

• 研究前沿(Regular Articles) •

化繁为简：视觉集合感知的神经机制*

孙焕翔^{1#} 张帆^{1,2#} 李思嘉¹ 张秀玲¹ 蒋毅^{3,4}

(¹ 东北师范大学心理学院, 长春 130024) (² 晋中信息学院, 山西 晋中 030800)

(³ 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)

(⁴ 中国科学院大学心理学系, 北京 100049)

摘要 集合感知是视觉系统高效地从复杂的外部世界中提取均值、方差等概要信息的过程, 这对于人类适应环境具有重要意义。对其神经机制的研究有助于理解视觉系统如何实现高效的抽象表征。本文总结了集合感知的时间进程, 综述了这种整合机制的理论模型和实证证据, 并区分了集合编码与成员或个体编码的功能及神经基础。在现有研究成果的基础上, 提出了“粗略-细节-校准”的整合模型: 大脑在加工不同水平的视觉特征时, 可能依次存在领域通用与特异性机制, 早期依赖于通用性的大细胞通路的粗略加工, 随后是特异性的、依赖于各特征脑区小细胞通路的相对精细表征, 最后通过前馈-反馈循环迭代进行校准。未来研究可关注视觉集合感知的神经通路与具体脑区、前馈与反馈的角色、信息编码的通用性与特异性, 以及发育与经验对集合感知的影响。

关键词 集合感知, 统计概要表征, 知觉整合, 时间进程, 神经机制

分类号 B842

1 引言

想象一下, 你走进厨房, 目光不经意间扫过桌上的一盘苹果, 即使没有去刻意分辨每个苹果的细节, 你也能瞬间捕捉到这盘苹果的大小和成熟度。这种“一目了然”的能力绝非偶然, 而是视觉系统的一种高效加工方式——集合感知(Ensemble Perception)。当面对一组相似刺激时, 视觉系统能够迅速且准确地提取集合的整体统计属性(如平均值、方差等), 而不是处理单个刺激属性的细节(Ariely, 2001; Whitney & Yamanashi Leib, 2018)。这种机制充分展现了人类视觉系统在处理大量信息时的高效性。

集合感知又称集合编码(Ensemble Coding), 这一领域自 Ariely (2001)的开创性工作后得以迅速发展, 早期研究者通常使用“统计概要表征

(Statistical Summary Representation)”术语来描述这一现象(如: 全可等, 2015)。在视科学研究中, “集合感知”和“统计概要表征”均指对大量刺激的集合进行统计加工、整合的认知过程, 在许多研究报告中常被互换使用(Alvarez, 2011; Corbett et al., 2023)。但也有研究者指出, 两个概念间存在细微差异: 集合感知更强调研究对象是大量、集群的刺激, 而统计概要表征则更突出表征过程中的统计计算过程, 前者可包括那些不适用于统计指标描述的复杂属性, 其概念外延更为宽泛, 因此可视为一个更广义的上位概念(励奇添, 陈文锋, 2022)。值得注意的是, 对于“Ensemble Perception”, 国内学界尚未形成统一的译法。部分研究者将其译为“集群感知”(如: 励奇添, 陈文锋, 2022)、“集合感知”(如: 刘丁睿, 2021)、“整体感知”(如: 郝爽等, 2023), 或“整体知觉”(如: 赵冰洁等, 2024)。但是, 为避免与计算机科学与产业经济学领域的“集群”(cluster)产生混淆, 以及与认知科学中另外的重要概念“整体加工”(holistic processing)进行区分, 本文建议使用“集合感知”这一译法。此外, 如果未来的研

收稿日期: 2025-07-01

* 吉林省教育厅科学研究项目(JJKH20241389KJ)资助。

孙焕翔和张帆为本文的共同第一作者

通信作者: 张秀玲, E-mail: zhangxl556@nenu.edu.cn

究报告中仅涉及刺激集合的“均值”或“方差”，可直接使用“平均表征”(average representation)和“方差表征”(variance representation)术语。

集中趋势(如平均值)和离散趋势(如方差)是集合感知的两个核心指标(全可等, 2015; Norman et al., 2015)。平均表征是描述刺激集合集中趋势的重要代表。Ariely (2001)通过不同大小圆点集合的均值识别任务,发现个体可以快速而准确地表征集合均值,且这一表征不受集合大小(即个体数目)的影响,但却难以准确表征集合内的单个成员。随后的研究发现,平均表征广泛存在于从低水平到高水平的多种视觉特征中,适用于如刺激大小(Haberman & Suresh, 2021)、空间位置(Sun et al., 2021)、深度(Wardle et al., 2012)、朝向(Parkes et al., 2001)、颜色和对比度(Rajendran et al., 2021)、亮度(Bauer, 2009)等低水平特征,以及面孔情绪(Haberman et al., 2009)、面孔性别(Haberman & Whitney, 2007)、面孔身份(de Fockert & Wolfenstein, 2009; Davis et al., 2021)、目光朝向(Florey et al., 2016)、生物运动(Sweeny et al., 2013; Nguyen et al., 2021)和可信度(Marini et al., 2023; Chwe & Freeman, 2024)等高水平特征。

方差表征则是描述集合离散趋势的重要指标。Haberman, Lee 和 Whitney (2015)首次从方差角度探讨了面孔情绪的集合感知。研究使用调整匹配法,要求被试调整一组情绪不同的面孔集合的方差以匹配另一组集合,结果发现被试能够准确表征集合的方差,且不受集合大小的影响。后续研究进一步表明,与平均表征类似,视觉系统也可以表征多种视觉特征的方差,包括颜色(Maule & Franklin, 2020; Ward et al., 2016)、亮度(Tong et al., 2015)、大小(Tokita et al., 2016)以及面孔性别和种族(Phillips et al., 2018)。这些研究说明,视觉系统在处理集合离散趋势方面同样表现出高效性。然而,目前针对集合离散趋势的研究仍较为有限,其是否具有与平均表征类似的跨水平、跨特征的知觉普遍性尚不明确。

总之,视觉集合感知是一种应对繁杂视觉输入的高效加工方式,对其神经机制的探讨有利于我们理解视觉系统对均值等统计信息的神经计算过程。本文深入探讨了其加工早期和晚期的时间进程、信息整合的复杂机制(如表征层级以及计算机制),辨析了集合整体与单个刺激信息编码的神经基础差异

和理论模型,并在此基础上提出了视觉统计加工的整合假设。最后,提出未来研究方向,希望能为进一步揭示集合感知的神经机制提供思路和启发。

2 集合感知的时间进程

2.1 时间进程上早期的自动化加工证据

在注意需求方面,近年来的事件相关电位(Event-Related Potential, ERP)研究考察了对面孔情绪的平均表征是否依赖注意资源。Ji 等人(2018)采用线索-目标范式,要求被试在有效或无效线索条件下判断面孔集合的平均情绪是积极还是消极的。结果显示,即使在注意受限的无效线索条件下,被试仍可以有效提取平均情绪,且未观察到反映空间选择性注意的 N2pc (N2-posterior-contralateral)成分,表明平均情绪的提取对注意资源的依赖性较低。值得注意的是,在有效线索条件下,集合任务与个体任务诱发的对侧延迟负波(Sustained Posterior Contralateral Negativity, SPCN)成分(其波幅通常反映视觉短时记忆负荷)无显著差异,意味着多个面孔信息被压缩成一个“单一”对象存储于视觉短时记忆中(Ji et al., 2018)。

无效线索条件下,虽然注意资源减少,但也存在注意泄露(Attentional Leakage)的可能。在此基础上, Ji 等人(2024)进一步探究了在更严格的注意资源操控条件下,平均情绪信息是否仍能被自动提取。该研究采用 oddball 范式,要求被试完成一项与外周呈现的面孔情绪集合无关的中央注视任务。结果发现,尽管注意完全集中于中央注视任务,被试仍能自动检测到外周面孔平均情绪的变化。这种自动检测表现出情绪效价的显著差异,平均情绪为负性的偏差刺激约 92 ms 便诱发了快速且持续的视觉失匹配负波(Visual Mismatch Negativity, vMMN);而平均情绪为正性的 oddball 刺激则在稍晚的 168 ms 诱发失匹配正波(mismatch positivity, MP);且多变量模式分析(Multivariate Pattern Analysis, MVPA)结果显示,大脑约在 90 ms 便可区分标准刺激与 oddball 刺激。这些发现表明,平均情绪提取过程可以在无需注意参与的情况下自动进行,并且正、负性平均情绪的加工可能涉及不同的神经机制。

此外,在加工速度和时间进程方面,另一项采用 oddball 范式的 ERP 研究探索了由多个线段组成的集合感知的时间进程。研究者区分了基于

集合中某个成员的 oddball 刺激与基于整个集合的 oddball 刺激, 发现与成员条件相比, 集合 oddball 刺激诱发的 P3b 成分潜伏期显著更早。此外, MVPA 结果显示, 与成员条件相比, 神经信号能够更早地分类集合条件下的标准刺激与 oddball 刺激(分类开始于刺激呈现后约 102 ms)。这一结果表明, 集合感知发生迅速, 且对集合的表征早于对个体成员属性的表征(Epstein & Emmanouil, 2021)。

然而, 近期一项平均朝向的研究对集合感知的完全自动化机制提出了挑战。Lukashevich 等人(2025)同样采用 oddball 范式, 设计了两个实验来检验注意在平均朝向识别中的必要性。实验一为注意条件, 要求被试明确注意并报告集合平均朝向的变化, 研究观察到了清晰的 P3 成分, 这是意识层面变化检测和注意参与的典型标志。实验二为非注意条件, 要求被试将注意完全集中于中央注视点的变化、忽略背景集合刺激, 即使集合的平均朝向发生了与实验一中相同的变化, 研究也未能探测到 vMMN 成分。多变量模式分析(Multivariate Pattern Analysis, MVPA)的结果与 ERP 结果一致, 无法在未注意条件下成功解码平均朝向的变化。作者认为, 至少对于中等程度差异的平均朝向而言, 其识别与变化检测并非完全自动化, 而是需要注意资源的投入。这一发现与 Ji 等人(2024)关于平均情绪自动加工的结果形成了有趣的对比, 提示不同类型的统计概要信息如面孔情绪和线条朝向可能依赖不同的认知机制: 因为面孔情绪更具有进化的适应性, 可能与杏仁核等快速情绪处理通路有关, 更容易在非注意条件下形成平均表征; 另外, 自动加工的程度可能受差异突显性的影响, 如负性情绪与正性情绪的差异可能大于 40 度与 60 度朝向的差异。

此外, Im 等人(2021)及 Roberts 等人(2019)的研究虽然主要探讨集合与个体编码神经基础的分离, 但也对集合感知的时间进程进行了一定分析, 其研究结果中亦涉及了集合感知不同阶段的神经活动特点(详见本文 4.1 节)。需要强调的是, 关于集合感知自动化加工与注意的关系, 前人基于行为研究的综述已有较为丰富的讨论(可参阅: 全可等, 2015; Alvarez, 2011; Whitney & Yamanashi Leib, 2018), 故本文不再赘述。

2.2 时间进程上晚期的精确性加工证据

近年来, 也有一些研究使用神经生理技术探

讨了集合感知晚期的时间进程, 揭示其构建可能是一个逐步精细加工的动态过程。

Yashiro 等人(2024)结合逆向编码模型(inverted encoding model, IEM)和脑电(Electroencephalography, EEG)数据, 重构了视觉系统在不同时间点对平均朝向的神经表征, 并将其与被试的行为反应(所报告的平均朝向感知)进行了关联。他们发现, 尽管同质条件朝向的神经表征在刺激呈现后相对较早的时间窗口内(174~828 ms)内便能够得到显著解码, 但异质条件平均朝向的神经表征强度却是在一个较晚的时间段(400~700 ms)内才逐渐得到增强, 且其强度仅在更晚期(600~700 ms)与被试的行为表现显著相关。由此, 研究者推测集合感知可能结合了并行处理与序列处理, 集合信息在短时间内通过快速的前馈处理被粗略提取, 而更精确的表征则需要迭代循环的前馈-反馈机制来进行细化(Yashiro et al., 2024)。

Gong 等人(2025)最近的一项脑磁图(Magnetoencephalography, MEG)研究结合 MVPA 和 IEM 技术, 揭示了类似的时间动态。研究考察了集合朝向(无任何成员朝向与平均朝向相同)的神经表征, 发现尽管集合朝向信息在刺激呈现后约 112 ms 即可被解码, 但只有出现在 600 ms 的解码峰值和行为正确率显著相关, IEM 分析显示, 真正表征集合的平均朝向的神经活动则在更晚的时间点(约 370 ms 后)才出现, 而且表现出与成员信息简单相加不同的独特神经表征模式。

两项研究共同提示, 尽管对集合信息的初步提取可能较快, 但形成精确、区别于个体项目相加的集合感知, 则需要更长的处理时间, 可能通过递归、迭代甚至更复杂的加工过程以精细化表征(Yashiro et al., 2024; Gong et al., 2025)。

3 集合感知的整合机制

视觉系统如何从繁杂的个体刺激信息中整合出集合感知? 这是一个核心的科学问题。早期的理论模型为理解这一过程提供了基础框架, 后续研究则揭示了更为复杂的神经计算机制, 超越了简单的线性平均。

3.1 层级汇集假设(Signal Pooling Hypothesis): 描述性的理论假设

Whitney 带领他的团队在集合感知领域进行了大量的研究, 做出了突出贡献。他们基于该领

域大量的行为研究证据,提出了关于大脑进行集合感知的神经通路和表征过程的理论设想,我们称其为层级汇集假设(Signal Pooling) (Herman & Whitney, 2012; Whitney et al., 2014)。作为早期的重要理论设想,层级汇集为视觉系统如何高效处理大量信息提供了一种计算原理层面的解释。然而,该假设也存在一定的问题,如集合感知的具体整合机制、注意和意识对其的影响,以及跨领域刺激的加工机制等方面仍需要进一步的揭示。主要内容如下:

(1) 集合感知的神经机制可能是通过线性整合或群体编码而实现神经信号的汇集。例如,对于一组由不同朝向组成的线段集合,对各个朝向敏感的神经元在初级视觉皮层(Primary Visual Cortex, V1)被激活,神经信号沿视觉层级自下而上逐层汇集,最终会形成对集合整体的感知。这一设想以视觉系统层级表征为基础,依赖神经元感受野逐层增大的特点(Rousselet et al., 2004),并指出汇集能通过平均化有效降低噪声,提高表征精度;(2) 集合感知依赖于内隐的、平行的加工机制。其表征过程不受集合中项目数量的影响(Chong & Treisman, 2003),也可以在有限注意资源下进行(Wolfe et al., 2015)。(3) 不同集合感知发生于视觉系统传导的不同水平层级、不同传导通路上。如平均亮度、颜色和朝向等低水平特征的汇集可能发生在 V1,甚至是皮层下区域;而形状和面孔等高水平特征可能定位于腹侧通路,平均运动和位置则可能与背侧通路相关。

近年来的神经影像学研究了沿视觉加工通路逐层汇集的观点。例如,研究者采用功能性磁共振成像(functional Magnetic Resonance Imaging, fMRI)和 IEM 研究光栅集合的平均朝向表征,发现尽管集合任务与个体任务的整体血氧水平依赖(Blood-Oxygen-Level-Dependent, BOLD)信号并无显著差异,但通过 IEM 所得到的神经反应却显示出差异。在任务相关条件下,枕叶和额顶叶都存在对平均朝向(集合任务)以及个体朝向(个体任务)的选择性反应。值得注意的是,平均朝向的选择性反应在 V1 并不显著,却在 V2 和 V3 中表现显著,且从 V1 到 V2 再到 V3 呈现显著的线性增长。该研究结果表明,视觉系统在多个层次上对信息进行汇集,从而形成了对统计概要信息的神经表征,支持了逐层汇集的观点(Tark et al., 2021)。

然而,也有证据发现,层级汇集过程并非低层级的信号在高层级的简单线性平均,而是一种高阶(Allik et al., 2022)或高维(Jia et al., 2022)的整合。另需注意的是,当我们对高水平视觉刺激(如面孔)形成平均表征时,大脑需要存在接收多个面孔信息输入的高级神经元。这意味着大脑不仅需要大量表征单个面孔的神经元,还需数目更加庞大的“统计概要神经元”。Whitney 团队后来也承认,简单的层级汇集模型难以解释客体水平的集合感知(Whitney & Yamanashi Leib, 2018)。因此,当涉及整体和个体关系时(尤其在解释拥挤效应时),稀疏选择模型(Chaney et al., 2014)和逆层级理论(Hochstein et al., 2015)可能更适用一些。这时,统计信息的加工可能更多地依赖于神经元群体编码(population coding),即多个表征不同面孔的神经元被同时激活,从而在群体水平上形成了对集合整体的表征。

此外,层级汇集假设强调,集合感知是一个自动化的过程,它依赖于平行加工,且不受注意资源和集合中项目数量的限制。尽管自动化加工存在一定的支持证据。例如,拥挤效应(crowding)的研究发现,即使被试对个别成员的注意和意识受到了影响,但是集合感知表现仍良好(Parkes et al., 2001; Fischer & Whitney, 2011),以及对单侧空间忽视综合症(unilateral spatial neglect, USN)患者的研究发现,即使因右侧顶叶损伤而导致左侧空间忽视,患者对侧视觉信息的平均表征能力仍相对保留(Yamanashi Leib et al., 2012; Pavlovskaya et al., 2015)。但这一观点仍有争议。一些研究指出,集合感知的精度可能随项目数量增加而下降,可能存在容量限制或早期编码噪声的影响(Marchant et al., 2013; Utochkin & Tiurina, 2014),且注意可能通过提高个体项目权重的方式调节集合感知(Ying, 2022)。

最后,层级汇集假设认为,在不同视觉加工水平的刺激间存在多重的机制而非单一机制。但是,目前有关这个问题的实证结果尚不一致(Chang, Cha & Gauthier, 2024; Herman, Brady & Alvarez, 2015),这仍是一个活跃的研究领域(详见本文 6.3 节)。

3.2 整合的计算模型

尽管层级汇集或传统的线性平均假设为理解集合感知提供了一个基础框架,但其仅是描述层

面的理论概述,为了更好地解释整合的机制,计算建模领域的进展提供了更具体的神经计算路径和假设。

Utochkin 等人(2024)提出的群体响应模型(population response model)可以理解为层级汇集假设的模型版本,认为集合感知是大脑中特定特征选择性神经元群体响应经过汇集的结果,通过一个两层的神经网络模型即可模拟这一过程。第一层为感受野比较小的特征层,编码个体刺激的低层级神经元的信号,经过加权的突触连接以前馈的方式传递给第二层;第二层为感受野比较大的汇集层(池化层),对传入信号进行整合,形成一个反映刺激集合统计特性(如均值、方差、分布形态)的整体群体响应调谐曲线(tuning curves)。这种基于群体编码的汇集机制本身就能通过前馈机制在神经层面实现对统计信息的有效表征,而无需进行“求和再除以数量”的算术运算。这一模型得到了 Iakovlev 和 Utochkin (2023)研究的进一步支持。他们采用偏态分布(skewed distributions)的朝向集合进行研究,发现被试估计的平均朝向会系统性地偏向分布的众数(mode),且偏离程度不仅取决于均值与众数的距离,还受到整个分布形状的影响。重要的是,在群体响应模型中,神经信号在特征空间中汇集,使得响应峰值自然地倾向于分布中项目更密集的区域(即众数附近),其偏离程度则受到神经元调谐曲线宽度的影响。这为降低离群值权重的“稳健平均”(robust averaging)提供了一种神经上似乎合理的解释,也表明群体响应模型可以作为理解包括凸显性加权和注意调节在内的非均匀整合现象的神经基础(Iakovlev & Utochkin, 2023; Utochkin et al., 2024)。

此外,Robinson 和 Brady (2023)提出的感知加总模型(Perceptual Summation model),则基于记忆的目标混淆竞争理论(Target Confusability Competition, TCC)框架来揭示集合表征的机制,并得到实验数据的验证。他们认为,集合感知反映的是对各个项目特征诱发的分布式激活模式的直接求和,其表征基于概率而非点估计。该模型只有一个从个体项目记忆任务中估计的信噪比参数(d'),而无自由参数。加总过程被认为发生在早期编码阶段而非记忆阶段,依赖于个体表征而非独立于个体表征的自动过程(Robinson & Brady,

2023)。

$$r_{ENS} = \operatorname{argmax} \left(\sum_{i=1}^N f(x)_i d' + \sigma_{noise} \right).$$

注: r_{ENS} 代表模型对集合感知任务(Ensemble Task, 简称 ENS)中报告的特征值的预测结果,即模型预测的被试所选择的那个代表集合平均特征的值。 d' (d prime)为信噪比参数,量化了每个单独记忆表征的强度或清晰度,并根据记忆任务的需求(如记忆负荷、编码时间、记忆间隔等因素)来调整 $f(x)_i$ 所代表的激活模式的强度。(引自 Robinson & Brady (2023). *Nature Human Behaviour*, pp. 1638–1651)

计算模型为集合感知提供了一个潜在的神经计算框架,群体响应模型通过神经网络模拟了统计信息(包括对偏态分布的稳健估计)如何从个体信息中高效生成;感知汇总模型则强调了早期知觉响应的简单线性汇总。

3.3 非简单线性平均的证据

越来越多的证据表明,大脑整合个体信息形成集合感知的过程远比预期复杂。有研究对前述 Whitney 的汇集假设所持有的线性汇集机制以及无自由参数的线性汇总模型提出了挑战,认为集合项目的整合可能涉及更复杂的非加性整合机制(nonadditive integration)或非均匀加权(non-uniform weighting) (Choi & Chong, 2020; Jia et al., 2022; Wang et al., 2023; Gong et al., 2025; Tiurina et al., 2024; Kanaya et al., 2018)。

一方面,整合过程可能并非简单的基于成员表征的线性加和。为探究大脑中是否存在处理集合感知的专门神经机制, Jia 等人(2022)采用频率标记脑电(Steady-State Visual Evoked Potential, SSVEP)方法追踪大脑对以一定频率周期性变化的圆环集合平均大小的集合表征。他们在顶枕电极上观察到对整体平均大小的神经反应,并通过时间响应函数(Time Response Function, TRF)分离了个体大小以及个体间的相互作用(包括全局交互和局部交互)的神经反应。结果显示,只有全局交互直接贡献了整体的大小感知,表明大脑中存在专门表征集合整体大小的特定神经机制,且这种机制对于个体信息的整合过程不是简单线性平均的,而是表征了个体之间的相对关系,表现为一种高维的整合(Jia et al., 2022)。其研究团队的另一项研究使用基于回归框架的事件相关电位(regression-based Event-Related Potentials, rERPs)方法,进一步提出集合感知依赖于非加性整合机

制(Wang et al., 2023)。他们从 EEG 信号中分离了三个信息层级的神经响应(个体项目、局部交互、全局交互),结果发现,只有全局交互的神经表征与行为层面的感知精准度显著相关,且分布式注意可以增强全局交互的早期神经响应。因此,线性平均不足以捕捉刺激间的交互作用对整体的贡献,无法准确地反映刺激集合内部的复杂关系(Jia et al., 2022; Wang et al., 2023)。Gong 等人(2025)的研究也为这种超越简单线性叠加的整合机制提供了直接的神经证据。他们使用逆向编码模型(IEM)分析 MEG 数据发现,对于异质朝向集合,基于 IEM 重构出的大脑平均朝向表征与基于对各个成员朝向反应进行线性相加预测出的表征存在显著差异。具体来说,实际表征在平均朝向处表现出更高的保真度(fidelity)和响应强度。此外,基于源重构的神经活动通过贝叶斯概率解码(Bayesian probabilistic decoding)分析,也证实了异质集合诱发的神经活动模式与模拟简单相加的模式不同,前者在平均朝向处具有更高的朝向响应强度。这些发现有力地表明,大脑在形成集合表征时,并不仅仅是将个体信息进行简单的线性求和,而是涉及了额外的、更复杂的整合过程(Gong et al., 2025)。

另一方面,整合过程中的加权可能并非均匀的(non-uniform weighting),注意、项目凸显性和刺激在视野中的空间位置在其中扮演了重要角色。例如,更凸显(如更大、更高频)的项目可能被赋予更高权重,导致感知平均值偏向这些项目,产生“放大效应(amplification effect)”(Kanaya et al., 2018)。同时,选择性注意也能主动调节这一加权过程,不仅可以直接增加被注意项目的权重,还可能通过“知觉放大(perceptual enlargement)”(即增加被注意项目的权重)来影响最终的平均值估计,这种效应在注意被提前引导至特定项目时尤为明显(Choi & Chong, 2020)。此外,这种加权过程在视野空间上也并非均匀进行,Tiurina 等人(2024)发现,在估计集合平均朝向时,被试表现出明显的空间各向异性(spatial anisotropies),具体表现为对呈现在中央凹(fovea)位置刺激的强烈偏向,并且这种偏向的程度随着刺激不确定性的增加(如刺激朝向为倾斜或集合方差增大时)而增大。

人类的认知过程往往比模型复杂,行为和神

经科学的证据揭示出非加性整合、基于凸显性和注意的加权,以及刺激的空间位置都在最终形成集合感知的过程中扮演着重要角色。这些实证研究揭示了集合感知过程中整合机制的复杂性,简单的线性平均或均匀加权假设面临挑战。未来的研究需要进一步整合这些不同取向的发现(如整合机制的计算方式、时间阶段、注意调节、空间权重等),以构建更完善的集合感知神经整合模型。

4 特异于单个刺激编码的神经基础

集合编码和单一项目的编码是否有本质不同,似乎是集合感知领域最著名的争论(Ariely, 2001; Corbett et al., 2023)。对集合整体的处理和对单个刺激的处理在神经基础和加工特性上是否存在差异?大脑如何区分并处理集合的整体属性与集合中的成员信息?本节将探讨集合编码与个体编码以及集合成员编码的神经机制可能分离的证据,包括其潜在机制的理论框架、加工的优先顺序、涉及的神经通路、对不同信息的敏感性以及在工作记忆中的表征。

4.1 集合和个体编码的神经基础

已有研究表明,大脑中存在对集合属性进行集合表征的特定神经机制(Roberts et al., 2019; Jia et al., 2022)。近年来,研究者通过多种神经影像技术深入探讨大脑如何处理集合刺激与单一刺激的信息,分析相关脑区的激活模式及其功能分工。例如,集合编码和个体编码在背侧通路和腹侧通路之间存在功能分离(Im et al., 2017; Im et al., 2021);刺激的空间分布(Sama et al., 2024)、不同的空间频率(Zhao, Shen et al., 2023),也分别对集合信息与个体刺激产生不同影响。此外,还有研究关注工作记忆存储的集合表征涉及的脑区(如前额叶和枕顶叶)及其模式差异(Oh et al., 2019)。

首先,Roberts 等人(2019)使用 EEG 探索面孔身份集合感知的神经基础,探究对面孔集合信息的整体加工是否存在特定神经机制。实验中给被试呈现集合面孔或单个面孔,集合条件的 P1 波幅更小, N1 和 N2 的潜伏期更短。使用线性支持向量机(Support Vector Machine, SVM)的多变量模式分析进行神经解码,发现神经信号不但可以区分不同的单张面孔,也可以区分平均身份不同的集合面孔,有趣的是,神经信号难以区分平均身份

相同而成员不同的两个集合(结果达到边缘显著水平)。集合表征解码的时间进程不同于单个面孔表征,体现出知觉信息随时间逐渐累积的过程。通过神经信号的图像重建可以准确地可视化集合表征的内容。在这一研究团队的另一篇文章中(实际上使用的面孔集合数据来自 Roberts 等(2019)的研究),基于线性支持向量机和递归特征消除进行多变量特征选择,发现神经信号(在时域上和频域上)可以区分不同的面孔集合,相较于单个面孔和词汇,面孔集合不仅仅依赖于颞枕区的电极通道,也依赖于中央区的电极通道(Nemrodov et al., 2020)。

此外,为考察面孔集合情绪编码和单一面孔情绪编码的感知差异,研究者采用 fMRI 技术探索了两类激活脑区的分离。发现位于背侧通路(dorsal stream)的顶内沟和额上回参与面孔集合整体情绪感知,而位于腹侧通路(ventral stream)的梭状皮层参与个体面孔的情绪感知,同时这项研究中还发现了面孔集合的情绪编码具有大脑右半球偏侧化优势(Im et al., 2017)。2021 年该团队使用 MEG 技术重复了之前的发现,即背侧通路参与了集合面孔情绪的整体加工,腹侧通路参与对单个面孔情绪进行识别与分辨。重要的是,MEG 揭示出背侧通路可以对集合面孔情绪进行非常快速的反应(刺激呈现后 68 ms),可能依赖大细胞通路快速的信息输入来形成对集合面孔情绪的集合感知(Im et al., 2021)。

Zhao, Shen 等人(2023)的研究进一步探讨了不同空间频率对面孔集合情绪感知的影响,经过不同的滤波生成高频(HSF)、低频(LSF)和宽频(BSF)的面孔集合刺激,测试被试对面孔集合的快乐、中性和恐惧情绪的识别,发现低频条件可以促进集合恐惧面孔的识别。fMRI 结果揭示了空间频率在群体情绪感知中的差异性作用:高频信息主要激活腹侧视觉通路(如梭状回、双侧枕中回和枕下回),处理细节信息但难以支持整体情绪识别;宽频信息条件下则观察到右侧额下回被激活,表明整合多频信息需要更多的认知资源。

为进一步探讨集合内部信息处理的分工, Sama 等人(2024)采用了一种新颖的中心-周围范式(一个中心面孔被 6 个周围面孔环绕),结合 EEG 和 MVPA 技术,揭示了面孔集合中心面孔

与周围集合信息的神经表征动态时程,发现对中心面孔的神经解码在时间进程上显著早于对周围集合面孔的解码。而且有趣的是,中心面孔的解码时间进程与单独呈现的单个面孔非常相似,而单个面孔和集合面孔的解码时间模式存在广泛的差别,这提示大脑可能以相对独立的神经机制来处理注意焦点下的个体信息(即使其处于集合中)和更广泛的集合整体信息。这一发现为集合编码和个体编码以及集合周围信息和中心焦点信息在神经时间动态上的分离提供了直接证据,并强调了注意在区分这两种加工过程中的潜在作用(Sama et al., 2024)。此外,该研究通过独立操控面孔的形状和表面属性,发现两个属性在集合表征中皆扮演重要角色,并且后者的贡献更大一些,这可以体现在较高的分类正确率上。

最后,也有研究者考察工作记忆表征是否像视觉系统的层级表征一样是结构化的,高级脑区表征统计概要信息,低水平脑区表征繁杂的感觉信息。使用线条朝向的集合刺激,采用逆向编码模型(IEM)对 EEG 信号进行解码,结果支持结构化表征假设。具体为在额中央区不仅存在对简单的特征的稳定编码(stable coding, 可时间泛化),也存在对集合平均的稳定编码,重要的是额中央区活动甚至和行为的集合判断存在显著相关;在顶枕区存在对特征的动态编码(dynamic coding, 不可时间泛化),只有在目标调整任务中存在对集合平均的稳定编码,新旧判断中不表征平均信息,即受到任务要求的调节。总之,在视觉层级上越高级的脑区工作记忆表征的内容越抽象(Oh et al., 2019)。

4.2 逆层级理论(Reverse Hierarchy Theory)

Hochstein 及其同事提出的逆层级理论(Reverse Hierarchy Theory, RHT),一定程度上(Hochstein & Ahissar, 2002; Hochstein et al., 2015)为集合感知的内部机制提供了基本的理论框架。有观点认为, RHT 作为一个广泛且全面的理论框架,解释了快速加工集合信息(如“先看到森林而非树木”)的现象,是当前解释集合感知最有效的模型(Corbett et al., 2023)。

与传统的“自下而上”模型形成对比, RHT 强调了视觉感知首先把握集合的“要旨”(gist)。其核

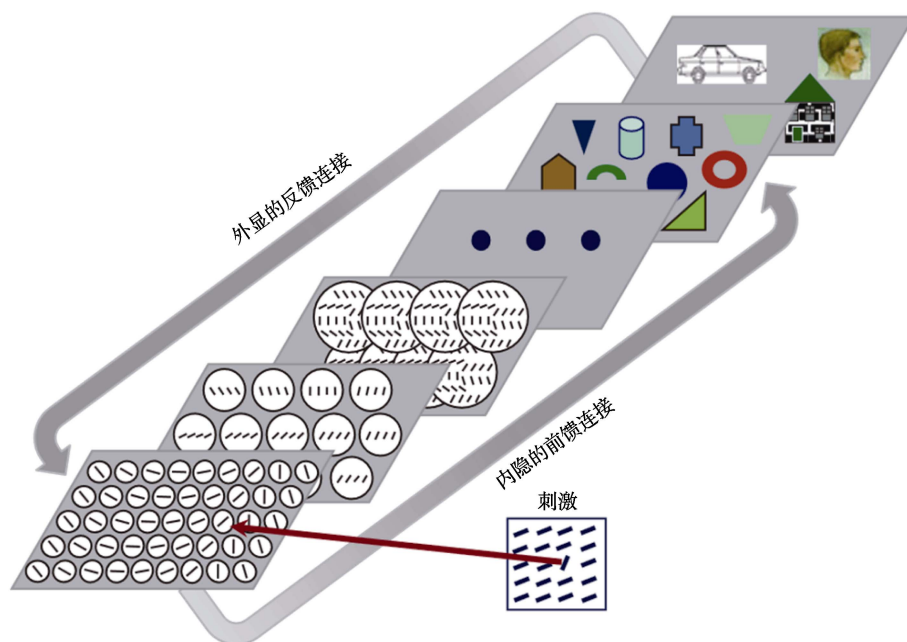


图 1 传统层级结构和逆层级理论(RHT)。

注：传统观点认为，视觉输入首先由低层级视觉皮层区域的神经元接收和处理，对刺激的简单几何形状做出反应。随后，信息自下而上逐层传递，表征全局特征。最终，信息在更高层级的皮层区域得到整合，形成抽象形状、物体和类别的表征，这一过程并不涉及反馈。逆层级理论则提出，前馈路径仅内隐地加工特征信息，有意识的视觉整体感知在高层级皮层形成，表现为对场景要点的概要表征。随后，意识感知通过反馈连接，自上而下返回至相应的低层级皮层，以形成更详细的表征。(资料来源 Hochstein & Ahissar, *Neuron* (pp. 791–804) (2002), 已获得改编许可)

心观点是：视觉系统的处理遵循沿视觉皮层的层级“快速前馈与细节反馈”的模式(见图 1)。视觉信息的初始处理是内隐(implicit)、前馈(feedforward)的，表征局部特征的信息自下而上传递，并在较高层级的脑区形成对集合整体的概要表征(即“vision with a glance”)；随后，有意识的感知遵循一个逆层级路径，当需要加工局部细节信息时，意识感知通过反馈(feedback)连接，以自上而下的方式返回到较低层级的脑区，形成个体的成员的精确表征(即“vision with scrutiny”)，局部的加工可以对最初的视觉表征近似值进行修正并补充细节信息(Hochstein & Ahissar, 2002)。Hochstein 等人(2015)基于 RHT 框架，进一步指出这种“全局优先、局部细化”的处理进程与集合感知的快速、整体性加工的特点密切相关。

实证研究也为 RHT 关于集合优先于局部得以表征的假设提供了支持。例如，田欣然等人(2021)通过平均辨别任务和吸引力评价任务也间接揭示了这种整体优先的加工特性。研究发现，被试对平均吸引力的感知依赖于平均的面孔表征，

当面对大容量集合(如 12 张面孔)时，合成平均表征的主导作用更为显著，而当集合较小(如 4 张面孔)时，更多细节加工资源被允许，个体面孔信息的影响力增强，合成平均表征的主导性则相应减弱。近年来的 ERP 研究同样得出支持性结论。Epstein 和 Emmanouil (2021)使用 oddball 范式的研究发现，相比于成员 oddball (由集合中特定线段的朝向定义)，被试对集合 oddball (由集合平均朝向定义)的反应更快，集合 oddball 条件的脑电神经反应的潜伏期更早，且在解码中表现出更早、正确率更高，泛化模式更强的特点。Liu 等人(2023)通过操纵集合面孔的呈现时间以及项目数量，进一步揭示了这种加工优先级的时间依赖性：在较短时间(如 50 ms)下，被试对平均面孔情绪的判断优于对单个面孔情绪，而在较长时间(如 750 ms)下则相反；此外，ERP 结果显示，反映项目个体化的 N2pc 成分在短时条件下不受集合内面孔数量影响，但在长时条件下则随面孔数量的增加而增强。

因此，RHT 为理解大脑处理集合信息与成员

信息时可能存在的不同加工机制与优先级提供了一个有解释力的理论框架。集合信息与成员信息在神经机制和时间进程的分离, 不仅解释了行为层面观察到的整体优先效应, 也暗示集合编码和个体编码在神经基础上的差异。但我们仍需认识到, 集合感知的复杂机制可能并非单一理论所能完全涵盖。将 RHT 视为解释集合感知的有效视角的同时, 也应结合其他理论模型和相关的实证研究, 以构建更全面的理解(Corbett et al., 2023)。

5 方差表征神经机制的初探

方差表征有助于人类识别和评估信息的多样性、不确定性和可靠性, 也有助于引导基于统计突显性的注意偏向, 帮助我们快速定位“异常”, 从而灵活地应对复杂的外部世界。深入探究方差表征的神经机制, 将揭示大脑如何编码刺激集合的离散程度或异质性, 有助于阐明视觉系统表征群体多样性与规律性信息的内在机制。然而, 目前有关方差表征的神经科学研究仍较为有限。

视觉方差的表征可能发生于早期视觉通路, 甚至早在皮层下结构中就已经开始。例如, Norman 等人(2015)发现, 方差的适应依赖于视网膜坐标(相对于注视点位置), 而非空间视觉坐标。为进一步探究方差编码的神经基础, 该研究对一位视觉区脑损伤患者进行了测试, 发现与正常被试相比, 该患者在光栅朝向方差的表征正确率上并无显著差异。该患者仅左半球 V1 区未受损, 很多视觉区受损, 如双侧腹内侧枕颞皮质受损, 右半球 V1 受损, 且双侧 V2、V3、V4 部分区域受损, 这表明方差表征可能发生在视觉系统的较早阶段, 如 V1(Norman et al., 2015)。此外, 动物研究也提供了方差表征依赖于早期视觉加工通路的证据, 例如, 猫的外侧膝状体(lateral geniculate nucleus, LGN)能够对亮度的标准差产生特异性反应(Bonin et al., 2006)。亦有行为研究提示, 皮层下结构可能在方差表征中扮演重要角色。研究者通过立体镜分别向被试双眼呈现刺激, 发现相较于呈现于不同眼(双眼分视)条件, 当两个刺激序列先后呈现于同一只眼(单眼呈现)时, 被试对方差的判断准确性更高。这种“同眼优势效应”(same-eye effect)表明, 视觉方差信息在信号到达 V1 之前, 就可能已在皮层下结构中得到了初步处理(Zeng et al., 2024)。

近期的电生理研究则为方差的自动加工机制提供了直接证据。陈子龙和季球妍(2025)的一项 ERP 研究采用反向 oddball 范式, 要求被试将注意集中于中央注视点进行辨别任务, 同时在外周呈现与任务无关的面孔集合, 通过操纵集合内面孔情绪强度的一致性来区分高、低情绪方差(两者互为标准刺激和 oddball 刺激)。结果发现, 当平均情绪为中性时, 仅有作为 oddball 刺激的高情绪方差的面孔集合诱发了 vMMN 成分, 而当平均情绪为愤怒时, 高、低情绪方差的 oddball 刺激分别能诱发晚期的视觉失匹配正波(visual mismatch positivity, vMMP)和 vMMN 成分, 表明大脑能够在非注意条件下自动地加工面孔情绪集合的方差信息。而当平均情绪为高兴时, 则未观察到显著效应, 说明情绪类型这种社会性信息可以调节大脑对方差变化的敏感性。在时间进程上, MVPA 的结果进一步表明, 在所有条件下, 大脑能够在刺激呈现后极早期(约 70~90 ms)开始对不同方差水平的刺激进行区分。

此外, 刺激集合方差的高低不同可能会导致激活脑区的强度差异。一方面, 部分脑区的活动强度与方差水平呈正相关。例如, Michael 等人(2015)的 fMRI 研究发现, 视觉皮层和上顶叶的 BOLD 信号会随着刺激集合的方差增高而增强。类似地, Allard 等人(2021)通过改变光栅集合的朝向差异区分高方差朝向组与低方差朝向组, 让被试判断这些光栅集合的平均朝向相对于竖直朝向是顺时针还是逆时针倾斜。fMRI 结果显示, 高方差朝向的光栅集合在右上额叶和左中额回的脑区激活程度显著高于低方差朝向的集合。该研究中所观察到的高级脑区的激活可能和被试需要完成集合平均朝向的辨别任务有关(Allard et al., 2021)。另一方面, Michael 等人(2015)的研究揭示了相反的活动模式, 即背内侧前额叶皮层(dorsomedial Prefrontal Cortex, dmPFC)和前脑岛(anterior insula)的 BOLD 信号随着方差的增大反而减弱, 在刺激更同质化(低方差)时激活水平更强。综合来看, 不同脑区在方差表征中可能扮演着不同的角色, 脑区的激活模式不仅与任务要求以及刺激的物理突显性有关(Allard et al., 2021; Michael et al., 2015), 也与刺激本身的社会性意义(如情绪)紧密相联(陈子龙, 季球妍, 2025)。

然而, 从早期的动物实验、一位患者的脑损

伤研究,再到后来的行为、脑成像与电生理研究,我们尚无法从仅有的研究中得出关于人类离散趋势表征脑机制的确定结论。未来的研究可采用不同的实验范式与技术手段,为揭示方差的神经基础提供新的思路。例如,将要求被试直接比较或估计集合方差的显性判断(Explicit Judgment)范式与 fMRI 技术相结合,可以识别出哪些脑区的活动强度随刺激方差系统性地变化。此外,将经典的神经适应(Neural Adaptation)范式用于方差表征任务,并结合 fMRI 适应性抑制分析,是一个检验是否存在方差特异性调谐神经元群体的有效手段。同时, MVPA 方法能够通过解码分布式神经活动模式,为我们揭示方差信息的具体表征模式提供更进一步的见解。最后,将高时间分辨率的 EEG/MEG 技术应用于上述范式,则有望揭示方差信息被提取和处理的完整时间进程。值得注意的是,未来的研究在解决这一问题时,需要注意区分不同视觉水平的刺激可能带来的影响,同时也需关注平均表征与方差表征之间的相互作用(见本文 6.3 节“通用性机制还是特异性机制”)。

6 研究展望

本文系统回顾了视觉集合感知的神经机制研究,涵盖了其基本特性、时间进程、整合机制、集合与个体编码的分离等内容。未来研究除了需要关注集合编码的时间进程、反馈信号的意义,更需要关注不同视觉通路功能,以及发育和经验可能的影响。同时,也建议研究者对集合感知和统计概要表征的神经机制进行区分。最后,在现有研究的基础上,我们提出集合感知遵从“粗略-细节-校准”(Coarse-Fine-Refine)的多阶段加工过程,并辨析了该模型与领域内主要理论的异同(见本文 6.6 节)。

6.1 集合感知与反馈信号

不同理论模型对集合感知是否涉及反馈信号持有不同的观点。一方面,前馈主导模型(如信号汇集假设和群体响应模型),主张集合感知基于一个前馈加工的过程。神经信号沿视觉层级由低阶向高阶脑区逐级汇集,最终形成对集合统计属性的稳定表征。反馈的作用则被限定在较为次要的功能上,例如标记或识别集合中的极端值(Whitney et al., 2014; Utochkin et al., 2024)。另一方面,阶段加工模型则认为自上而下的反馈至关

重要。例如,逆层级理论提出,视觉感知遵循“先森林后树木”的整体优先原则,前馈形成对森林的整体意识,反馈形成对成员树木的意识(Hochstein et al., 2015)。同时,预测编码理论(Predictive Coding, PC)也强调反馈的重要作用(Gilbert & Sigman, 2007; He et al., 2012; Shipp, 2016)。根据预测编码,集合感知可能包含反馈过程,即对多个客体组成的集合或群体,我们可能会先形成一个关于集合统计属性的预期,随后预期信息反馈至视觉皮层,与真实的视觉输入进行比较。两者间的差异构成了预测误差,视觉系统通过不断修正预期以最小化该误差。逆层级理论和预测编码理论都可以解释以往研究中观察到的集合整体优先现象,即人们对集合整体的平均表征优于对成员的表征(Chong & Treisman, 2003; Li et al., 2016)。

尽管仍需要一些证实集合感知中反馈重要作用的直接神经证据,但来自知觉整合和特征捆绑等相关领域的研究提供了间接支持。例如, Liu 等人(2017)对感知整体模式的脑活动进行研究,发现对视野中的离散刺激形成整合的过程中,背侧视觉通路的后顶内沟(posterior Intraparietal Sulcus, IPS)会快速激活(100 ms 以内),并通过反馈调节早期视觉皮层(Early Visual Areas, EVAs)的活动。在特征捆绑的任务中,研究者也发现了从 V5 到 V2 的反馈连接(Zhang et al., 2014)。

6.2 不同视觉加工通路在集合感知中的功能

首先,关于集合感知是依赖腹侧通路还是背侧通路,现有的研究证据相对匮乏。Whitney 等人(2014)认为,集合感知并非由单一脑区或单一通路负责。例如,朝向等低水平视觉特征依赖于枕叶,物体运动或位置信息依赖于顶叶背侧通路;而面孔或形状等高水平刺激则更依赖于颞叶腹侧通路等。然而,实证研究虽然部分支持多通路加工的观点,但对不同脑区的作用却有不同揭示,感知面孔情绪集合时更依赖于背侧视觉通路(Im et al., 2021; Zhao, Shen et al., 2023),而对朝向集合的表征则不局限于枕叶,从早期视觉区域延伸至顶叶皮层(Tark et al., 2021)。同样,与背侧、腹侧通路紧密相关的大细胞(Magnocellular, M)与小细胞(Parvocellular, P)在集合感知中的作用也存在争议(Im et al., 2021; Lee & Chong, 2021)。支持大细胞参与的证据来自 Im 等人(2021)的 MEG 研究,

他们观察到,在面孔集合刺激呈现后约 68 ms 背侧通路便出现快速激活,其时间特征与大细胞快速传导的特性高度吻合。然而,一项采用快速闪烁适应范式(flicker adaptation paradigm)的行为研究却得出了相反的结论,抑制大细胞活动不仅未削弱集合感知的精度,反而提升了其精确性,并据此认为集合感知更依赖于小细胞所提供的精细信息(Lee & Chong, 2021)。未来研究需要关注腹侧通路和背侧通路以及大细胞和小细胞是否在集合感知过程中扮演不同的角色、是否具有不同的重要性。

此外,关于 V1 是否能处理统计信息,学界尚存争议。V1 区域的神经元感受野极小,往往不能同时覆盖一个集合中的多个项目,因此有理由相信集合感知发生于 V1 之后的更高级视觉区域。然而,支持这一观点的直接证据相对有限:Joo 等人(2009)发现,视觉统计信息的神经计算发生于双眼融合及双眼抑制之后,即处理阶段不早于 V1。与之相对,支持 V1 或更早期阶段参与的证据则来自多项研究报告。一方面,从早期的动物生理研究(Bonin et al., 2006),到针对纹外皮层及高级视觉区受损患者的观察(Norman et al., 2015),均支持方差表征发生于 V1。另一方面,近期的行为研究发现,在双眼信息融合之前大脑即可表征和提取圆形大小的均值与方差信息,暗示集合感知甚至可能起源于 V1 之前的皮层下结构(Zhao, Zeng et al., 2023; Zeng et al., 2024)。

重要的是,现有的神经影像学对于 V1 的研究结论亦不一致:Tark 等人(2021)使用 fMRI 和 IEM 分析发现,尽管当平均朝向在个体任务中作为无关特征时,枕叶 V1/V3 区域表现出了显著的集合朝向选择性反应,但综合考虑到 V1 感受野的大小限制,作者最终认为 V1 并不直接参与集合感知,而更强调 V3 区的作用。与此不同,Gong 等人(2025)利用高时间分辨率的 MEG,结合 IEM 和源重构(source reconstruction)分析表明,V1、V2、V3 对平均朝向编码均有贡献。他们从时间维度揭示了 V1 的延迟激活,提出 V1 可能通过两种机制实现对集合信息的编码:一是依赖于早期视觉区域内部的横向连接(horizontal connections),以促进个体项目之间的相互作用;二是利用循环信号的迭代或递归过程,优化集合内特征的权重分配,以实现精确的集合感知。

从行为证据到神经证据,我们难以从中得出关于人类集合感知脑机制的确定结论,不同研究结果的差异很可能来自于刺激特征本身所带来的差异。未来研究在解决这一问题时,也需要注意区分不同刺激水平可能造成的影响,同时需进一步回答平均表征和方差表征是否具有不同的神经机制。

6.3 通用性机制还是特异性机制

一个核心且悬而未决的问题是,视觉系统加工不同水平(如朝向、面孔身份)、同一水平不同视觉特征(如颜色、朝向)或不同统计指标(如均值、方差)的统计概要信息时,是依赖于领域通用的(domain-general)神经机制,还是存在多个领域特异的(domain-specific)机制?目前围绕这一问题的研究主要集中在行为层面,且结论存在诸多分歧(Chang, Cha, McGugin et al., 2024; Haberman, Brady & Alvarez, 2015)。

首先,早期基于个体差异方法的研究发现在高、低不同视觉水平,即跨领域(between-domain, high vs. low)之间任务成绩不相关,支持特异性机制(Haberman, Brady & Alvarez, 2015),但近期采用潜变量模型的研究则发现单一潜变量可以解释跨水平任务的表现,支持通用性假设(Chang, Cha & Gauthier, 2024)。而跨领域的方差适应后效研究则暗示方差表征可能存在通用机制(Maule & Franklin, 2020)。其次,在同一视觉水平内部(within-domain),即高或低水平各自领域内,低水平的颜色与亮度(Rajendran et al., 2021)、大小与朝向(Yörük & Boduroglu, 2020)可能存在特异性机制,但也有研究发现大小与朝向存在共享机制(Kacin et al., 2021; Yang et al., 2018);高水平特征(如面孔身份与情绪)可能存在分离(Haberman & Whitney, 2009; Kwon & Chong, 2023),但对非面孔复杂物体的研究则发现通用能力的存在(Chang & Gauthier, 2022)。最后,关于不同统计指标,如均值与方差,部分研究揭示了两者的独立加工过程(Norman et al., 2015; Michael et al., 2015),如对大小和朝向的均值和方差估计并无显著相关性(Khvostov & Utochkin, 2019; Yang et al., 2018);然而,更多研究揭示了两者的交互影响和共享机制(Chang, Cha & Gauthier, 2024; Hansmann-Roth et al., 2021; Jeong & Chong, 2020; Michael et al., 2014; Tong et al., 2015),如采用较大样本量的二

择一迫选任务中,被试对圆点集合的均值大小和方差大小判断表现出显著相关(Cha et al., 2022),且对颜色方差的知觉适应效应依赖于颜色均值(Maule & Franklin, 2020)。

行为层面的矛盾结果凸显了仅依赖行为指标推断潜在机制的局限性。行为相关性无法在时间进程上区分早期视觉特征提取和稍后期统计加工的心理过程,也难以区分共享的神经计算过程与其他潜在混淆因素。因此,要真正解决集合感知信息编码机制的“特异性”和“通用性”之争,未来的研究迫切需要超越行为层面的相关分析,直接采用神经科学的技术手段进行深入探究。

事实上,神经证据揭示大细胞相关的背侧通路可能在集合感知中扮演着关键角色,无论是光栅刺激还是面孔刺激,均发现了顶叶区域的显著激活(Tark et al., 2021; Im et al., 2017, 2021)。因此,我们提出在集合编码的早期,存在依赖于大细胞的领域通用性机制,随后是依赖于小细胞系统和各个脑区的特异性机制(见后文“粗略-细节-校准”模型)。这一模型整合了领域通用性机制和特异性机制的争议。未来可以利用基于 MVPA 的 fMRI 技术,结合表征相似性分析(Representational Similarity Analysis, RSA),直接比较不同类型刺激(高/低水平)、不同特征或不同统计指标(均值/方差)所激活的脑区模式是否存在重叠、是否存在相似的神经表征,从而判断其神经基础是共享还是分离。EEG/MEG 的高时间分辨率则有助于揭示集合感知在不同类型任务中,加工时间进程上的异同。有望更清晰地揭示集合感知在不同维度上的神经机制的差异与共性,为理解视觉系统高效处理复杂信息的机制提供更坚实的证据。

6.4 神经系统发育和经验影响

集合感知是自动性比较高的心理过程,我们会很自然地预期基因的影响大于经验的影响。遗憾的是,现有的少量研究并不能解决这个问题。

首先,在典型的神经发育过程中,集合感知能力随年龄增长而逐步提升。研究表明,尽管 4~5 岁的儿童已经能对物体集合的平均大小进行表征,但其表征效率却低于成人,这提示集合感知能力是一个逐渐发展成熟的过程(Sweeny et al., 2015)。Karaminis 等人(2018)在面孔情绪的集合感知研究中也观察到类似的发展趋势,儿童在平均情绪辨别、单一面孔情绪辨别的表现均劣于成人。这些

发现共同表明,随着视觉系统和相关认知功能的成熟,个体提取和利用统计概要信息的能力会得到增强。

其次,文化背景和个体经验同样塑造着集合感知。Zhao 等人(2024)在跨文化背景下考察了面孔情绪的集合感知,发现了显著的“他族效应”(other-ethnicity effect),被试对自己族裔面孔的情绪感知更为精确,但在不同种族混合的群体中倾向于高估其他种族面孔的情绪权重。这项发现与 Peng 等人(2021)观察到英国被试对自己性别面孔存在更强的平均表征倾向(而中国被试则无此现象)的研究结果相呼应,共同表明文化经验和社会群体认同可能调节着对面孔这类高级社会性刺激的集合感知机制。此外,奖赏价值也许并不影响平均本身,而只影响集合感知晚期的意识表征阶段(Dodgson & Raymond, 2020)。同样,特定的长期经验乃至相关的临床障碍也可能影响集合感知。Chang 等人(2025)针对网络游戏障碍(Internet Gaming Disorder, IGD)个体的一项 ERP 研究发现,IGD 患者在早期自动加工阶段对负面情绪存在注意偏向,且在后期的认知调控和抑制干扰方面则存在缺陷。这表明,长期的特定行为模式(如过度游戏)可能导致对社会性情绪信息的集合感知方式发生改变。

除了典型发育过程中的变化,集合感知的个体差异及其在特定神经发育障碍中的表现也开始受到关注,如“自闭症谱系障碍(Autism Spectrum Disorder, ASD)”和“注意缺陷多动障碍(Attention-Deficit/Hyperactivity Disorder, ADHD)”个体。Karaminis 等人(2018)发现,ASD 儿童(通常被认为在社会信息加工中存在困难)与典型发育儿童的任务表现并无显著差异,但计算建模结果表明,其集合感知能力与其非言语推理能力相关,暗示 ASD 儿童可能采用不同的认知策略来完成集合感知。而针对 ADHD 个体,Yuan 等人(2025)一项采用 rERPs 方法的大样本研究发现,ADHD 患者的注意缺陷症状与早期全局信息加工和晚期局部信息加工的神经反应减弱均存在显著相关,提示 ADHD 的核心认知缺陷可能涉及统计概要信息形成过程中的多个层面。

因此,集合感知能力并非一成不变,其形成和发展也不完全由先天因素决定,神经系统的发育成熟和个体经验亦对其产生重要影响。前人研

究揭示了发育异常与环境因素在塑造集合感知能力上可能的交互作用, 并为理解其个体差异和潜在的干预可能提供了新的视角。未来需进一步采用纵向设计追踪其发展轨迹, 并深入探讨不同经验因素影响集合感知的具体神经机制, 例如, 通过结合神经影像技术考察经验如何重塑集合感知的神经环路。

6.5 集合感知的神经活动是否等同于统计概要表征的神经活动

虽然多数研究者将两个概念混用, 如Whitney团队认为, 当人们观察一个集合时, 可以自动提取统计概要信息, 对集合信息的自动化编码等同于统计概要表征(Whitney et al., 2014; Whitney & Yamanashi Leib, 2018)。这种等同在行为研究中存在合理性, 但在神经科学研究中却存在问题。原因在于, 大脑在处理集合信息时, 集合成员数量的增加会导致神经信号的累加, 即活动水平增强。例如, 研究发现, 随着集合中项目数量的增加, N170 波幅随之增加(Puce et al., 2013); 甚

至在初级视觉皮层中, 集合中多个项目所引起的C1 波幅可能是其成员项目的线性叠加(Chen et al., 2016)。这种由于数量增加而导致的神经活动增强, 并不是统计概要表征的神经机制, 后者依赖于对集合内容的全局整合。

然而, 目前一些关于集合感知神经机制的研究, 大多通过比较集合和个体项目加工的差异来推论集合感知的神经机制(Im et al., 2017, 2021; Roberts et al., 2019), 这种神经活动差异不但包含了统计概要表征的脑机制, 也包含因为数量上的差别导致的神经活动的差异, 即心理过程不单一纯粹。因此, 未来研究需要明确意识到集合感知与统计概要表征在神经机制上的差异。

6.6 粗略-细节-校准(Coarse-Fine-Refine)模型

基于现有研究, 我们提出集合感知遵从“粗略-细节-校准”的多阶段过程(见图 2): “粗略”阶段依赖以大细胞通路驱动的背侧视觉通路, 是一个广泛分布的、跨领域的通用性机制, 能够在注意资源有限的情况下快速提取场景或集合“要点”

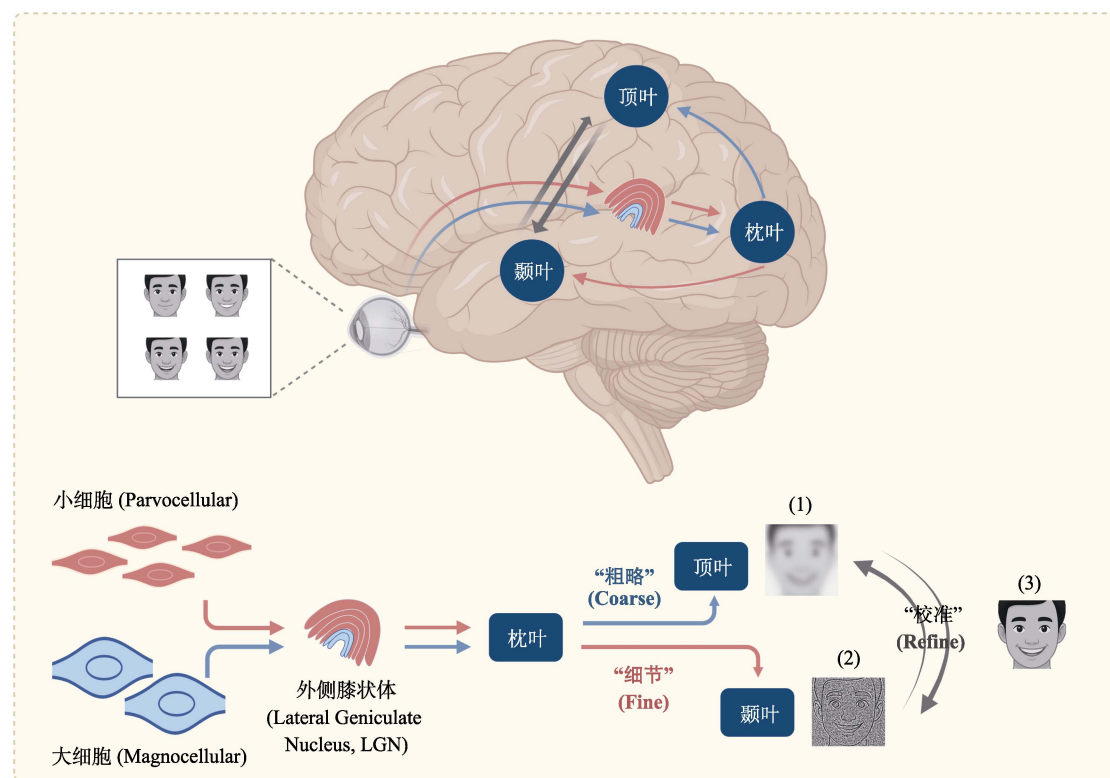


图2 视觉集合感知“粗略-细节-校准”模型示意图(以面孔情绪刺激为例): (1)“粗略”阶段加工低频率信息, 快速形成概要性的表征; (2)“细节”阶段加工高频率信息, 形成相对精细化的表征; (3)“校准”阶段通过前馈-反馈循环迭代形成的最终的精确化表征。

形成粗略的概要性表征或预期;随后的“细节”阶段则依赖于以小细胞通路驱动的腹侧视觉通路,是针对特定刺激特征的相对精细化加工,并体现出领域特异性;最后,粗略表征或预期反馈传递至各刺激的特异性脑区,通过前馈-反馈循环迭代对最初的概要性表征进行校准,实现最终精确化的集合感知。

对模型的推论如下:

首先,精确的集合感知的形成经历了多个阶段,该观点是在整合关于集合感知的时间进程系列神经证据的基础上所提出的。现有研究发现,集合信息可在刺激呈现后的极早期被大脑提取,并与个体/成员刺激的加工在神经活动上存在差异(Epstein & Emmanouil, 2021; Im et al., 2021; Roberts et al., 2019)。此外,早期脑电成分的出现反映了集合感知快速,甚至在一定程度上自动化的加工特点(Ji et al., 2018; Ji et al., 2024; Epstein & Emmanouil, 2021; 陈子龙, 季隼妍, 2025)。然而,使用 IEM 对 EEG/MEG 信号进行解码的研究揭示,虽然早期的神经信号确实存在,但稳定、精确的集合感知却在较晚阶段才与行为表现显著相关,这一过程可能涉及递归、迭代等更复杂的加工(Yashiro et al., 2024; Gong et al., 2025)。这些证据综合表明,集合感知是一个从快速、粗略的初步加工到晚期、精确的深度加工的多阶段过程。

其次,集合感知中的粗略信息和细节信息分别由快速的大细胞和慢速的小细胞通路传导,是基于以往理论的合理推测。以往针对客体 and 自然场景的高效识别过程的研究提出了由粗到细(Coarse to Fine)的观点,认为视觉系统首先通过大细胞快速处理低空间频率的粗略信息,并将其投射至高级皮层(如额顶叶),形成初步的“要点”感知或预期(Bar, 2003, 2004; Bar et al., 2006; Kveraga et al., 2007);随后,小细胞通路负责传递高空间频率信息,对细节信息进行精细化处理(Kveraga et al., 2007; Kandel et al., 2021; Kauffmann et al., 2014)。虽然集合信息不是客体或自然场景,但可以将之理解为由多个客体构成的特殊场景,大细胞高时间分辨率、低空间分辨率的生理特性,有助于对集合信息进行快速、前注意的加工,在背侧通路(如顶内沟)中形成统计概要的粗略信号。已有神经证据分别揭示了大细胞通路和小细胞通路在集合感知中的重要作用(Im

et al., 2017, 2021; Lee & Chong, 2021, 详见本文 6.2 节),由粗到细的观点恰好统一了这两种互相对立的证据。

再次,依赖于大细胞通路的粗略表征是领域通用的,而依赖于小细胞通路的细节表征是领域特异性的,这是综合现有证据的推论。研究发现,高水平的情绪面孔集合可以激活腹侧通路(Im et al., 2017),低水平的光栅朝向集合则激活了枕叶(Tark et al., 2021),似乎支持了领域特异性假设(Haberman, Brady & Alvarez, 2015; Whitney et al., 2014)。然而,不容忽视的是,无论是高水平的情绪面孔集合还是低水平的线段朝向集合,都能引起额顶叶区域的显著激活(Im et al., 2017, 2021; Tark et al., 2021; Gong et al., 2025; Michael et al., 2015),这便不能用单一的领域通用性机制或领域特异性机制来解释。结合大量研究揭示的大细胞通路善于在各种类型视觉刺激加工中发挥快速粗略表征的作用(Bar, 2003, 2004; Bar et al., 2006; Kveraga et al., 2007),以及传导速度稍慢的小细胞通路承担对细节进行精细化分析与识别的功能(Kveraga et al., 2007; Kandel et al., 2021; Kauffmann et al., 2014),我们有理由推测:集合感知可能在“粗略”阶段存在一个早期的、领域通用的神经基础;随后针对不同水平的特定刺激特征,在不同脑区进行差异化表征(特异性加工),完善了对“细节”的表征。该观点整合了集合感知究竟依赖于一种领域通用性机制还是多个特异性机制的争议(详见本文 6.3 节)。

最后,额顶网络形成的粗略表征或预期,可以自上而下地作用于各刺激特异性脑区的表征细节的前馈信号上(Gilbert & Sigman, 2007; Shipp, 2016),通过前馈-反馈循环迭代对初步的概要性印象进行校准,该观点同样是在预测编码理论和现有神经证据基础上提出的理论推论。当任务要求更高的精确度,或初步感知与预期不符时,额叶皮层等高级脑区便会通过反馈连接进行干预。例如,通过反馈连接,注意资源将被聚焦于集合中的一个或多个关键项目(如极端值或离群值),从而在最终的感知中导致项目被赋予不同的权重(Choi & Chong, 2020; De Fockert & Marchant, 2008)。此外,对离群值的抑制是反馈校准的另一个体现。研究发现,视觉系统能够稳健地计算出剔除离群值后的平均值(De Gardelle & Summerfield,

2011), 而这一过程被认为是一个相对缓慢、迭代的加工过程, 且高度依赖于反馈机制(Epstein et al., 2020)。通过前馈-反馈的循环迭代加工, 视觉系统不仅能够整合集合信息, 还能灵活地调整表征, 最终实现精确、稳健的集合感知。

综合来看, “粗略-细节-校准”模型是在预测编码理论和由粗到细理论观点的基础上(Gilbert & Sigman, 2007; He et al., 2012; Shipp, 2016; Bar, 2003, 2004; Bar et al., 2006; Kveraga et al., 2007), 针对集合感知自身的特性和最新的研究证据发展而来的, 与集合感知领域的主要理论观点均存在一些差别。首先, “粗略-细节-校准”模型包含了前馈和反馈过程, 而信号汇集假设和群体响应模型均强调前馈的直接作用。信号汇集假设是基于行为研究证据所提出的描述性假设, 提出了神经信号沿视觉层级汇集的可能性(Haberman & Whitney, 2012; Whitney et al., 2014); 群体响应模型则通过模拟神经元群体的调谐曲线, 为层级汇集提供了更具体的计算路径(Utochkin et al., 2024; Iakovlev & Utochkin, 2023)。“粗略-细节-校准”中的“细节”过程很可能通过层级汇集的方式实现。事实上, 从顶层到底层的反馈连接是可能的(Utochkin et al., 2024), 纯前馈的层级模型不足以解释离群值抑制等需要迭代加工的非线性现象(Gong et al., 2025; Wang et al., 2023)。所以在细节加工完成后, 开始了前馈-反馈循环迭代的“校准”过程。其次, “粗略-细节-校准”模型也在侧重点上有别于逆层级理论。逆层级理论主张在前馈阶段形成有意识的概要信息, 在反馈阶段将注意资源分配至局部细节上, 实现局部成员的表征(Hochstein & Ahissar, 2002; Hochstein et al., 2015)。虽然该理论也强调反馈的作用, 但是反馈的作用是实现局部表征, 遵循“先整体后局部”的过程。相比之下, 我们模型对应于逆层级理论的“整体”表征阶段, 即统计信息的形成本身便经历了多个动态阶段。

“粗略-细节-校准”假设整合了集合感知在时间进程、神经通路和计算机制上的多方面证据, 为理解视觉系统如何在效率与精度之间取得平衡提供了一个统一的视角。这虽为我们理解大脑如何高效处理视觉集合信息提供了更丰富的理解, 但仍需实证研究的进一步检验。

综上所述, 对视觉集合感知神经机制的探索虽已取得显著进展, 但仍处于深入发展的关键阶

段。未来的研究需要整合认知心理学、神经科学和计算建模的多元视角与方法, 聚焦于阐明其精确的神经通路与时间进程、前馈与反馈的协同作用、信息编码的通用性与特异性、发展与经验的影响, 并厘清不同概念之间的差异。这些努力将极大地加深我们对视觉系统如何从复杂信息中“化繁为简”实现高效抽象表征的理解。

参考文献

- 陈子龙, 季璟妍. (2025). 多面孔情绪变异性的自动加工: 来自视觉失匹配成分的证据. *心理学报*, 57(9), 1553-1571. <https://journal.psych.ac.cn/xlxb/CN/10.3724/SP.J.1041.2025.1553>
- 郝爽, 叶倩君, 何蔚祺. (2023). 群体面孔情绪的整体编码及其影响因素. *心理科学*, 46(01), 50-56. <https://jps.ecnu.edu.cn/CN/Y2023/V46/I1/50>
- 励奇添, 陈文锋. (2022). 集群表征在不同刺激及属性间的差异. *科学通报*, 67(21), 2463-2472. <https://doi.org/10.1360/TB-2021-1068>
- 刘丁睿. (2021). 集合表征对知觉决策调节效应的认知神经机制 [博士学位论文]. 中国科学院大学, 北京.
- 田欣然, 侯文霞, 欧玉晓, 易冰, 陈文锋, 尚俊辰. (2021). 基于合成平均刺激的平均表征机制——来自平均面孔吸引力的证据. *心理学报*, 53(7), 714-728. <https://doi.org/10.3724/sp.J.1041.2021.00714>
- 全可, 唐薇, 陈文锋, 傅小兰. (2015). 统计概要表征的内容与机制. *心理科学进展*, 23(10), 1723-1731. <https://doi.org/10.3724/sp.J.1042.2015.01723>
- 赵冰洁, 何婕, 刘颖, 杨邵峰, 王峥, 张琪涵, 白学军. (2024). 凸显刺激对整体知觉集合数量效应的影响. *心理与行为研究*, 22(2), 212. <https://psybeh.tjnu.edu.cn/CN/Y2024/V22/I2/212>
- Allard, R., Ramanoël, S., Silvestre, D., & Arleo, A. (2021). Variance-dependent neural activity in an involuntary averaging task. *Attention, Perception, & Psychophysics*, 83(3), 1094-1105. <https://doi.org/10.3758/s13414-020-02223-8>
- Allik, J., Toom, M., Naar, R., & Raidvee, A. (2022). How are local orientation signals pooled? *Attention, Perception, & Psychophysics*, 84(3), 981-991. <https://doi.org/10.3758/s13414-022-02456-9>
- Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, 15(3), 122-131. <https://doi.org/10.1016/j.tics.2011.01.003>
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 12, 157-162. <https://doi.org/10.1111/1467-9280.00327>
- Bar, M. (2003). A cortical mechanism for triggering top-down facilitation in visual object recognition. *Journal of Cognitive Neuroscience*, 15(4), 600-609. <https://doi.org/10.1162/089892903321662976>
- Bar, M. (2004). Visual objects in context. *Nature Reviews*

- Neuroscience*, 5(8), 617–629. <https://doi.org/10.1038/nrn1476>
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., ... Halgren, E. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences*, 103(2), 449–454. <https://doi.org/10.1073/pnas.0507062103>
- Bauer, B. (2009). The danger of trial-by-trial knowledge of results in perceptual averaging studies. *Attention, Perception, & Psychophysics*, 71(3), 655–665. <https://doi.org/10.3758/APP.71.3.655>
- Bonin, V., Mante, V., & Carandini, M. (2006). The statistical computation underlying contrast gain control. *Journal of Neuroscience*, 26(23), 6346–6353. <https://doi.org/10.1523/JNEUROSCI.0284-06.2006>
- Cha, O., Blake, R., & Gauthier, I. (2022). Contribution of a common ability in average and variability judgments. *Psychonomic Bulletin & Review*, 29(1), 108–115. <https://doi.org/10.3758/s13423-021-01982-1>
- Chaney, W., Fischer, J., & Whitney, D. (2014). The hierarchical sparse selection model of visual crowding. *Frontiers in Integrative Neuroscience*, 8, 73. <https://doi.org/10.3389/fnint.2014.00073>
- Chang, Q., Hao, B., Fan, C., Luo, W., & He, W. (2025). Ensemble coding of crowd facial emotion in Internet gaming disorder under the emotional interference condition: An ERP study. *Journal of Behavioral Addictions*, 14(2), 817–830. <https://doi.org/10.1556/2006.2025.00027>
- Chang, T. Y., Cha, O., & Gauthier, I. (2024). A general ability for judging simple and complex ensembles. *Journal of Experimental Psychology: General*, 153(6), 1517–1536. <https://doi.org/10.1037/xge0001582>
- Chang, T. Y., Cha, O., McGugin, R., Tomarken, A., & Gauthier, I. (2024). How general is ensemble perception?. *Psychological Research*, 88(3), 695–708. <https://doi.org/10.1007/s00426-023-01883-z>
- Chang, T. Y., & Gauthier, I. (2022). Domain-general ability underlies complex object ensemble processing. *Journal of Experimental Psychology: General*, 151(4), 966–972. <https://doi.org/10.1037/xge0001110>
- Chen, J., Yu, Q., Zhu, Z., Peng, Y., & Fang, F. (2016). Spatial summation revealed in the earliest visual evoked component C1 and the effect of attention on its linearity. *Journal of Neurophysiology*, 115(1), 500–509. <https://doi.org/10.1152/jn.00044.2015>
- Choi, J., & Chong, S. C. (2020). Contextual cueing in target absent trials by distractor–distractor associations. *Journal of Experimental Psychology: Human Perception and Performance*, 46(12), 1458. <https://doi.org/10.1037/xhp0000867>
- Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, 43(4), 393–404. [https://doi.org/10.1016/S0042-6989\(02\)00596-5](https://doi.org/10.1016/S0042-6989(02)00596-5)
- Chwe, J. A. H., & Freeman, J. B. (2024). Trustworthiness of crowds is gleaned in half a second. *Social Psychological and Personality Science*, 15(3), 351–359. <https://doi.org/10.1177/19485506231164703>
- Corbett, J. E., Utochkin, I., & Hochstein, S. (2023). *The pervasiveness of ensemble perception: Not just your average review*. Cambridge University Press.
- Davis, E. E., Matthews, C. M., & Mondloch, C. J. (2021). Ensemble coding of facial identity is not refined by experience: Evidence from other-race and inverted faces. *British Journal of Psychology*, 112(1), 265–281. <https://doi.org/10.1111/bjop.12457>
- De Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, 70(5), 789–794. <https://doi.org/10.3758/PP.70.5.789>
- de Fockert, J., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology*, 62(9), 1716–1722. <https://doi.org/10.1080/17470210902811249>
- De Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*, 108(32), 13341–13346. <https://doi.org/10.1073/pnas.1104517108>
- Dodgson, D. B., & Raymond, J. E. (2020). Value associations bias ensemble perception. *Attention, Perception, & Psychophysics*, 82(1), 109–117. <https://doi.org/10.3758/s13414-019-01744-1>
- Epstein, M. L., & Emmanouil, T. A. (2021). Ensemble statistics can be available before individual item properties: Electroencephalography evidence using the oddball paradigm. *Journal of Cognitive Neuroscience*, 33(6), 1056–1068. https://doi.org/10.1162/jocn_a_01704
- Epstein, M. L., Quilty-Dunn, J., Mandelbaum, E., & Emmanouil, T. A. (2020). The outlier paradox: The role of iterative ensemble coding in discounting outliers. *Journal of Experimental Psychology: Human Perception and Performance*, 46(11), 1267. <https://doi.org/10.1037/xhp0000857>
- Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*, 106(3), 1389–1398. <https://doi.org/10.1152/jn.00904.2010>
- Florey, J., Clifford, C. W., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports*, 6, 32210. <https://doi.org/10.1038/srep32210>
- Gilbert, C. D., & Sigman, M. (2007). Brain states: Top-down influences in sensory processing. *Neuron*, 54(5), 677–696. <https://doi.org/10.1016/j.neuron.2007.05.019>
- Gong, X., He, T., Wang, Q., Lu, J., & Fang, F. (2025). Time course of orientation ensemble representation in the human brain. *Journal of Neuroscience*, 45(7). <https://doi.org/10.1523/JNEUROSCI.1688-23.2024>
- Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology: General*, 144(2), 432–446. <https://doi.org/10.1037/xge0000053>

- Haberman, J., Harp, T., & Whitney, D. (2009). Averaging facial expression over time. *Journal of Vision*, 9(11), 1–1. <https://doi.org/10.1167/9.11.1>
- Haberman, J., Lee, P., & Whitney, D. (2015). Mixed emotions: Sensitivity to facial variance in a crowd of faces. *Journal of Vision*, 15(4), 16. <https://doi.org/10.1167/15.4.16>
- Haberman, J., & Suresh, S. (2021). Ensemble size judgments account for size constancy. *Attention, Perception, & Psychophysics*, 83(3), 925–933. <https://doi.org/10.3758/s13414-020-02144-6>
- Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, 17(17), R751–R753. <https://doi.org/10.1016/j.cub.2007.06.039>
- Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, 35(3), 718–734. <https://doi.org/10.1037/a0013899>
- Haberman, J., & Whitney, D. (2012). Ensemble perception: Summarizing the scene and broadening the limits of visual processing. In J. Wolfe & L. Robertson (Eds.), *From perception to consciousness: Searching with Anne Treisman* (pp. 339–349). Oxford University Press. <https://doi.org/10.1093/acprof:osobl/9780199734337.003.0030>
- Hansmann-Roth, S., Kristjánsson, Á., Whitney, D., & Chetverikov, A. (2021). Dissociating implicit and explicit ensemble representations reveals the limits of visual perception and the richness of behavior. *Scientific Reports*, 11(1), 3899. <https://doi.org/10.1038/s41598-021-83358-y>
- He, D., Kersten, D., & Fang, F. (2012). Opposite modulation of high-and low-level visual aftereffects by perceptual grouping. *Current Biology*, 22(11), 1040–1045. <https://doi.org/10.1016/j.cub.2012.04.026>
- Hochstein, S., & Ahissar, M. (2002). View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5), 791–804. [https://doi.org/10.1016/S0896-6273\(02\)01091-7](https://doi.org/10.1016/S0896-6273(02)01091-7)
- Hochstein, S., Pavlovskaya, M., Bonneh, Y. S., & Sorofer, N. (2015). Global statistics are not neglected. *Journal of Vision*, 15(4), 7. <https://doi.org/10.1167/15.4.7>
- Iakovlev, A. U., & Utochkin, I. S. (2023). Ensemble averaging: What can we learn from skewed feature distributions? *Journal of Vision*, 23(1), 5. <https://doi.org/10.1167/jov.23.1.5>
- Im, H. Y., Albohn, D. N., Steiner, T. G., Cushing, C. A., Adams, R. B., Jr., & Kveraga, K. (2017). Differential hemispheric and visual stream contributions to ensemble coding of crowd emotion. *Nature Human Behaviour*, 1, 828–842. <https://doi.org/10.1038/s41562-017-0225-z>
- Im, H. Y., Cushing, C. A., Ward, N., & Kveraga, K. (2021). Differential neurodynamics and connectivity in the dorsal and ventral visual pathways during perception of emotional crowds and individuals: A MEG study. *Cognitive, Affective, & Behavioral Neuroscience*, 21(4), 776–792. <https://doi.org/10.3758/s13415-021-00880-2>
- Jeong, J., & Chong, S. C. (2020). Adaptation to mean and variance: Interrelationships between mean and variance representations in orientation perception. *Vision Research*, 167, 46–53. <https://doi.org/10.1016/j.visres.2020.01.002>
- Ji, L., Chen, Z., Zeng, X., Sun, B., & Fu, S. (2024). Automatic processing of unattended mean emotion: Evidence from visual mismatch responses. *Neuropsychologia*, 202, 108963. <https://doi.org/10.1016/j.neuropsychologia.2024.108963>
- Ji, L., Rossi, V., & Pourtois, G. (2018). Mean emotion from multiple facial expressions can be extracted with limited attention: Evidence from visual ERPs. *Neuropsychologia*, 111, 92–102. <https://doi.org/10.1016/j.neuropsychologia.2018.01.022>
- Jia, J., Wang, T., Chen, S., Ding, N., & Fang, F. (2022). Ensemble size perception: Its neural signature and the role of global interaction over individual items. *Neuropsychologia*, 173, 108290. <https://doi.org/10.1016/j.neuropsychologia.2022.108290>
- Joo, S. J., Shin, K., Chong, S. C., & Blake, R. (2009). On the nature of the stimulus information necessary for estimating mean size of visual arrays. *Journal of Vision*, 9(9), 7–12. <https://doi.org/10.1167/9.9.7>
- Kacin, M., Gauthier, I., & Cha, O. (2021). Ensemble coding of average length and average orientation are correlated. *Vision Research*, 187, 94–101. <https://doi.org/10.1016/j.visres.2021.04.010>
- Kanaya, S., Hayashi, M. J., & Whitney, D. (2018). Exaggerated groups: Amplification in ensemble coding of temporal and spatial features. *Proceedings of the Royal Society B: Biological Sciences*, 285(1879), 20172770. <https://doi.org/10.1098/rspb.2017.2770>
- Kandel, E. R., Koester, J. D., Mack, S., & Siegelbaum, S. A. (Eds.). (2021). *Principles of neural science* (6th ed.). McGraw-Hill.
- Karaminis, T., Neil, L., Manning, C., Turi, M., Fiorentini, C., Burr, D., & Pellicano, E. (2018). Reprint of “Investigating ensemble perception of emotions in autistic and typical children and adolescents”. *Developmental Cognitive Neuroscience*, 29, 97–107. <https://doi.org/10.1016/j.dcn.2018.02.003>
- Kauffmann, L., Ramanoël, S., & Peyrin, C. (2014). The neural bases of spatial frequency processing during scene perception. *Frontiers in Integrative Neuroscience*, 8, 37. <https://doi.org/10.3389/fnint.2014.00037>
- Khvostov, V. A., & Utochkin, I. S. (2019). Independent and parallel visual processing of ensemble statistics: Evidence from dual tasks. *Journal of Vision*, 19(9), 3. <https://doi.org/10.1167/19.9.3>
- Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition*, 65(2), 145–168. <https://doi.org/10.1016/j.bandc.2007.06.007>
- Kwon, D., & Chong, S. C. (2023). The relationship between ensemble representations of facial information. *Vision*

- Research*, 203, 108156. <https://doi.org/10.1016/j.visres.2022.108156>
- Lee, J., & Chong, S. C. (2021). Quality of average representation can be enhanced by refined individual items. *Attention, Perception, & Psychophysics*, 83(3), 970–981. <https://doi.org/10.3758/s13414-020-02139-3>
- Li, H., Ji, L., Tong, K., Ren, N., Chen, W., Liu, C. H., & Fu, X. (2016). Processing of individual items during ensemble coding of facial expressions. *Frontiers in Psychology*, 7, 1332. <https://doi.org/10.3389/fpsyg.2016.01332>
- Liu, L., Wang, F., Zhou, K., Ding, N., & Luo, H. (2017). Perceptual integration rapidly activates dorsal visual pathway to guide local processing in early visual areas. *PLoS Biology*, 15(11), e2003646. <https://doi.org/10.1371/journal.pbio.2003646>
- Liu, R., Ye, Q., Hao, S., Li, Y., Shen, L., & He, W. (2023). The relationship between ensemble coding and individual representation of crowd facial emotion. *Biological Psychology*, 180, 108593. <https://doi.org/10.1016/j.biopsycho.2023.108593>
- Lukashevich, A., Sigurdardottir, H. M., Kudriavtsev, N., & Utochkin, I. (2025). The role of attention in basic ensemble statistics processing. *Neuropsychologia*, 208, 109086. <https://doi.org/10.1016/j.neuropsychologia.2025.109086>
- Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica*, 142(2), 245–250. <https://doi.org/10.1016/j.actpsy.2012.11.002>
- Marini, F., Sutherland, C. A., Ostrovska, B., & Manassi, M. (2023). Three's a crowd: Fast ensemble perception of first impressions of trustworthiness. *Cognition*, 239, 105540. <https://doi.org/10.1016/j.cognition.2023.105540>
- Maule, J., & Franklin, A. (2020). Adaptation to variance generalizes across visual domains. *Journal of Experimental Psychology: General*, 149(4), 662–675. <https://doi.org/10.1037/xge0000678>
- Michael, E., de Gardelle, V., Nevado-Holgado, A., & Summerfield, C. (2015). Unreliable evidence: 2 sources of uncertainty during perceptual choice. *Cerebral Cortex*, 25(4), 937–947. <https://doi.org/10.1093/cercor/bht287>
- Michael, E., de Gardelle, V., & Summerfield, C. (2014). Priming by the variability of visual information. *Proceedings of the National Academy of Sciences*, 111(21), 7873–7878. <https://doi.org/10.1073/pnas.1308674111>
- Nemrodov, D., Ling, S., Nudnou, I., Roberts, T., Cant, J. S., Lee, A. C. H., & Nestor, A. (2020). A multivariate investigation of visual word, face, and ensemble processing: Perspectives from EEG-based decoding and feature selection. *Psychophysiology*, 57(3), e13511. <https://doi.org/10.1111/psyp.13511>
- Nguyen, T. T. N., Vuong, Q. C., Mather, G., & Thornton, I. M. (2021). Ensemble coding of crowd speed using biological motion. *Attention, Perception, & Psychophysics*, 83(3), 1014–1035. <https://doi.org/10.3758/s13414-020-02163-3>
- Norman, L. J., Heywood, C. A., & Kentridge, R. W. (2015). Direct encoding of orientation variance in the visual system. *Journal of Vision*, 15(4), 3. <https://doi.org/10.1167/15.4.3>
- Oh, B. I., Kim, Y. J., & Kang, M. S. (2019). Ensemble representations reveal distinct neural coding of visual working memory. *Nature Communications*, 10(1), 5665. <https://doi.org/10.1038/s41467-019-13592-6>
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7), 739–744. <https://doi.org/10.1038/89532>
- Pavlovskaya, M., Soroker, N., Bonneh, Y. S., & Hochstein, S. (2015). Computing an average when part of the population is not perceived. *Journal of Cognitive Neuroscience*, 27(7), 1397–1411. https://doi.org/10.1162/jocn_a_00791
- Peng, S., Liu, C. H., & Hu, P. (2021). Effects of subjective similarity and culture on ensemble perception of faces. *Attention, Perception, & Psychophysics*, 83(3), 1070–1079. <https://doi.org/10.3758/s13414-020-02133-9>
- Phillips, L. T., Slepian, M. L., & Hughes, B. L. (2018). Perceiving groups: The people perception of diversity and hierarchy. *Journal of Personality and Social Psychology*, 114(5), 766–785. <https://doi.org/10.1037/pspi0000120>
- Puce, A., McNeely, M. E., Berrebi, M. E., Thompson, J. C., Hardee, J., & Brefczynski-Lewis, J. (2013). Multiple faces elicit augmented neural activity. *Frontiers in Human Neuroscience*, 7, 282. <https://doi.org/10.3389/fnhum.2013.00282>
- Rajendran, S., Maule, J., Franklin, A., & Webster, M. A. (2021). Ensemble coding of color and luminance contrast. *Attention, Perception, & Psychophysics*, 83(3), 911–924. <https://doi.org/10.3758/s13414-020-02136-6>
- Roberts, T., Cant, J. S., & Nestor, A. (2019). Elucidating the neural representation and the processing dynamics of face ensembles. *The Journal of Neuroscience*, 39(39), 7737–7747. <https://doi.org/10.1523/JNEUROSCI.0471-19.2019>
- Robinson, M. M., & Brady, T. F. (2023). A quantitative model of ensemble perception as summed activation in feature space. *Nature Human Behaviour*, 7(10), 1638–1651. <https://doi.org/10.1038/s41562-023-01602-z>
- Rousset, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences*, 8(8), 363–370. <https://doi.org/10.1016/j.tics.2004.06.003>
- Sama, M. A., Nestor, A., & Cant, J. S. (2024). The neural dynamics of face ensemble and central face processing. *Journal of Neuroscience*, 44(7), e1027232023. <https://doi.org/10.1523/JNEUROSCI.1027-23.2023>
- Shipp, S. (2016). Neural elements for predictive coding. *Frontiers in Psychology*, 7, 1792. <https://doi.org/10.3389/fpsyg.2016.01792>

- fpsyg.2016.01792
- Sun, P., Chu, V., & Sperling, G. (2021). Multiple concurrent centroid judgments imply multiple within-group salience maps. *Attention, Perception, & Psychophysics*, 83(3), 934–955. <https://doi.org/10.3758/s13414-020-02197-7>
- Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of Experimental Psychology: Human Perception and Performance*, 39(2), 329–337. <https://doi.org/10.1037/a0028712>
- Sweeny, T. D., Wurnitsch, N., Gopnik, A., & Whitney, D. (2015). Ensemble perception of size in 4-5-year-old children. *Developmental Science*, 18(4), 556–568. <https://doi.org/10.1111/desc.12239>
- Tark, K. J., Kang, M. S., Chong, S. C., & Shim, W. M. (2021). Neural representations of ensemble coding in the occipital and parietal cortices. *Neuroimage*, 245, 118680. <https://doi.org/10.1016/j.neuroimage.2021.118680>
- Tiurina, N. A., Markov, Y. A., Whitney, D., & Pascucci, D. (2024). The functional role of spatial anisotropies in ensemble perception. *BMC Biology*, 22(1), 28. <https://doi.org/10.1186/s12915-024-01822-3>
- Tokita, M., Ueda, S., & Ishiguchi, A. (2016). Evidence for a global sampling process in extraction of summary statistics of item sizes in a set. *Frontiers in Psychology*, 7, 711. <https://doi.org/10.3389/fpsyg.2016.00711>
- Tong, K., Ji, L., Chen, W., & Fu, X. (2015). Unstable mean context causes sensitivity loss and biased estimation of variability. *Journal of Vision*, 15(4), 15. <https://doi.org/10.1167/15.4.15>
- Utochkin, I. S., Choi, J., & Chong, S. C. (2024). A population response model of ensemble perception. *Psychological Review*, 131(1), 36–57. <https://doi.org/10.1037/rev0000426>
- Utochkin, I. S., & Tiurina, N. A. (2014). Parallel averaging of size is possible but range-limited: A reply to Marchant, Simons, and De Fockert. *Acta Psychologica*, 146, 7–18. <https://doi.org/10.1016/j.actpsy.2013.11.012>
- Wang, T., Zhao, Y., & Jia, J. (2023). Nonadditive integration of visual information in ensemble processing. *Iscience*, 26(10). <https://doi.org/10.1016/j.isci.2023.107988>
- Ward, E. J., Bear, A., & Scholl, B. J. (2016). Can you perceive ensembles without perceiving individuals?: The role of statistical perception in determining whether awareness overflows access. *Cognition*, 152, 78–86. <https://doi.org/10.1016/j.cognition.2016.01.010>
- Wardle, S. G., Bex, P. J., Cass, J., & Alais, D. (2012). Stereoacuity in the periphery is limited by internal noise. *Journal of Vision*, 12(6), 12–12. <https://doi.org/10.1167/12.6.12>
- Whitney, D., Haberman, J., & Sweeny, T. D. (2014). From textures to crowds: Multiple levels of summary statistical perception. In J. S. Werner, L. M. Chalupa, & M. E. Burns (Eds.), *The new visual neurosciences* (pp. 695–710). MIT Press.
- Whitney, D., & Yamanashi Leib, A. (2018). Ensemble perception. *Annual Review of Psychology*, 69, 105–129. <https://doi.org/10.1146/annurev-psych-010416-044232>
- Wolfe, B. A., Kosovicheva, A. A., Leib, A. Y., Wood, K., & Whitney, D. (2015). Foveal input is not required for perception of crowd facial expression. *Journal of Vision*, 15(4), 11. <https://doi.org/10.1167/15.4.11>
- Yamanashi Leib, A., Landau, A. N., Baek, Y., Chong, S. C., & Robertson, L. (2012). Extracting the mean size across the visual field in patients with mild, chronic unilateral neglect. *Frontiers in Human Neuroscience*, 6, 267. <https://doi.org/10.3389/fnhum.2012.00267>
- Yang, Y., Tokita, M., & Ishiguchi, A. (2018). Is there a common summary statistical process for representing the mean and variance? A study using illustrations of familiar items. *i-Perception*, 9(1), 2041669517747297. <https://doi.org/10.1177/2041669517747297>
- Yashiro, R., Sawayama, M., & Amano, K. (2024). Decoding time-resolved neural representations of orientation ensemble perception. *Frontiers in Neuroscience*, 18, 1387393. <https://doi.org/10.3389/fnins.2024.1387393>
- Ying, H. (2022). Attention modulates the ensemble coding of facial expressions. *Perception*, 51(4), 276–285. <https://doi.org/10.1177/03010066221079686>
- Yörük, H., & Boduroglu, A. (2020). Feature-specificity in visual statistical summary processing. *Attention, Perception, & Psychophysics*, 82(2), 852–864. <https://doi.org/10.3758/s13414-019-01942-x>
- Yuan, J., Pan, H., Sun, Y., Wang, Y., & Jia, J. (2025). Neural responses to global and local visual information processing provide neural signatures of ADHD symptoms. *International Journal of Psychophysiology*, 112582. <https://doi.org/10.1016/j.ijpsycho.2025.112582>
- Zeng, T., Zhao, Y., Cao, B., & Jia, J. (2024). Perception of visual variance is mediated by subcortical mechanisms. *Brain and Cognition*, 175, 106131. <https://doi.org/10.1016/j.bandc.2024.106131>
- Zhang, X., Qiu, J., Zhang, Y., Han, S., & Fang, F. (2014). Misbinding of color and motion in human visual cortex. *Current Biology*, 24(12), 1354–1360. <https://doi.org/10.1016/j.cub.2014.04.045>
- Zhao, D., Shen, X., Li, S., & He, W. (2023). The impact of spatial frequency on the perception of crowd emotion: An fMRI study. *Brain Sciences*, 13(12), 1699. <https://doi.org/10.3390/brainsci13121699>
- Zhao, Y., Zeng, T., Wang, T., Fang, F., Pan, Y., & Jia, J. (2023). Subcortical encoding of summary statistics in humans. *Cognition*, 234, 105384. <https://doi.org/10.1016/j.cognition.2023.105384>
- Zhao, Z., Yaoma, K., Wu, Y., Burns, E., Sun, M., & Ying, H. (2024). Other ethnicity effects in ensemble coding of facial expressions. *Attention, Perception, & Psychophysics*, 86(7), 2412–2423. <https://doi.org/10.3758/s13414-024-02920-8>

Simplify complexity: The neural mechanisms underlying ensemble perception

SUN Huanxiang¹, ZHANG Fan^{1,2}, LI Sijia¹, ZHANG Xiuling¹, JIANG Yi^{3,4}

(¹ School of Psychology, Northeast Normal University, Changchun 130024, China)

(² Jinzhong College of Information, Jinzhong 030800, China)

(³ State Key Laboratory of Brain and Cognitive Science, Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China)

(⁴ Department of Psychology, University of Chinese Academy of Sciences, Beijing 100049, China)

Abstract: Ensemble perception, the process by which the visual system extracts summary statistical information (for example, mean and variance) from groups of similar objects at a glance, is critical to human adaptive functioning. The precise neural mechanisms underlying ensemble perception remain elusive. The present work reviews the temporal dynamics of ensemble perception, evaluates the leading theoretical models with key empirical findings, and further distinguishes the neural substrates between ensemble and single-item processing. Based on existing evidence, the review proposes a “Coarse-Fine-Refine” model, which suggests that during ensemble perception, neural processing proceeds from domain-general, coarse representations mediated by the magnocellular pathway to more detailed, domain-specific representations by parvocellular routes within feature-selective brain regions. Finally, recurrent feedback loops enable iterative calibration of the summary estimate. The review also outlines promising future research directions, including (1) locating the precise neural circuits and cortical areas of ensemble processing, (2) delineating the contributions of feedforward versus feedback signaling, (3) clarifying the domain-general and domain-specific coding in ensemble perception, (4) examining the role of experience and neural development in shaping ensemble perception.

Keywords: ensemble perception, statistical summary representation, perceptual integration, temporal dynamics, neural mechanism