165–1187, DOI: 10.1162/089976600300015547 (2000). 34. Alais, D. & Burr, D. 腹语者效应源于近最优双模态整合。《当代生物学》14, 257–262 (2004). 35. Ernst, M. O. & Banks, M. S. 人类以统计最优方式整合视觉和触觉信息。《自然》415, 429–433 (2002). 36. Butler, J. S., Smith, S. T., Campos, J. L. & Bülthoff, H. H. 为确定航向而对视觉和前庭信号进行贝叶斯整合。视觉杂志 10, 23–23 (2010). 37. Butler, J. S., Campos, J. L., Bülthoff, H. H. & Smith, S. T. 立体视觉在视觉—前庭整合中的作用。观看与知觉 24, 453–470 (2011). 38. Gu, Y., Angelaki, D. E. & DeAngelis, G. C. 猕猴 MSTd 中多感觉线索整合的神经相关物。自然—神经科学 11, 1201–1210 (2008). 39. Drugowitsch, J., DeAngelis, G. C., Klier, E. M., Angelaki, D. E. & Pouget, A. 反应时任务中的最优多感觉决策。eLife 3, e03005, DOI: 10.7554/eLife.03005 (2014). 40. Wallace, M. T. 等。跨时间与空间统一多感官信号。实验脑研究 158, 252–258, DOI: 10.1007/s00221-004-1899-9 (2004). 41. Körding, K. P. 等。多感觉知觉中的因果推断。公共科学图书馆·综合 2, e943, DOI: 10.1371/journal.pone.0000943 (2007). 42. Sato, Y., Toyoizumi, T. & Aihara, K. 贝叶斯推断解释统一知觉和腹语者后效应：识别视听刺激的共同来源。神经计算 19, 3335–3355 (2007). 43. MacNeilage, P. R., Banks, M. S., Berger, D. R. & Bülthoff, H. H. 视觉线索对重力惯性力进行消歧的贝叶斯模型。实验脑研究 179, 263–290 (2007). 44. Terada, Y. & Toyoizumi, T. 混沌神经动力学通过采样促进概率计算。美国国家科学院院刊 121, e2312992121 (2024). 45. Rideaux, R., Storrs, K. R., Maiello, G. & Welchman, A. E. 多感觉神经元如何解决因果推断问题。美国国家科学院院刊 118, e2106235118 (2021). 46. Froudist-Walsh, S. 等。跨尺度和物种的认知机制性计算模型。生物精神病学 DOI: https://doi.org/10.1016/j.biopsych.2026.08.017 (2026). 47. Beck, J. M., Latham, P. E. & Pouget, A. 采用除法归一化的神经回路中的边缘化。《神经科学杂志》31, 15310–15319 (2011). 48. Ohshiro, T., Angelaki, D. E. & DeAngelis, G. C. 多感官整合的归一化模型。Nat. Neurosci. 14, 775–782, DOI: 10.1038/nn.2815 (2011).

49. Parise, C. V. & Ernst, M. O. 相关性检测作为多感官整合的一般机制。Nat. Commun. 7, DOI: 10.1038/ncomms11543 (2016). 50. Pesnot Lerousseau, J., Parise, C. V., Ernst, M. O. & van Wassenhove, V. 通过时间分辨编码模型识别人脑中的多感觉相关性计算。Nat. communications 13, 2489 (2022). 51. Parise, C. V. & Ernst, M. O. 多感官整合以单模态瞬态通道的相关输入为基础。eLife 12, RP90841 (2025). 52. Hassenstein, B. & Reichardt, W. 象鼻虫 Chlorophanus 运动知觉中对时间、顺序和符号的系统理论分析。Zeitschrift für Naturforschung B 11, 513–524 (1956). 53. Magosso, E., Cuppini, C., Serino, A., Di Pellegrino, G. & Ursino, M. 通过神经网络模型对上丘多感官整合进行理论研究。Neural Networks 21, 817–829, DOI: 10.1016/j.neunet.2008.06.003 (2008). 在线发表：2008年6月22日。54. Ursino, M., Cuppini, C., Magosso, E., Serino, A. & Di Pellegrino, G. 上丘中的多感官整合：神经网络模型。《计算神经科学杂志》26, 55–73 (2009). 55. Bishop, C. M. & Bishop, H. 深度学习：基础与概念（Springer，2024）。56. Jonas, E. & Kording, K. P. 神经科学家能理解微处理器吗？《公共科学图书馆·计算生物学》13, e1005268 (2017). 57. Fakhar, K. & Hilgetag, C. C. 对人工神经网络进行系统性扰动：迈向量化大脑中的因果贡献。公共科学图书馆·计算生物学 18, e1010250 (2022). 58. Paszke, A. 等。PyTorch：一种命令式、高性能深度学习库。神经信息处理系统进展 32 (2019). 59. Ghosh, M., Béna, G., Bormuth, V. & Goodman, D. F. 非线性融合对于广泛一类多感官任务均为最优。公共科学图书馆·计算生物学 20, e1012246 (2024). 60. Farahmandi, A., Abedi Khoozani, P. & Blohm, G. 采用加性前馈网络实现跨参考系的多感官整合。bioRxiv 2024–11 (2024). 61. Talsma, D. 预测编码与多感官整合：对多感官心智的一种注意理论解释。Front. integrative neuroscience 9, 19 (2015). 62. Chandrasekaran, C. 多感官整合的计算原理与模型。《神经生物学观点》43, 25–34 (2017). 63. Senkowski, D. & Engel, A. K. 用于多感官整合的多时间尺度神经动力学。Nat. Rev. Neurosci. 25, 625–642 (2024). 64. Bolhasani, E., Aboutalebi, S. H. & Merrikhi, Y. 循环神经网络多感官整合计算模型：批判性综述与未来方向。Adv. Intell. Syst. 8, 2500147 (2026). 65. Elman, J. L. 在时间中寻找结构。Cogn. science 14, 179–211 (1990). 66. Yamashita, I., Katahira, K., Igarashi, Y., Okanoya, K. & Okada, M. 用于多感官整合的循环网络——识别视听刺激的共同来源。Front. Comput. Neurosci. 7, 101 (2013). 67. Anil, S., Goodman, D. F. & Ghosh, M. 跨通道和时间融合多感官信号。公共科学图书馆·计算生物学 21, e1013125 (2025). 68. Ma, W. J., Kording, K. P. & Goldreich, D. 知觉与行动的贝叶斯模型：导论（MIT Press，2023）。69. Wierda, T. S., Dora, S., Pennartz, C. M. & Mejias, J. F. 在针对多感觉决策进行训练的循环神经网络中，涌现出多样且灵活的行为策略。公共科学图书馆·计算生物学 21, e1013559 (2025). 70. Tugsbayar, M., Li, M., Muller, E. B. & Richards, B. 自上而下反馈至关重要：类脑连接基元对视听整合的功能影响。bioRxiv 2024–10 (2024). 71. Ghosh, M. & Goodman, D. F. 部分循环可实现稳健且高效的计算。Commun. AI & Comput. 1, 11 (2026). 72. Gadzicki, K., Khamsehashari, R. & Zetzsche, C. 多模态卷积神经网络中的早期融合与晚期融合。载于2020年IEEE第23届国际信息融合会议（FUSION），1–6（IEEE，2020）。73. Barnum, G., Talukder, S. & Yue, Y. 论早期融合在多模态表示学习中的优势。arXiv 预印本 arXiv:2011.07191 (2020)。

74. Vaswani, A. 等. 注意力即一切. Adv. neural information processing systems 30 (2017). 75. Nagrani, A. 等. 多模态融合的注意力瓶颈 . Adv. neural information processing systems 34, 14200– 14213 (2021). 76. Edelman, G. M. & Gally, J. A. 生物系统中的退化性与复杂性. Proc. national academy sciences 98, 13763–13768 (2001). 77. Marder, E. & Taylor, A. L. 用多个模型捕捉生物神经元和网络的变异性. Nat. neuroscience 14, 133–138 (2011). 78. Oess, T., Löhr, M. P., Schmid, D., Ernst, M. O. & Neumann, H. 从近乎最优的贝叶斯整合到神经形态硬件：多感官整合的神经网络模型. Front. Neurorobotics 14, 29 (2020). 79. Blohm, G. 等. 生成式对抗协作：会议组织者和参与的科学家的实用指南. arXiv preprint arXiv:2402.12604 (2024). 80. Peters, B. 等. 生成式对抗协作：科学论述的新模型. Trends Cogn. Sci. 29, 1–4 (2025). 81. Marr, D. 视觉：对人类视觉信息表征与处理的计算研究 (W. H. Freeman, 旧金山, CA, 1982). 82. Lengyel, M. 交锋反驳：Marr的三个分析层次并非神经科学的有效框架. The J. Physiol. 602 (2024). 83. Pillow, J. W. 交锋反方观点：Marr的三个分析层次并非神经科学的有效框架. The J. Physiol. 602, 1915–1917 (2024). 84. Glaser, J. I. 等. 神经解码的机器学习. eNeuro 7 (2020). 85. Ji-An, L., Benna, M. K. & Mattar, M. G. 利用小型循环神经网络发现认知策略. 《自然》644, 993–1001 (2025年). 86. Richards, B. A. 等. 神经科学的深度学习框架. Nat. neuroscience 22, 1761–1770 (2019). 87. Doerig, A. 等. 神经联结主义研究纲领. Nat. Rev. Neurosci. 24, 431–450 (2023). 88. Patterson, A., Neumann, S., White, M. & White, A. 强化学习中的实证设计. arXiv preprint arXiv:2304.01315 (2023). 89. Crilly, A. 等. 在科学领域驾驭人工智能的十条简单规则. PLoS computational biology 21, e1013259 (2025). 90. Eckstein, M. K., Summerfield, C., Daw, N. D. & Miller, K. J. 神经—认知混合模型揭示记忆如何塑造人类奖励学习. Nat. Hum. Behav. 1–16 (2026). 91. Romera-Paredes, B. 等. 利用大语言模型进行程序搜索所获的数学发现. 《自然》625, 468–475 (2024年). 92. Castro, P. S. 等. 从人类与动物行为中发现符号认知模型. bioRxiv 2025–02 (2025).

致谢

MG 获得 Schmidt Sciences LLC 和帝国理工学院研究奖学金的支持。