

## 《心理学报》审稿意见与作者回应

题目：多种情感与认知任务驱动下大脑可泛化神经表征的发育模式

作者：郝磊，许天委，周文龙，杨杰，彭思雅，刘明兰，徐家华，王延培，谭淑平，高家红，  
贺永，陶沙，董奇，秦绍正

### 第一轮

#### 审稿人 1 意见：

该研究从跨不同情感与认知域的角度印证了交互式特异化理论的构想，引入了多认知域和子域对应的多层级结构信息，考察了儿童多种情感认知能力发展的可泛化神经表征脑发育机制。理论基础扎实，研究方法具有较高的创新性，使我们对儿童青少年的脑智发育规律有了全新视角的理解。但是，我仍有一些针对文章各部分的修改建议，希望在发表前对该研究工作的改进有所帮助。

回应：我们感谢审稿人积极、热情的评价，很高兴审稿人认可我们研究中理论基础的扎实和在方法上的创新。根据审稿人的评论，我们进行了认真的修改，相信这些改进使得手稿得到了很大的改进。

意见 1：在引言部分，需要对神经特异化和神经泛化进行明白的定义，以方便读者更好的理解研究的核心思想。

回应：针对神经特异化 (Neural specialization)，在修改稿中我们已在正文的第二段进行了概念的界定，内容为：“在发展认知神经科学领域，神经特异化的概念最早是由弗朗兹·约瑟夫·加尔 (Franz Joseph Gall, 1758-1828) 提出的，他认为大脑是思维的载体，思维是由不同的心智能力组成的，而不同的心智能力各自位于特定的大脑皮层中 (Kanwisher, 2010; Leigh, 1977)。之后形成的“成熟理论” (The maturational perspective) 也同样认为，大脑特定区域的成熟与新的感觉、运动或高级认知功能的出现是有直接联系的 (Johnson, 2001; Kanwisher, 2010)。” (769 行-773 行)；针对神经泛化 (Neural generalization)，由于这是一个较新的概念，此前很少有研究将神经泛化发育与心理发展结合探索，因此还未形成权威性的概念界定，这

也是本文在理论前沿进行拓展和方法上进行创新的核心表现,我们在修改稿中也对此概念的描述进行了更新,内容为:“从脑智发育的视角出发,神经泛化即某一特定神经系统在发育中逐渐成熟从而支持多种情感与认知发展的过程。”(850行-851行)

**意见 2:** 该研究并未使用功能连接或功能网络相关的分析方法,却在多处使用了神经网络的专业词汇,容易给读者造成方法上的误解,请作者针对本研究使用的研究方法进行改进。

**回应:** 在本研究中,确实未使用脑网络相关的数据分析方法,“神经网络”是指不同情感与认知功能对应的多脑区组成的神经系统,确实易与注意网络(此处网络代表不同注意功能)产生概念混淆,因此,我们在修改稿中不再使用“神经网络”一词,改用“神经系统”代替。

**意见 3:** 请作者在方法部分对各情感和认知条件的神经建模进行清楚明确的图示化展示,使读者能够更清晰的理解本研究中多情感与认知的本质。

**回应:** 我们已在修改稿中增加了各情感与认知条件的图示(图2),并在“2.3 单变量一般线性模型分析”部分改进了神经建模的描述,内容为:“为评估 ANT, WM 和 EM 中与任务相关的大脑激活反应,我们分别构建了三个任务中各实验条件的回归因子。对于 ANT,将条件(无提示、双侧提示、中心提示、空间提示)建模为独立的事件相关回归因子,包括警觉(双侧提示 v.s. 无提示)和定向(空间提示 v.s. 中心提示)条件;对于 WM 任务,将两个条件(1-Back 和 2-Back)建模为单独的回归变量;对于 EM 任务,将情绪和方向识别控制条件建模为单独的回归变量。”(963行-967行),以方便读者清晰的知晓本研究中所使用的情感与认知任务都包含了哪些具体的条件。

**意见 4:** 图6的结果呈现需要更加明确,左侧的小提琴图与右侧的脑区展示,两者的对应关系不明确,可以考虑换一种意义明确的呈现方式。

**回应:** 感谢审稿人提出图6呈现方式的问题,我们对此进行了改进,在图6右侧各感兴趣区标记了名称,以便读者更够更直观的对应各小提琴图结果针对的脑区。同时也改进了该结果部分的文字叙述,改进的内容为:“我们在检查跨情感与认知领域的神经表征泛化水平的发育性差异时发现,成人在多元需求额顶系统(双侧 IPS + 双侧 FEF:  $p = 0.0018$ )中表现出了显著高于随机水平的跨情感与认知的神经泛化水平,特别是在双侧 IPS 中( $p = 0.004$ ),但是,在控制分析的 ROI(包括双侧 dACC、双侧 DLPFC 和双侧 AI,  $p_s \geq 0.991$ )中却未发现统计上显著的皮层区域。”(1056行-1059行)

**意见 5:** 顶内沟皮层的神经泛化规律作为本研究的核心结果，并没有进行针对性的讨论，请作者予以补充。

**回应:** 我们在修改稿中，已增加关于多元需求系统中顶内沟皮层的讨论。增加内容为：“多元需求额顶系统共同参与着多种情感与认知功能过程(Cole & Schneider, 2007; Dosenbach et al., 2006; Duncan & Owen, 2000; Fedorenko et al., 2013; Hugdahl et al., 2015)，本研究的结果在注意、工作记忆和情绪任务中印证了这一现象。只有成人在多元需求额顶系统中，特别是顶内沟皮层区域，表现出了显著的跨情感与认知领域的神经泛化水平，此外，三组被试在多元需求额顶系统也表现出了显著的发育性差异，说明了儿童该脑区的组构性信息编码加工机制还未成熟，在支持不同情感与认知任务时未能执行稳定的差异性共激活模式(Reverberi et al., 2012)，其表现就是虽然在各过程均被激活，但是却未能对不同层级情感或认知功能相应的神经表征信息进行有效的信息组织。”（1106 行-1114 行）

**意见 6:** 针对研究结果的讨论，作者认为神经泛化的基础可能与认知灵活切换有关，本研究使用的多个情感与认知任务，并非是在一个任务中进行了多个情感与认知任务，因此不涉及多个认知间的灵活切换，需要在文章的局限性中有所提及。

**回应:** 审稿人提出的这个问题使我们对本研究的结果有了新的思考，已有研究发现多元需求额顶系统在调用不同认知过程时扮演着支持认知灵活切换的功能，本研究使用的多个情感与认知任务，并非是在一次任务中进行了多个情感与认知过程，因此在对结果进行解释时，并不能从认知灵活性的角度出发进行讨论，其基础更有可能是神经系统组构性的编码机制支持着脑智发育过程的神经泛化的。简而言之，多元需求额顶系统的认知灵活性机制可以作为本研究问题提出的支撑，而不能作为对研究结果的解释。我们已在修改稿中删除了相关内容，着重从神经系统组构性的编码机制角度进行解释与讨论，并在本研究的局限性部分进行了补充，补充的内容为：“此外，本研究使用了多个情感与认知任务，并非是在一次任务中进行了多个情感与认知过程，因此，本研究并不能说明多元需求额顶系统的泛化发育能够更好的支持认知灵活切换。”（1148 行-1149 行）

.....

**审稿人 2 意见:**

在这项关于“多种情感认知任务驱动下的可泛化神经表征模式”的研究中，虽有其理论

和方法上的创新，但亦存在不少潜在问题与局限。首先，论文在引言部分对于某些神经模块在发育过程中支持特定情感或认知功能成熟的假设支撑不足，实验数据主要支持了神经表征信息整合的后半部分假设。其次，关于被试的筛选，剔除了大量儿童被试（只保留了 33%），可能影响了样本的代表性和实验结果的普适性，且未详细报告筛选前后被试的任务绩效差异。对于数据处理和统计分析部分，表述不够清晰，写作上也存在不规范之处，比如对跨年龄组比较的描述不清，以及图表内容的重复，这些问题都可能影响读者对研究结果的理解和评价。总体而言，本研究具有一定的创新点，但在理论支持、方法论、结果解释及写作规范性方面还有改进的空间。以下是具体建议和意见。

**回应：**感谢审稿人针对我们研究提出的积极和客观的专业性评价，很高兴审稿人认可我们的研究在理论和方法上的创新。根据审稿人的评论，我们对文章进行了认真仔细的修改，以提升本研究在理论支持、方法论、结果解释及写作规范性方面的水平。

**意见 1：**前言部分最后一段。“我们推测某些神经模块既可以在发育进程中支持特定情感或认知功能的成熟，又可以在差异化的神经表征共激活模式下，通过组结构性编码机制对不同层级情感或认知功能相应的神经表征信息进行分离或整合，从而支持不同情感与认知功能的成熟”。文中的实验结果主要支撑后半部分假设，对于前半部分“推测某些神经模块既可以在发育进程中支持特定情感或认知功能的成熟”支撑不够充分。

**回应：**感谢审稿人指出我们的实验结果主要支撑后半部分假设而缺乏支持前半部分假设的问题，本研究的最大创新点是基于交互式特异化理论提出了脑智发育过程中除了神经特异化过程，还存在着神经泛化的过程，以支持着多情感与认知功能的发展成熟，拓展了脑智发育理论的前沿，而本研究的重点主要是针对神经泛化的探索。对于审稿人指出的问题，主要原因是在本研究开展之前，我们梳理了领域内的文献，发现已经有许多的研究探索了关于神经特异化的脑发育机制，并且领域内已经达成共识，认为脑智发育过程中是存在神经特异化过程的。例如，一些皮层区域选择性地对视觉呈现的单个类别做出响应(Kanwisher, 2010)，梭状回面孔识别区只对面部图片刺激做出选择性响应(Kanwisher et al., 1997; McCarthy et al., 1997)，旁海马空间加工区只对地点场所图片刺激做出选择性响应(Epstein & Kanwisher, 1998)，以及纹体区（extrastriate body area, EBA）只对身体和身体部位图片刺激做出选择性响应(Downing et al., 2001)，此外，腹侧颞枕皮层（ventrotemporal occipital cortex, VTOC）中也发现了特异于视觉数字识别的特异化(Abboud et al., 2015; Hannagan et al., 2015; Shum et al., 2013; Yeo et al., 2017)。基于交互式特异化理论的观点，神经模块既可以在发育进程中支

持特定认知功能的成熟，又可以在多神经生物模式下共存以支持不同认知功能的成熟，前半句关于神经特异化已有文献支持，因此，我们将研究的重点放在了较少研究支持的后半句，也就是关于脑智发育过程中的神经泛化过程。

针对审稿人指出的问题，我们将前言部分最后一段进行了修改，以减轻可能造成的问题。修改后的内容为：“综合多元需求系统和交互式特异化理论构想(Johnson, 2011)，已有研究已经充分证明特定神经系统可以在发育进程中支持特定情感或认知功能的成熟，我们推测，还存在着一种特殊的神经系统，可以在差异化的神经表征共激活模式下，通过组构性编码机制对不同情感或认知功能相应的神经表征信息进行组织，从而支持不同情感与认知领域的成熟。也就是说，儿童大脑发育过程中，除神经特异化进程外，可能还存在着另外一种潜在的支持多情感与认知发展的脑发育进程。”（860 行-865 行）

**意见 2：**第 2.1 节，本实验共招募 548 位儿童，104 名成年人，最终入组 89 名成人和 182 名儿童。只保留了 33% 的儿童被试，剔除比例是否太高，未被筛除的被试是否有某种认知或者情感上的特质影响实验结果？需要报告筛除被试和未被筛除被试之间的任务绩效差异。

**回应：**感谢审稿人提出被试剔除比例较高的问题，我们非常理解审稿人对排除标准较为严格的担忧，想借此机会澄清我们对采用相对严格标准的考虑。与成人研究不同，18 岁以下的儿科神经影像学检查通常极具挑战性，我们目前的研究以及国际上类似的儿童研究均会遇到这样的问题(Zhang, 2019)，因为扫描过程中头部运动过多以及扫描任务意外中止。特别是，头部运动对 fMRI 数据收集的质量和后续数据分析的可靠性有着至关重要的影响。为了确保数据完整性可靠性，我们选择了相对严格的纳入标准。为了让审稿人和读者更清楚地了解数据排除程序，我们可视化了被试筛选流程图（如图 R1 所示）。

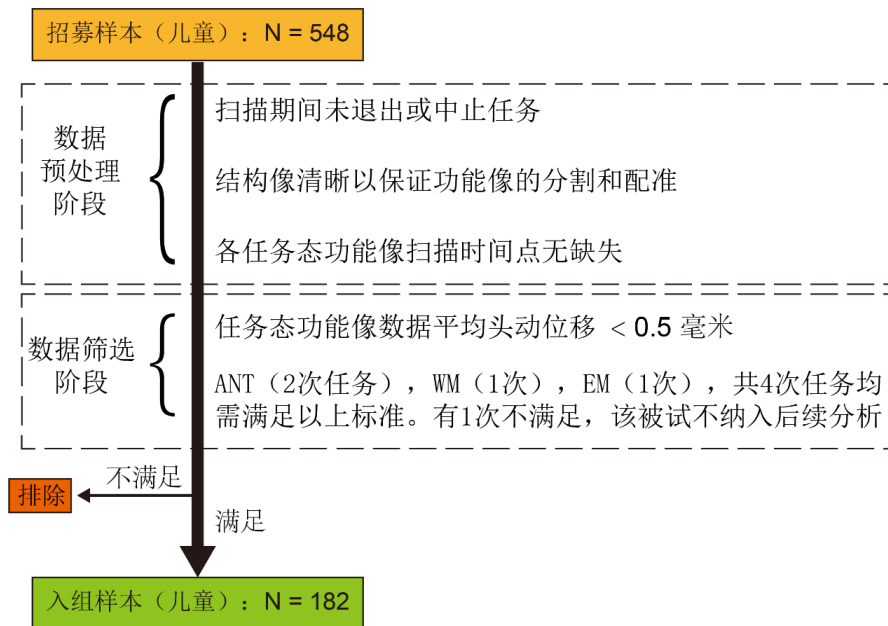


图 R1: 儿童样本被试筛选流程图 (成人样本执行相同流程)

为了解决审稿人对过于严格排除标准的担忧, 分别针对三个单独的任务, 我们对比了筛选被试 (未入组样本, 但需有完整的行为数据) 和未被筛选被试 (入组样本, N=182) 之间的任务绩效差异。在三个任务中, 均未发现显著的组间差异, 统计结果见表 R1。

表 R1: 筛选被试和未被筛选被试之间的任务绩效差异对比

任务	条件	样本量	t 值	p 值
ANT1	警觉分数	182/267	1.320	0.188
	定向分数	182/267	1.005	0.316
ANT2	警觉分数	182/252	1.646	0.101
	定向分数	182/252	1.363	0.174
WM	1-Back 正确率	182/238	0.351	0.726
	2-Back 正确率	182/238	1.622	0.106
EM	情绪正确率	182/221	0.125	0.901
	控制正确率	182/221	0.331	0.741

儿童大样本研究: Zhang, Y., Padmanabhan, A., Gross, J. J., & Menon, V. (2019). Development of human emotion circuits investigated using a big-data analytic approach: stability, reliability, and robustness. *Journal of Neuroscience*, 39(36), 7155-7172.

**意见 3:** 第 2.5 节, 最后一段。用 $\beta$  值衡量“不同过程之间共享的空间模式”可泛化水平是否合适? 原文中, 神经相异性 ( $Y$ ) =  $\Sigma$  (模型相异性 \*  $\beta$ ) + C。其中“ $Y$  是通过 1 减去不同情感认知域下子条件相关的空间激活模式向量两两配对之间的皮尔逊相关系数 ( $1-r$ )”。当  $Y$  值较低时, 意味着被试在做不同任务时具有“共享的空间模式”, 但如果  $Y$  值普遍较高, 是否意味着不同过程之间不存在“共享的空间模式”?

**回应:** 对于用 $\beta$  值衡量可泛化水平的问题, 我们使用的方法参考了 Kragel 博士等人 2018 年发表于 Nature Neuroscience 的研究 (Kragel et al., 2018), 该研究介绍了使用层级化建模的方式探索内侧额叶皮层在不同认知任务进程下的潜在神经泛化机制, 我们通过对该方法进行变式以适应我们的实验数据, 探索了儿童跨情感与认知发展的神经泛化发育基础。Kragel 对该研究方法进行了详细的介绍, 其中, 关于使用 $\beta$  值衡量可泛化水平的描述为: “Parameter estimates ( $\beta$ ) from the model provide estimates of generalizability, with the interpretation of a significant (nonzero) parameter estimate depending on the nature of the regressor. ... Thus, we refer to these parameter estimates as generalization indices, as they reflect the extent to which patterns of brain activation generalize across subdomains, studies, or subjects.”

此外, 根据模型构建的原理图示 (图 4), 关于  $Y$  值, 我们使用的是表征相异性的方法进行计算的, 衡量的是个体与个体之间在各个条件下实际大脑激活模式的相异性, 代表的是观测到的真实值。因此, 可以肯定的是其值越低, 就意味着被试在做不同任务时神经激活的模式越相似, 但如果  $Y$  值较高, 就意味着被试在做不同任务时神经激活的模式越不相似。关于  $\beta$  值, 代表的是理论模型的估计值, 对于“共享的空间模式”是否存在以及程度如何, 则需要通过 $\beta$  值来估计。

**意见 4:** 讨论部分, 第一段。阐述了儿童居于更弱的跨情感认知域的神经泛化程度, 但是对于这一现象没有提供可能的解释, 建议补充。

**回应:** 感谢审稿人提出本研究对儿童更低神经泛化水平的可能解释, 根据研究结果, 我们对该讨论进行了补充。补充的内容为: “多元需求额顶系统共同参与着多种情感与认知功能过程(Cole & Schneider, 2007; Dosenbach et al., 2006; Duncan & Owen, 2000; Fedorenko et al., 2013; Hugdahl et al., 2015), 本研究的结果在注意、工作记忆和情绪任务中印证了这一现象。只有成人在多元需求额顶系统中, 特别是顶内沟皮层区域, 表现出了显著的跨情感与认知领域的神经泛化水平, 此外, 三组被试在多元需求额顶系统中也表现出了显著的发育性差异, 说明了儿童该脑区的组构性信息编码加工机制还未成熟, 在支持不同情感与认知任务时未能

执行稳定的差异性共激活模式(Reverberi et al., 2012), 其表现就是虽然在各过程均被激活, 但是却未能对不同层级情感或认知功能相应的神经表征信息进行有效的信息组织。”(1106行-1114行)

**意见 5:** 讨论部分, 第二段。“共同的神经激活皮层区域——多元需求额顶系统, 包含双侧的顶内沟、额眼区域、腹侧额叶皮层和前辅助运动区均出现了跨情感认知域的神经共激活模式, 说明了该系统对多种情感认知任务的支持作用。”多元需求额顶系统有多个脑区都出现了跨任务的共激活, 但是只有 IPS 和 FEF 存在儿童组和成人组的可泛化水平差异, 对于这一结果缺乏进一步的讨论。

**回应:** 感谢审稿人提出缺乏对该结果的讨论, 我们对此内容进行了补充并修改了讨论第二段的上述声明。修改的内容为: “我们发现了多元需求额顶系统(双侧顶内沟和额眼区域)在三组被试中均出现了跨情感与认知领域的神经共激活模式, 另外, 右侧腹侧额叶皮层(ventral frontal cortex, VFC)和左侧前辅助运动区(pre-supplementary motor area, pSMA)也表现出了部分组别的共激活现象, 但未在所有组别出现。后续研究发现, 多元需求额顶系统 IPS 和 FEF 存在三个组别的可泛化水平差异, 说明了 IPS 和 FEF 存在差异性共激活现象, 通过神经泛化作用支持着多种情感与认知功能的发展成熟。虽然腹侧额叶皮层和前辅助运动区也部分存在共激活现象, 但是一个脑区在多个任务下被激活并不代表该脑区存在神经泛化, 还可能存在其他的神经支持机制, 对于这些机制的探索为我们未来的研究提供了新的思路和方向。”(1086行-1093行)

**意见 6:** 方法部分的写作累赘。例如“相关的对比参数估计图像最初是在个体水平上生成的…随后进行了组水平激活分析。”同样的语句重复了三遍, 请考虑重新组织语言, 更简洁地描述。

**回应:** 我们对该部分内容进行了重新组织, 以保证更简洁地描述该方法。修改后的内容为: “对于 ANT, 将条件(无提示、双侧提示、中心提示、空间提示)建模为独立的事件相关回归因子, 包括警觉(双侧提示 v.s. 无提示)和定向(空间提示 v.s. 中心提示)条件; 对于 WM 任务, 将两个条件(1-Back 和 2-Back)建模为单独的回归变量; 对于 EM 任务, 将情绪和方向识别控制条件建模为单独的回归变量。对于以上任务各条件, 相关的对比参数估计图像最初是在个体水平上生成的, 为了保证后续激活和建模分析中各年龄组人数的平衡, 我们将儿童分为两个组, 分别是儿童低年龄组(7至9岁, N=91), 儿童高年龄组(9至12



岁, N=91), 以及成人组 (19 至 24 岁, N=89), 随后按照分组进行了组水平激活分析, 用于寻找不同情感与认知领域下多元需求额顶系统的神经共激活模式和神经泛化水平的计算。”(964 行-972 行)

**意见 7:** 方法部分有表述不清楚的地方, 例如“自动获取了几组生理的回归序列数据, 以进行基于成分的干扰信号噪声校正(Behzadi et al., 2007)。”自动获取了哪些生理量应该明确。

“在对进行过预处理的图像(使用截止时间为 128s 的离散余弦滤波器)的时间成分和解剖成分进行高通滤波后, 进行了主成分的估计。对于每次噪声校正成分分解, 保留了具有最大奇异值的 k 个分量, 因此保留成分的时间序列足以解释干扰变量(脑脊液、白质等) 50% 的变异。”该部分好像在描述如何提取噪声成分, 但是描述含糊, 什么是“时间成分和解剖成分”? 每次噪声校正意味着做了多次吗? K 个分量意味着每被试都不同? 保留成分还是噪声成分? 干扰变量(脑脊液、白质)是如何定义的? “超出 0.5 毫米 FD 值或 1.5 倍标准化 DVARS 阈值的图像被标注为运动异常值 …”作者对这些异常值做了什么处理? 是插值还是进行回归? 还是去掉?

**回应:** 感谢审稿人对影像数据预处理部分提出的问题, 由于儿童身心发育还未成熟, 对自身的控制能力较弱, 尤其是扫描期间的头动控制, 因此, 影像数据预处理非常棘手的问题之一就是如何减轻头动对数据质量的影响。针对此问题, 我们的研究使用的是最新的“fmriprep”工具包进行的预处理流程, 以尽量减少混杂变量(例如头动及其派生变量等)的影响。该工具包由斯坦福大学的 Esteban 博士开发, 发表于国际权威期刊 Nature Methods, 具有一定的参考性。

针对审稿人提出的问题, 我们通过认真仔细的阅读相关文献, 对表述不清的地方进行了补充。针对第一个问题, 改进内容为: “另外, 自动获取了一系列生理噪声回归量, 包括时间(tCompCor)和解剖(aCompCor)成分, 时间成分是皮层下区域 Mask 中前 5% 的可变体素计算出的六个成分。解剖成分是在将上述 Mask 与在 T1w 空间中计算的 CSF 和 WM 的交集内计算的六个成分, 用以进行基于成分的干扰信号噪声校正(Behzadi et al., 2007)。”(926 行-930 行)。针对第二个问题, 改进内容为: “该校正过程是对进行过预处理的图像(使用截止时间为 128s 的离散余弦滤波器)的时间成分和解剖成分进行高通滤波后, 进行了主成分的估计。对于每个 CompCor 分解, 保留了具有最大奇异值的 k 个分量(算法自动计算, 每个被试会有所差异), 使得噪声成分的时间序列足以解释干扰 Mask(脑脊液、白质等)中 50% 的变异, 干扰 Mask(脑脊液、白质)是基于解剖像进行自动化分割得到的。”(930 行

-934 行)。针对第三个问题,改进内容为:“超出 0.5 毫米 FD 值或 1.5 倍标准化 DVARS 阈值的图像被标注为运动异常值,随后通过组合相关变换(包括头部运动变换矩阵、可用的敏感度失真校正以及与解剖和输出空间的联合配准)进行插值。”(934 行-936 行)。

**意见 8:** 控制分析需要做论证: “三个用于控制分析的脑区,包括双侧的前辅助运动区( pre-supplementary motor area,pSMA)、背外侧前额叶皮层( dorsolateral prefrontal cortex, DLPFC)和前部脑岛( anterior insula,AI)(图 2)”。PreSMA 看起来也在多元需求网络当中,为什么被当作控制分析的脑区呢?请说明理由。对 DLPFC 的 AI 的选择也应作说明。

感谢审稿人提出关于控制分析 ROI 的问题。首先,对于前辅助运动区而言,由于我们是在“domain general”这一术语下载的大脑激活图提取强度最高的前 50%的体素得到的,在原始激活图上,这一区域包含了前辅助运动区和背侧前扣带回( dorsal anterior cingulate cortex, dACC),而实际上最终经过阈值得到的 ROI 更多体素是位于 dACC 脑区的,尤其是在左侧。

**回应:** 由于我们的失误,将此脑区命名为前辅助运动区,因此在修改稿中我们将其更正为 dACC。最终控制分析的脑区包括 dACC、DLPFC 和 AI,这三个脑区分别是特异于注意、工作记忆和情绪的非常具有代表性的脑区,因此作为本研究的控制分析 ROI。为解决该问题,我们对这部分内容进行了修正,修正后的内容为:“然后,在此基础上我们只取强度最高的前 50%的体素以过滤掉低相关的区域,最终选定了多元需求额顶系统的代表性皮层区域,包括双侧的顶内沟( intraparietal sulcus, IPS)和双侧的额眼区域( frontal eye field, FEF),此外,我们还选取了支持注意、工作记忆和情绪功能的代表性脑区作为本研究的控制分析 ROI,包括双侧的侧前扣带回( dorsal anterior cingulate cortex, dACC)(Petersen & Posner, 2012)、背外侧前额叶皮层( dorsolateral prefrontal cortex, DLPFC)(Barbey et al., 2013)和前部脑岛( anterior insula, AI)(Uddin et al., 2017)。”(983 行-989 行)

**意见 9:** 写作有表达不规范的地方:如“为了获取跨年龄组全脑的同质性,我们将儿童分为两个组,分别是儿童中期低年龄组( 7 至 9 岁, N=91),儿童晚期高年龄组( 9 至 12 岁, N=91)以及所有成年人( 19 至 24 岁, N=89)归为一组。”首先,为了获取跨年龄组全脑的同质性表意不明,其次,“以及所有成年。。”和前面两句不搭配。

**回应:** 感谢审稿人指出的问题,我们在修改稿中对这部分内容进行了改进,改进后的内容为:“对于以上任务各条件,相关的对比参数估计图像最初是在个体水平上生成的,为了保证后续激活和建模分析中各年龄组人数的平衡,我们将儿童分为两个组,分别是儿童低年龄组(7

至 9 岁, N=91), 儿童高年龄组 (9 至 12 岁, N=91), 以及成人组 (19 至 24 岁, N=89), 随后按照分组进行了组水平激活分析, 用于寻找不同情感与认知领域下多元需求额顶系统的神经共激活模式和神经泛化水平的计算。”(967 行-972 行)

**意见 10:** 图 3 和图 1 表达的信息有较大的重复, 建议重新组织。

**回应:** 图 3 的内容确实与图 1 有重复, 因此我们将图 3 进行了删除, 并改进了这部分内容的文字描述以及对应到了图 4 中, 更新后的内容为: “该模型被组织为三个层级结构(图 4 左), 包括不同的情感/认知条件作为第一层, 跨情感/认知任务作为第二层, 以及可泛化的跨情感与认知领域作为最高的第三层(测量的是跨不同情感与认知领域的多体素激活模式泛化程度)。”(999 行-1001 行)

**意见 11:** 如何定义认知子域可能是本研究比较重要的一个问题。把 1back 和 2back 定义为可以独立为认知子域需要商榷。首先 1back 和 2back 在加工过程方面基本是一致的, 只不过 2back 比 1back 需要更多的工作记忆存储空间。将他们并列为认知子域可能欠妥当。

**回应:** 感谢审稿人提出认知子域等专业词汇在本研究中的定义和使用问题, 1Back 和 2Back 在加工过程方面基本是一致的, 定义为不同的认知子域确实欠妥。使用跨认知域、认知子域等概念的初衷是为了强调层级上的关系, 但由于本研究使用的多种情感与认知任务难以严格的定义为认知子域, 因此, 我们决定结合审稿人 2 和审稿人 3 的意见以及论文的主题, 将“认知域”统一更改为“情感与认知领域”, 将“认知子域”统一更改为“情感/认知任务”。

.....

**审稿人 3 意见:**

以往的儿童发展研究更多关注脑功能的特异性, 该研究认为脑功能的可泛化性也可能是脑智发育的一种重要机制。该研究通过层级化神经表征建模的方法考察了多元需求额顶系统在注意网络测试 (ANT)、工作记忆任务 (WM) 和情绪匹配任务 (EM) 中神经泛化水平的年龄效应, 结果发现低年龄儿童、高年龄儿童和成年人的神经泛化水平依次升高。对比来说, 该现象在另一组控制脑区 (双侧前辅助运动区、背外侧前额叶皮层、前部脑岛) 上并没有出现。据此, 作者认为, 多元需求额顶系统可能以结构性的信息组织方式来支撑儿童认知和情感功能的发展。这一结果有助于我们进一步理解儿童脑智发育的神经机制、寻找脑智发育的神经标记。总体来说, 该研究在研究问题上具有创新性, 在研究方法上具有前沿性, 也让我

学习到了很多。这篇文章整体上写得比较规范，但仍然有些重要的细节有待完善，在一些具体问题上我仍然有些疑惑和担心，在它们被较好地解决后，我才能够建议发表。

**回应：**感谢审稿人认为本论文在研究问题上的创新性和研究方法上的前沿性，能够感受到审稿人花费了大量时间与精力仔细地阅读了我们的论文，提出的修改意见准确的抓住了本研究的问题与局限，修改完善后将显著提升本论文的专业性。根据审稿人的评论，我们对论文进行了认真仔细的修改和完善。

### 自检报告

**意见 1：**没有做好匿名。自检报告中第 9 点，数据链接中显示了作者的账户 drhaol，代码里的注释显示了作者的信息和邮箱，“2.1 被试”里的“具体任务细节请参考(Hao et al., 2021)”，基本可以确认作者的身份。作为审稿人，我也没必要假装没看见这些。作者发表过一些列高水平期刊，这非常值得赞赏，这很大程度上增加了我对本项研究中方法部分的信心。只是从匿名审稿的角度来说，这或许是一个小失误。顺便说一下，三个实验任务，在本文中应该有足够介绍，以使得本文本身能够独立成篇（stand on its own）。比如，这三个任务具体要求被试做什么，有哪些实验条件，记录什么数据等。在介绍完这些主要的信息后，作者当然可以再引用(Hao et al., 2021)以给读者更多信息，只是考虑到这个引用会泄漏作者信息，可以在审稿阶段对该引用匿名化处理，等正式发表时再补上。

**回应：**感谢审稿人提出的匿名化问题，由于我们考虑欠周，导致了在自检报告和正文中出现了作者信息暴露。我们根据审稿人的建议，重新进行了匿名化处理，造成的问题已然发生，请编辑和审稿人接受我们的歉意。首先，我们删除了数据链接中分析代码的相关作者信息和邮箱。其次，关于三个实验任务，我们在修改稿中增加了实验范式的图示化（图 2）和流程描述（940 行-960 行），以帮助读者了解具体的情感与认知任务信息，同时删除了“具体任务细节请参考(Hao et al., 2021)”内容，如果有幸能够发表，我们再进行补充。

**意见 2：**自检报告中的第 2 个问题：“作者已经投稿或发表的文章中是否采用了与本研究相同的数据？如果是，请把文章附上审查。”作者回答的是“否”。不过延续我在上一条意见中的理解，作者目前使用的数据应该也在 Hao et al., (2021)中（部分）被使用了吧？如果确实没有使用，请更明确地说清楚，使用了相同的实验程序但招募了完全不同的被试。

**回应：**本研究中使用的数据来自一个正在持续进行数据采集的大型脑智发育项目，研究问题也是有多系列，被试样本量随时间的推移是在持续更新的，截止本研究开展数据分析时已

采集 548 名儿童被试，截止前一研究（Hao et al., (2021)）分析时采集的儿童被试数量为 250 名，当前研究使用的数据包含了部分前一研究中的样本，因此两个研究的数据并不完全相同，此外，研究问题也并不属于同一系列的相关研究。根据审稿人的建议，我们将在自检报告重新声明这一情况。（650 行-656 行）

## 摘要

**意见 3:** 在摘要中，作者写道：“结果表明：在学龄儿童期，多元需求额顶系统（包括顶内沟和额眼区）的活动共同参与本研究使用的情感认知任务，即跨任务共同激活……”。这是对图 5A 中红色区域的结果吗？我没理解错的话，这个结果实际上是三组人（低年龄儿童、高年龄儿童、成人）合并（平均）的结果，对吗？如果是这样，那么摘要中这句话中的“在学龄儿童期”可能不太准确，因为理论上存在一个可能性：图 5A 的结果实际上是由成人带来的，儿童没有这个现象。建议图 5A 的结果也分三组人呈现。

**回应:** 图 5A 是三组人（低年龄儿童、高年龄儿童、成人）合并（平均）的结果，实际上我们也考虑过审稿人提出的图 5A 结果可能是由成人带来问题，所以才在后续分析中选择了 NeuroSynth 平台进行独立 ROI 的定义，以避免双重浸渍（double dipping）可能造成的结果偏向。摘要的写作问题给审稿人带来了误解，我们进行了修正，并根据审稿人的建议重新绘制了图 5，分别呈现了三组被试的多任务共激活结果，以及改进了相关的文字描述。摘要中改进的文字内容为：“结果表明：本研究中的三组（儿童低年龄组、儿童高年龄组和成人组）被试均表现出了多元需求额顶系统（包括顶内沟和额眼区域）共同参与多情感与认知任务的现象，即跨领域共激活；值得强调的是，学龄儿童多元需求额顶系统表现出了更低水平的跨领域神经表征可泛化性（generalizability），而控制分析的脑区则没有表现出组间的可泛化性差异。”（750 行-754 行）。正文中改进的文字内容为：“在三组被试中，我们发现双侧顶内沟、双侧额眼区域均出现了跨情感与认知领域的神经共激活模式，另外，右侧腹侧额叶皮层（ventral frontal cortex, VFC）和左侧前辅助运动区（pre-supplementary motor area, pSMA）也表现出了部分组别的共激活现象，但未在所有组别出现，… …”（1042 行-1045 行）

**意见 4:** 在摘要中，作者写道：“……而其它脑区则没有表现可泛化性差异。”这句话，作者实际想说的是控制脑区（双侧前辅助运动区、背外侧前额叶皮层、前部脑岛）没有表现可泛化性的年龄差异。但只读摘要的话，我会误以为除了多元需求额顶系统以外的其它脑区都没有这个现象。

回应：感谢审稿人指出，我们改进了文字表述，以避免可能造成的误解。改进的内容为：“……，而控制分析的脑区则没有表现出组间的可泛化性差异。”（753 行-754 行）

前言

意见 5：文章没有页码。我曾经跟自己的学生说我要确立一个原则，再看到没有页码的文章我会直接拒稿。这次考虑到这项研究的质量和价值，我还是选择打破这个原则，只是没想到这个原则打破得这么快。建议作者在符合期刊要求的前提下，再加上行号，这样便于审稿人评论时引用具体位置。

回应：感谢审稿人指出的页码问题，由于我们的疏忽，确实为编辑和审稿人引用具体位置带来了不便，请接受我们的歉意。文档现已增加了行号，并且我们也会在以后的投稿中形成增加行号的习惯。

意见 6：术语没有统一，而且难懂、容易有误解。比如，请问文中的“情感认知”与“情感和认知”是一个意思吗？具体来说，认知就是注意网络测试（ANT）和工作记忆任务（WM），情感就是情绪匹配任务（EM），是这样的吗？因为“情感认知”还可以被理解为“对情感的认知加工”，这就可能造成误解了。从概念上讲，情感是否也是一种认知？如果不是，请作者最好能在文中做好概念辨析，并且可以考虑统一表达为“情感和认知”，如果在中文句子框架下要省略“和”字，或许表达为“认知情感”会减少我说的这个误解。如果作者同意情感也是一种认知，那么或许整篇文章都用“认知”这个概念，是否会更清楚？此外，作者在图 1 的标题上写的是“跨情感认知域……”，但在该图的最上层缺写的是“跨认知域”，类似的情况在图 3 中也存在，看起来作者是同意情感也是一种认知的？请辨析。

回应：感谢审稿人提出由于术语统一不当造成的难懂和容易误解问题，本研的核心是检查不同情感与认知过程的神经泛化模式，关注的重点在于“不同…过程”，基于本研究使用的方法，不管是情感还是认知，都是用于评估泛化水平的不同过程，是平行的关系。结合审稿人的评论，我们针对修改如下：（1）“情感认知”与“情感与认知”在本研究的文字表述中是同一个意思，我们已经在全文统一使用“情感与认知”。（2）审稿人的理解完全正确，本研究中的认知是通过注意网络测试（ANT）和工作记忆任务（WM）获取的，情感是通过情绪匹配任务（EM）获取的。（3）我们在前言中增加了情感与认知的概念辨析，具体内容为：“多元需求系统（Multiple demand system）是一个以不同神经表征模式参与多种情感与认知任务的共激活神经系统，认知方面包括感知觉、注意、工作记忆和执行控制等过程，情感方

面包括情绪识别、情绪调节等过程(Adolphs, 2002; Camilleri et al., 2018; Duncan, 2010; Duncan & Owen, 2000; Sugiura et al., 2023)。”(813 行-816 行)。(4) 针对标题和图 3 中文字表述问题, 由于我们的失误, 未在提交前针对文字同步更新, 现已在修改稿中修正该问题。

**意见 7:** 另外, 我注意到作者在前言部分更偏重于认知部分的文献回顾, 如“新认知功能的出现取决于新大脑皮层区域的成熟……因此, 新情感认知功能的出现并不一定与被认为之前不成熟的“沉默”神经区域成熟时变得活跃有关, 或是基于发展认知神经科学的理论和实证研究进展……交互式特异化理论……用以指导情感认知发展和脑发育基础研究”, 并没有提及或举例说明情感部分, 那么请问作者把认知和情感放在一起研究的理由和目的是什么? Hao et al., (2021)还有奖赏任务, 为什么在这篇文中中没有纳入进来? 以及, 为什么选用这几个任务来考察认知和情感, 这在前言部分可能需要有足够的铺垫。

**回应:** 将认知和情感放在一起研究理由和目的, 与意见 6 的回复相类似, 本研究的核心是检查不同情感与认知过程的神经泛化模式, 关注的重点在于“不同…过程”, 如果条件允许的话, 我们希望能够纳入更多的情感与认知过程来验证本研究的假设, 以及探索实验结论是否普适于更多的情感与认知领域, 这也是本研究的局限性和未来更深入研究的方向, 我们会在修改稿的局限性部分明确该问题的存在, 内容为:“本研究仅采用了三种情感与认知领域的六个过程, 未来的研究应该纳入更多的情感与认知条件来验证本研究的假设, 探索实验结论是否普适于更多的领域。”(1144 行-1145 行), 并且在前言部分增加关于情感部分的文献回顾, 内容为:“新情感与认知功能的出现, 并不一定是由于神经发育过程中原本不成熟的“沉默”区域发展为成熟的“活跃”区域而形成的。例如, 婴儿在出生后的几个月就能够对情绪的表达进行感知, 左右脑均会变得活跃, 但是针对不同情绪则会发展出不同的左右脑偏向神经反应模式(Missana & Grossmann, 2015)。”(782 行-786 行)。“例如, 杏仁核与前额叶皮质作为情感过程的关键性神经环路, 在神经发育过程中会出现网络连接模式由正向负的“发展性转变”(Gee et al., 2013)。”(805 行-806 行)。关于 BART 奖赏任务未纳入本研究的原因, 是由于该任务的特殊设计导致的, 每个被试的冒险倾向程度是不同的, 每次打气球的次数也就不同, 因此会造成每个条件最终的试次数量不同, 并不像 ANT、WM 和 EM 三个任务试次是固定的, 纳入统一的模型中可能会导致条件数量权重不平衡, 所以未纳入本研究当中。最后, 选用这几个任务来考察不同情感与认知的原因, 是由于本实验所属的科研项目就只采集了这几个任务, 结合意见 6 中的问题, 我们通过梳理文献总结了多元需求系统参与的情感与认知过程, 发现在认知方面主要包括感知觉、注意、工作记忆和执行控制等过程, 情感方

面包括情绪识别、情绪调节等过程，因此本研究中选取的各情感与认知任务还是具有一定程度的代表性的，我们在眼前部分对此也进行了铺垫，内容为：“针对多元需求系统参与的情感与认知过程，我们选取了具体代表性的认知领域内的注意和工作记忆任务，以及情绪领域的情绪识别任务对上述问题进行了探索。”（867行-868行）

**意见 8：** 认知子域、跨认知子域、跨认知域，这几个概念我一开始看是有点“不明觉厉”的，看懂后才知道，其实就是认知子领域、跨认知子领域、跨认知领域的含义。从大同行、一般读者的角度，或许表达为“领域”（而不是“域”），虽然会多一个字，但好理解得多。当然，如果这个“研究域”都这么表达，已经成为约定俗称，而且作者也坚持认为这样更好，我也不反对，仅供参考。

**回应：** 感谢审稿人关于认知子域、跨认知子域、跨认知域等概念在本研究中的使用问题，由于这一概念还较少有研究关注，确实可能会给审稿人和读者带来一定程度上的理解负担，结合审稿人 2 和审稿人 3 的意见以及论文的主题，我们在修改稿中将“认知域”统一更改为“情感与认知领域”，将“认知子域”统一更改为“情感/认知任务”，以增加论文的可读性。

**意见 9：** 作者在前言第 3 段提到“交互式特异化理论”，并且提到“功能性脑发育是以脑区或神经系统间的相互作用为基础而发展成熟的”，好像是在说，特定功能的脑区之间的相互作用来实现的通用性。这一段里的“通用性”是 task-general 的意思吗？是一个脑区可以参与多个认知(子)领域加工的意思吗？这里的交互是什么意思？是脑区间的(功能)连接吗？可是本项研究并没有使用功能连接的方法来考察脑区间的“交互”，我不太明白这部分跟本研究之间的关系。另外，我甚至不太明白这个理论的逻辑，即为什么该理论认为多个特定功能的脑区交互就能发挥通用性的功能？“通用性”除了字面意思上的一个脑区可以干不同的任务之外，它在理论上意味着什么？请作者讲得更清楚。

**回应：** 基于交互式特异化理论，“通用性”是指适用于不同的情感与认知领域，在本研究中，至少是通用于注意、工作记忆和情绪识别领域的，或者也可以理解为“一般性”。这种说法我们参考了发展认知神经科学领域的研究者 Mark H. Johnson 的观点，之所以将交互式特异化这种广泛性的理论称为“架构 (Frameworks)”或“通用性理论 (General theories)”，因为其来源是基于对大量数据的思考和观点(Johnson, 2011; Morton & Frith, 1995)。以该理论为基础，在其框架内可以基于新的研究发现构建更具体、更详细的理论，这也就为本研究的开展提供了理论基础。当然，本研究的发现是基于三个情感与认知领域得来的，并没有基于大



量的不同领域数据，从被试数量上倒可以算作大样本研究，后续需要增加来自不同领域的更多任务，验证当前研究的实验结论是否通用于更多的领域。我们认为本研究的意义是在验证了交互式特异化理论的同时，又在小步子程度上拓展了该理论前沿。但是，不可否认这确实是本研究的局限，我们会在局限性部分对此进行声明（1144 行-1145 行），也会调整第 3 段的相关表述以方便审稿人和读者的理解，内容为：“基于发展认知神经科学的理论和实证研究进展，有研究者提出了用于解释人类功能性大脑发育机制的通用性理论——交互式特异化理论（The interactive specialization theory），旨在解释可观测行为发展的功能性大脑发育基础(Johnson, 2001, 2011)。之所以将这种广泛性的观点称为“架构（Frameworks）”或“通用性理论（General theories）”，因为其来源是基于对大量数据的思考和观点总结(Johnson, 2011; Morton & Frith, 1995)。在其框架内可以基于新的研究发现构建更具体、更详细的理论分支或细化模型，例如，可以通过评估来自不同情感与认知领域的脑发育性影像证据，确定在多大程度上将其视为候选领域的通用性发育理论。”（793 行-799 行）

关于“交互”，在 Johnson 看来是理论上的概念，交互的形式可以因为实证研究方法的不同而不同，可以是神经网络连接，也可以是神经的激活模式。Johnson 认为，人类在出生后的脑智发育过程中，一些皮质区域的激活模式会发生变化，它们通过交互和竞争以获得支持新能力的角色，区域之间活动相关的交互增强了皮质区域的功能和响应特性，使得它们的活动被限制在更窄的范围内（例如，最初由各种各样的视觉对象激活的区域，可能会将其响应限制在人脸上）。此外，婴儿期新行为能力的出现与多个区域活动的变化有关，而不仅仅是一个或多个区域活动的出现(Johnson, 2011)。在本研究中，我们试图引入“组构性”这一概念解释共激活的神经表征模式对于多情感与认知领域的支持，即在神经激活模式的方法框架下解释该发育现象，我们也将意见 10 中对“组构性”一词进行详细的解释。“交互”可能会给审稿人和读者造成脑区间连接的误解，我们在修改稿中将使用“协作”一词进行替换，但是“交互式特异化理论”仍然沿用直译的名字。

**意见 10：**前言倒数第二段，“组构性”和“表征泛化”、“泛化性”、“可泛化性”是一个意思吗？作者写道：“在计算水平上，无需训练即可利用已经学习过的完成一项任务的基本过程结构来正确执行新任务，该属性被称为“组构性”，已经被提议作为灵活性认知控制的基础原理(Cole, Laurent, et al., 2013)。”这句话读者好像也是“泛化性”的意思。目前这一段，我只理解到了一个脑区/神经元可能参与一种任务也可能参与多种任务，我不明白这怎么就是“组构性”了。我感觉到背后应该有很重要的理论思考，但我现在还没明白。请作者进一步

解释。

回应：“组构性”是 Reverberi 博士等研究者提出的一种神经表征编码规则，在本研究中，是用于解释“表征泛化”可能的神经元底层机制的。一个神经元在执行多种不同任务时可能被激活也可能未被激活，不同的任务可能被各自“私有”的神经元集群所支持，也可能涉及其相关神经系统中的每个神经元，又或介于这两种模式之间。如果多个任务需要一个共同的基础认知过程，那这不同的任务会使用一部分共享的神经元集群，而组构性 (Compositionality) 这一概念被研究者引入就为共享的神经元集群如何支持多个任务提供了途径 (Reverberi et al., 2012)，该规则随后又在机学习的神经网络模型中得到了印证 (Yang et al., 2019)。如图 R2 所示，将“组构性”规则进行了图示化，在中间的一行中，有两个简单的神经元活动模式编码规则的示例：规则 A 和规则 B。在顶部和底部的两行中，对于规则 A 和规则 B 形成的复合规则，有两种编码策略。第一种策略（顶部）使用的就是组构性编码策略，规则 AB 的编码模式是规则 A 和规则 B 的神经元模式的组合，这种组合方式就会规则 A 和 B 的共激活模式。另一种策略（底部）是使用独立性编码策略，规则 AB 的编码模式与规则 A 和规则 B 的模式无关。在具有组构性编码功能的神经环路中，新任务可以被表示为基本过程表征之和。“在计算水平上，无需训练……原理 (Cole, Laurent, et al., 2013)。”这句话是对于“组构性”这一概念的文献支持，对于本研究结果的解释意义不大，现已删除。

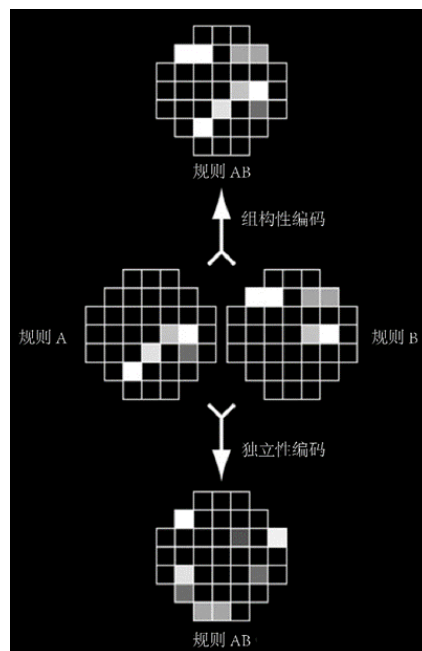


图 R2：神经表征组构性编码机制示意图

我们使用的神经表征相异性分析是基于功能性磁共振成像技术进行的，然而，目前的无创脑

功能成像技术的时空分辨率有限，是无法完全表征精确的神经元激活的，因此我基于当前研究的发现，进行了“组构性”编码机制可能是神经表征泛化的底层机制的理论思考。我们会调整第 3 段的相关表述以方便审稿人和读者的理解。

**意见 11:** “在人类神经影像研究水平上，位于额顶网络的前额叶皮层对许多情感与认知功能都及其重要(Fuster, 2015; Miller & Cohen, 2001; Stuss & Knight, 2013)，其原因在于它在多种任务表征中的核心作用(Cole et al., 2011; Sakai, 2008; Tschentscher et al., 2017; Wallis et al., 2001)，……”在这句话中的“其原因在于”前后是一个意思吧，都在说“重要”/“核心”作用，没有解释重要的原因。

**回应:** 这句话主要想交代额顶皮层参与了许多情感与认知过程，及其神经表征对组构性编码的支持。根据审稿人的建议，我们改进了这句话的表述，内容为：“在人类神经影像研究中，位于额顶皮层的前额叶参与了许多情感与认知功能，不同的神经表征模式可以发挥各种功能，例如通过向大脑的其他部分提供自上而下的指令，引导着认知过程正常进行(Fuster, 2015; Miller & Cohen, 2001; Stuss & Knight, 2013)，它在多种任务表征中还支持着规则的理解和运用等(Cole et al., 2011; Sakai, 2008; Tschentscher et al., 2017; Wallis et al., 2001)，此外，还有外侧前额叶皮层中复杂认知表征对组构性编码的有效支持(Cole et al., 2011; Reverberi et al., 2012)。”(840 行-845 行)

**意见 12:** 图 1 需要有更多的解释。目前孤零零地放在这里，图后面没有注释说明，正文中也没有足够的展开，读者很难知道从哪入手去看懂它。

**回应:** 图 1 是我们结合了多元需求系统、交互式特异化理论和本研究的实验设计提出的构想，试图描绘一个跨情感与认知领域发展的多层级神经表征泛化机制的模型。根据审稿人的建议，我们对其各层级进行了注释说明，内容为：“如图 1 所示，是我们结合了多元需求系统、交互式特异化理论和本研究的实验设计提出的构想，试图描绘一个跨情感与认知领域发展的多层级神经表征泛化机制的模型。”(865 行-867 行)，“注：第一层级对应不同的情感/认知条件，该层级的各条件由神经特异性激活支持；第二层级对应跨情感/认知任务的神经共激活，多元需求系统在神经激活水平上支持着不同任务过程；第三层级对应跨情感与认知领域的神经表征泛化，通过组构性编码机制支持着不同情感与认知功能的成熟。”(871 行-873 行)

方法

意见 13：“本研究共招募了 548 位正常发育的儿童被试……和 104 名健康的成年人……”，“因此纳入分析的被试需要完成所有任务并且满足头动剔除标准，平均头动位移（*framewise displacement, FD*）值小于 0.5 毫米（ANT 任务两次实验均需满足标准），最终满足入组标准的被试包括 89 名成人……和 182 名儿童……。”相当于只保留了约 33%的儿童和 85%的成人，这个缺失比例跟前人文献中的差不多吗？有没有办法检验这个缺失是完全随机的缺失，排除掉“泛化水平的年龄差异可能是因为泛化水平高的人在童年时期头动更大更容易被踢掉”这个替代解释。

回应：与成人研究不同，18 岁以下的儿科神经影像学检查通常极具挑战性，审稿人 2 也提出了类似的担忧，我们目前的研究以及国际上类似的儿童研究均会遇到这样的问题，这个缺失比例与前人文献中类似的儿童影像大样本研究是类似的(Zhang, 2019)，可以看到 Zhang 的研究中 13 岁以上的被试可以较好的控制头动，13 岁以下的被试则剩余样本量较少，这与本研究的年龄范围相。为了保证结果的可靠性，我们使用了较为严格的头动剔除标准，具体流程如“图 R1：儿童样本被试筛选流程图”所示。针对随机性的问题，我们对比了筛除被试（未入组样本，但需有完整的行为数据）和未被筛除被试（入组样本，N=182）之间的任务绩效差异。在三个任务中，均未发现显著的组间差异（如表 R1 所示），这一结果可以在一定程度说明缺失是随机的。

意见 14：“采集矩阵（*acquisition matrix*）为 256×224，视场为 256x224 平方毫米”。×与 x 要统一。

回应：感谢审稿人指出该问题，我们已经针对文字格式、规范性等问题进行了修正，并且重新阅读了全文，以保证其他内容中也没有这些问题。

意见 15：“对于 ANT，将六个条件（即无提示、双侧提示、中心提示、空间提示、方向一致靶刺激和方向不一致靶刺激）建模为六个独立的事件相关回归因子,相关的对比参数估计图像最初是在个体水平上生成的，包括警觉、定向和执行注意条件……。”尽管 Hao et al., (2021)中交代了警觉、定向和执行注意条件怎么计算（*contrast?*），但本文中并没有交代。另外 Hao et al., (2021)也没有交代“即无提示、双侧提示、中心提示、空间提示”具体是怎么做的，要被试做什么，我也没有看到反馈一屏的持续时间，我好像也没有看到一个 run 的试次数量和每个条件的试次数量。这些可能需要作者写清楚。

回应：感谢审稿人指出该问题，我们在修改稿中增加了 ANT、WM 和 EM 三个实验范式的

图示化（图 2）和流程描述（940 行-960 行），以及警觉和定向条件的计算方法（965 行-966 行），以帮助读者了解各情感与认知任务的反应要求及相关信息。

**意见 16：**“随后进行了组水平激活分析。”这是把所有被试合并在一起坐的组分析吗？图 5A 就是组分析的结果，是这样的吗？图 5A 中的重叠脑区，是指这六个条件中都激活了的区域吗？方法部分好像没有足够的交代。

**回应：**感谢审稿人指出我们忽略了激活分析中被试分组交代不清楚的问题，我们针对此问题已经进行了改进，对分组信息在方法部分进行了交代。改进的内容为：“对于以上任务各条件，相关的对比参数估计图像最初是在个体水平上生成的，为了保证后续激活和建模分析中各年龄组人数的平衡，我们将儿童分为两个组，分别是儿童低年龄组（7 至 9 岁，N=91），儿童高年龄组（9 至 12 岁，N=91），以及成人组（19 至 24 岁，N=89），随后按照分组进行了组水平激活分析，用于寻找不同情感与认知领域下多元需求额顶系统的神经共激活模式和神经泛化水平的计算。”（967 行-972 行）

**意见 17：**ANT、WM、EM 任务的事件（regressor）数量不一样（每个事件的条件数量是一样的吗？），Contrast 的数量也不一样，被试数也不一样，这会使得每个 Contrast 的统计检验力也不一样，这对图 5A 的结果有影响吗？这对多层级神经泛化模式分析会有影响吗？会有多大的影响？这可能不是一个大问题，但我想知道作者的观点，如果会，或许在讨论部分增加对这部分局限性的讨论，如果有补充分析来排除这个问题就更好了。

**回应：**我们非常赞同审稿人的担忧，但是要真正做到 regressor 数量一样可能需要重新设计任务。ANT 任务是事件相关设计，WM 和 EM 任务是 Block 设计，在 ANT 中每个提示条件共 48 个试次（两个 run 之和），WM 任务中 1-和 2-Back 各 4 个 Block，EM 任务中情绪和方向各 5 各 Block，总体上的时间大约还是相对一致的。这些对统计检验力和本研究中重要结果的影响确实是一个值得思考的问题，我们不敢妄下结论是否有影响，但是我们认为这确实可以作为后续研究中的关注点，用以加强我们结果的可靠性。我们将根据审稿人的建议，在修改稿中的局限性部分对此问题进行声明，内容为：“由于各情感与认知任务的设计不同，存在 regressor、contrast 数量不一致的情况，这对统计检验力和本研究中重要结果是否存在影响是值得商榷的，需要后续改进实验设计来验证此问题”（1145 行-1147 行）

**意见 18：**图 3，注释和图中的表述不对应。

回应：我们已经根据审稿人 2 的建议对图 3 进行了删除，部分信息已经整合到了图 1 和图 4 中。

意见 19：“我们将儿童分为两个组，分别是儿童中期低年龄组（7 至 9 岁，N=91），儿童晚期高年龄组（9 至 12 岁，N=91）以及所有成年人（19 至 24 岁，N=89）归为一组”。“晚期”读着有点奇怪，并且跟图 3、图 6 的表述也不一致。

回应：感谢审稿人的纠正，我们已经在修改稿中将三组被试统一为了“儿童低年龄组”、“儿童高年龄组”和“成人组”，图片中的相关内容也已同步更新。

意见 20：图 4 右边的三角形。这是本研究的真实数据画出来的吗？相比于 Kragel（2018）中的 Fig.1，这个三角形旋转了 45 度，这倒不是问题。然而，在相应的注释里，作者仍然写道“行代表每个受试者的每个情感认知过程”，旋转之后的“行”，读者可能就很难搞明白了。如果能在图上标出来，可能会更清楚。另外，值得一说的是，作者在文章中对方法的可视化，做得特别好。

回应：感谢审稿人对我们文章中方法可视化的认可。图 4 左侧的三角形是真实数据画出来的，由于被试和各任务条件结合到一起数量庞大，因此每个格子极小。而右侧为模型图示化，是通过欧氏距离计算出的模型相异性，与左侧是对应的，并不是真实数据画出来的。由于三角形旋转了 45 度，用“行”描述确实会让审稿人和读者产生疑惑，我们在图 4 中进行了标记并修改了文字以改进此问题，修改的文字为：“左侧矩阵代表一个脑区内所有受试者在不同情感与认知条件下的神经表征相异性，元素格代表所有受试者的每个情感与认知过程成对匹配的神经相异性，彩色条表示功能相异性层级中的相应级别。”（1017 行-1018 行）

意见 21：“为了检查发育性的变化，我们比较了三组受试者之间的参数估计的差异，每次我们使用独立样本 T 检验从 5000 次迭代中随机选择 100 个值进行比较，并获得 p 值，相应的显著性定义为差异显著结果（我们使用了  $p = 0.001$  的阈值决定是否显著）的总数量除以 1000。”我不理解，为什么组间差异要抽取 100 个值？这 100 个值怎么就获得 p 值了？如果是 t 检验，那只能得到一次显著或者不显著，它除以 1000 又有什么意义？请作者解释。另外，我头脑中设想的方法是：仍然使用 5000 次 bootstrapping 中每一次的组间差值来计算，即差值小于 0 的比率。或许作者已经考虑过了，但我（以及以我为代表的读者）可能很好奇为什么不这么做。

回应：感谢审稿人提出有关结果显著性定义的问题，随机选择 100 个值的原因是由于我们使用的方法在 5000 次迭代中本身就具有一定的随机性，并不是使用真实的测量值进行的 5000 次迭代，每次迭代会对每个情感与认知过程成对匹配的神经相异性进行随机化，因此所产生的参数估计值 ( $\beta$ ) 也有所不同，这就构成了 5000 次结果的样本分布，具体内容可以参考我们上传的代码 (Generalization/AnalyScripts/BootstrapROI\_CCA\_Cond6\_NS.m) 中第 101-120 行。因此，进行组间差异比较时，要随机抽取 100 个值进行 t 检验，这个过程重复 1000 次，因此最终显著性定义要用差异显著结果的总数量除以 1000。由于我们表述中没有交代这个过程要重复 1000 次，因此给审稿人带来了疑问，我们现已对这部分表述进行了改进，以避免给读者也造成类似问题。改进后的内容为：“为了检查发育性的变化，我们比较了三组受试者之间的参数估计的差异，每次我们使用独立样本 T 检验从 5000 次迭代中随机选择 100 个值进行比较并获得 p 值，这个过程会重复 1000 次，因此，相应的显著性定义为差异显著结果（我们使用了  $p = 0.001$  的阈值决定是否显著）的总数量除以 1000。这么做的原因在于我们使用的方法在 5000 次迭代中本身就具有一定的随机性，每次迭代会对每个情感与认知条件成对匹配的神经相异性进行随机化，因此所产生的参数估计值 ( $\beta$ ) 也有所不同，这就构成了 5000 次结果的样本分布因此，进行组间差异比较时，要随机抽取 100 个值进行 T 检验。”（1029 行-1035 行）

对于审稿人提出的使用差值小于 0 的比率检查泛化水平的发育性差异的方法，我们也进行了尝试，其结果和我们当前研究的结果基本上是一致的，如图 R3 所示，儿童在多元需求额顶系统中各个脑区仍然表现出了更低的跨情感与认知领域的神经泛化水平 ( $p \leq 0.05$ )。

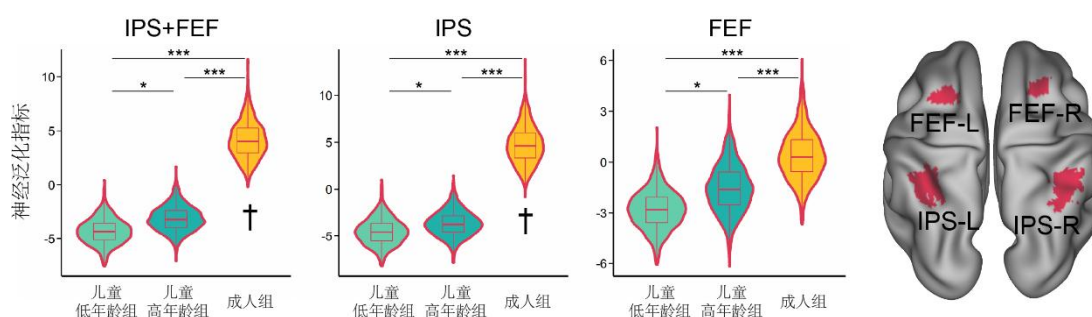


图 R3 跨情感与认知领域的神经泛化水平的发育效应

意见 22：方法 2.4 定义多元需求额顶神经系统部分，文章研究的是多情感认知任务，为什么选择的搜索术语是“domain general”，而不是“cognitive”或“emotion”，或者就是多元需求系统？

回应：我们在 Neurosynth 平台使用 Term-based meta-analyses 进行 ROIs 定义时，实际上首先考虑的就是“multiple demand”，但是未检索出相关的 Map；随后尝试用“multi”检索看是否有多任务、多领域相关的 Map，仍然未检索到；在输入“general”一词检索时，发现了“domain general”的相关 Map，并且其激活的脑区与 multiple-demand system 的关键脑区基本上是相匹配，因此选用了“domain general”的 Map 进行 ROIs 的定义。关于“cognitive”或“emotion”，我们研究中确实使用了这两个领域的任务，但是我们无法确定使用哪一个更合适，或许“cognitive”符合更多的任务领域，但将 emotion 归类于其下，可能会有部分研究者表示不赞同。

意见 23：作者选取双侧的前辅助运动区、背外侧前额叶皮层和前部脑岛作为控制分析的脑区，这三个脑区是怎么选出来的？根据前人模板还是 Neurosynth 术语？为什么是这三个脑区？

回应：在本研究中，多元需求额顶系统的代表性皮层区域（IPS 与 FEF）和注意（dACC）、工作记忆（DLPFC）和情绪（AI）功能的代表性脑区都是在“domain general”的 Map 中提取强度最高的前 50%的体素得到的。

关于前辅助运动区是由于命名失误造成的，在原始激活图上（图 3 左侧），这一区域包含了前辅助运动区和背侧前扣带回（dorsal anterior cingulate cortex, dACC），而实际上最终经过阈值得到的 ROI 更多体素是位于 dACC 脑区的，尤其是在左侧。由于我们的失误，将此脑区命名为前辅助运动区，因此在修改稿中我们将其更正为 dACC。最终控制分析的脑区包括 dACC、DLPFC 和 AI，这三个脑区分别是特异于注意、工作记忆和情绪的非常具有代表性的脑区，因此作为本研究的控制分析 ROI。

我们对这部分内容进行了改进，改进后的内容为：“然后，在此基础上我们只取强度最高的前 50%的体素以过滤掉低相关的区域，最终选定了多元需求额顶系统的代表性皮层区域，包括双侧的顶内沟（intraparietal sulcus, IPS）和双侧的额眼区域（frontal eye field, FEF），此外，我们还选取了支持注意、工作记忆和情绪功能的代表性脑区作为本研究的控制分析 ROI，包括双侧的侧前扣带回（dorsal anterior cingulate cortex, dACC）(Petersen & Posner, 2012)、背外侧前额叶皮层（dorsolateral prefrontal cortex, DLPFC）(Barbey et al., 2013)和前部脑岛（anterior insula, AI）(Uddin et al., 2017)。”（983 行-989 行）

结果



**意见 24:** “3.1 不同情感认知域下多元需求额顶系统的神经共激活”的结果更像是定性的结果，而且三组被试合并在了一起（我没理解错的话），或许三组分别呈现能提供更多的信息？

**回应:** 我们已根据意见 3 针对此问题进行了改进。

**意见 25:** 注意有警觉、定向和执行注意三个条件，为什么只呈现前两个的结果？以及，在“多层次神经泛化模式分析”为什么也只用前两个条件？

**回应:** 只用警觉和定向两个条件的原因是模型构建需要保持每个任务的条件数是一致的，因为 EM 只有情绪和方向两个条件，并且 WM 中的 0-Back 严格意义上讲也不能很好的测量工作记忆，结合以上原因，我们在整个研究中针对每个情感与认知任务就统一使用了两个条件。

**意见 26:** 既然已经在方法 2.4 部分自定义了多元需求额顶系统，在图 5B 里呈现前人文献中多元需求系统脑区分布图的目的是什么？是为了说明它跟图 5A 中的红色区域很像，是吗？这很有帮助，不过逻辑上稍微有些绕。读者可能会问，作者为什么不直接用那项研究的结果作为多元需求额顶系统？

**意见 27:** 我不确定作者是否已经获得原文出版社的授权，我稍微提醒一下。另外，图上有一些绿点、坐标、脑区简称，如果与本文主题无关，或许应该删除。如果作者能找到相应研究的.nii 文件，我更建议作者重新绘制一下，这样既避免了版权问题又更简洁。

**回应:** 如审稿人所说，在图 5B 里呈现前人文献中多元需求系统脑区分布，确实是为了对比图 5A 中的红色区域，用于说明我们发现的多任务共激活脑区分布与前人研究是一致的。使用 Neurosynth 平台的 Term-based meta-analyses 进行 ROIs 定义是因为该平台综合的研究数量更多，客观性方面也会更好，而且我们无法获取相应研究的.nii 文件。结合审稿人意见 3、26 和 27，同时由于我们也未得到原文出版社的授权，因此将图 5B 部分删除，图 5 改为分别呈现三组被试的多任务共激活结果。

**意见 28:** 图 6 的注释中提到了“置换检验”，但我好像搜索全文也都没看到对“置换检验”的方法描述。请作者补充。

**回应:** 感谢审稿人指出该问题，本研究中并未使用置换检验，修改稿中已针对此问题进行了修正。（1066 行）

**意见 29:** 图 6 里面有 IPL 和 FEF，还有 IPL+FEF，为什么还有两者相加的？方法部分好像

没有足够的交代。

回应：图 6 中的 IPL（现已经更正为 IPS）+FEF 是指多元需求系统的核心部分额顶网络，我们分别针对 IPS+FEF、IPS、FEF 进行跨情感与认知领域的神经泛化水平进行测量，想弄清楚是整个额顶网络还是其中的关键脑区存在神经表征泛化现象。从目前的结果来看，是 IPS 的效应更为显著，虽然 FEF 也存在一定程度的神经泛化现象，但未能达到显著性水平，只表现出了高于儿童的发育性差异。根据审稿人的建议，我们在方法部分对此进行了补充和改进，内容为：“对于多元需求额顶系统的代表性皮层区域，本研究试图探索是整个额顶网络还是其中的关键脑区存在神经表征泛化现象，因此分别针对 IPS+FEF、IPS、FEF 均进行了跨情感与认知领域的神经泛化水平进行测量。”（989 行-991 行）

意见 30：本研究的核心结果其实只有“3.2 跨情感认知域神经泛化水平的发育性差异”，这个结果很漂亮，但坦白讲，可能稍显单薄，尽管这个方法涉及的工作量不小。作者有没有可能从其它角度对该结果进行交叉验证、或者补充更多分析？比如，行为数据（反应时、正确率）的描述统计及三个年龄组的 t 检验、脑与行为表现的相关等。

回应：感谢审稿人对我们的研究结果的肯定，如意见 31 中我们的解释，目前还无法使用神经泛化指标进行个体水平的计算，这确实是该方法很大的局限性所在，未来我们也一定会思考如何解决这个问题，这对我们后续更加深入的探索儿童跨情感与认知领域发展的神经系统组织方式是非常关键的。根据审稿人的建议，我们汇报了相关的行为数据（反应时、正确率）与年龄的相关统计（图 R4），由于本研究的方法不能在全脑和 ROI 水平上计算个体的神经表征泛化水平，无法将行为表现与脑进行相关分析，因此对于核心结果的验证补充分析目前还没有思路，如果审稿人能够提供相关思路解决该问题，我们非常愿意进行补充验证。

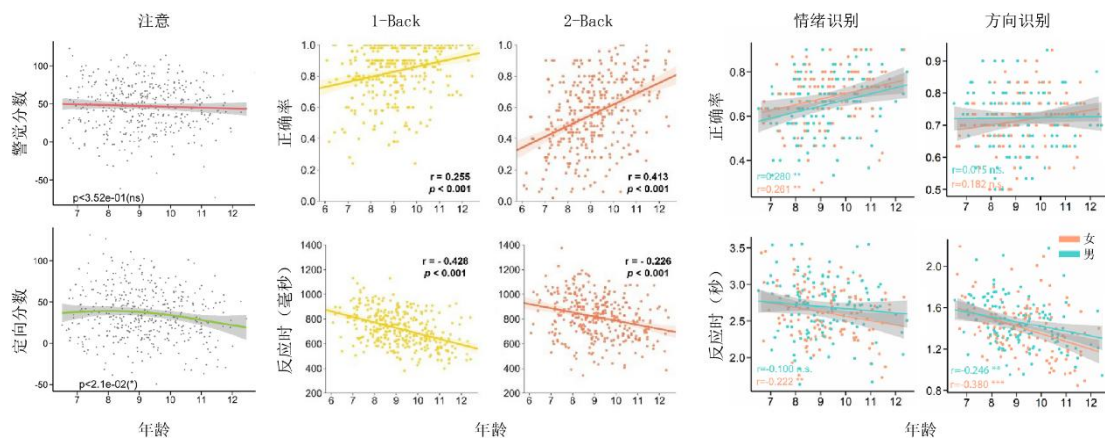


图 R4 行为数据与年龄的相关结果

**意见 31:** 作者在多处提到“异常发育的客观生物标记”。然而，坦白讲，目前基于组水平的结果，尽管通过 bootstrapping 可以得到显著性，但并不能得到每个个体的泛化水平的指标，这限制了它作为客观生物标记的可能性。考虑到本研究中的一名被试是完成了多个任务，这不同于 Kragel (2018) 里的数据是来自于不同研究的不同被试，建议作者尝试在每个个体内构建多层级神经泛化模型，这样可以计算每个被试的泛化性水平，然后将其与被试的月龄(不是年龄；如果可以获得的话)做相关。如果出现显著的正相关，那么是对目前图 6 结果的交叉验证，而且是真的可以作为一个客观生物标记物来使用。

**回应:** 感谢审稿人提出使用神经泛化指标测量个体水平来作为客观生物标记物的问题，实际上我们一开始阅读到 Kragel 博士这篇论文(Kragel et al., 2018)时也是这样想的，但是在了解神经表征泛化的方法原理之后，发现并不能进行个体水平的计算，我们也联系了该方法的作者 Kragel 博士讨论在个体水平上实现的可能性，但是并没有讨论出可行的解决思路，我们会在局限性部分对此问题进行声明，内容为：“由于本研究的分析时基于组水平的结果，尽管通过 bootstrapping 可以得到显著性，但并不能得到每个个体的泛化水平的指标，这限制了它作为客观生物标记的可能性。”(1149 行-1151 行)。我们非常期待在未来的研究中解决这个问题，这样也能更好的推广该方法到更多的研究领域。关于“异常发育的客观生物标记”等涉及个体水平应用的表述，我们也会进行删除以保证论文的严谨性。

## 讨论

**意见 32:** 讨论第 3 段，“发现了儿童在多元需求额顶系统中表现了更弱的跨情感认知域相关的多体素神经激活模式的泛化水平，特别是其内部的顶内沟皮层区域”，“特别是其内部的顶内沟皮层区域”这个观点的依据是什么？

**回应:** 对于多元需求额顶系统而言，本研究试图探索是整个额顶网络还是其中的关键脑区存在神经表征泛化现象，因此分别针对 IPS+FEF、IPS、FEF 均进行了跨情感与认知领域的神经泛化水平进行测量。“特别是其内部的顶内沟皮层区域”观点的依据，是在针对单独的 IPS 和 FEF 脑区神经表征泛化水平的分析中，只有 IPS 在成人组表现出了显著水平，图 6 中用†标记的，而 FEF 在三个组别均未达到显著，只表现出了各组别的组间差异。因此，我们推测，IPS 相对于 FEF 来说，在支持不同情感与认知任务时，能够更好的对不同层级情感或认知功能相应的神经表征信息进行组织。

**意见 33:** 讨论最后一段，仅凭目前的结果，似乎还不能推断出顶内沟皮层是多元需求额顶系统的核心节点以及它在情感认知功能发展的重要作用。

**回应:** 感谢审稿人的建议，以目前的结果来看，顶内沟皮层是多元需求额顶系统的核心节点以及它在情感认知功能发展中重要作用的结论稍有不妥，我们基于当前结果的部分推论也会慎重改进，以避免过度推论。针对审稿人的建议，我们对最后一段进行了重新的讨论，“基于前人研究中的理论和实证性证据，表明多元需求额顶系统在多种情感与认知任务的执行中起着关键作用，可能存在一种不同情感与认知功能的组构性编码机制，支持着神经信息的组织和协调以适应多种任务需求。基于本研究的发现，我们推断，多元需求额顶系统在促进从儿童到成年期的情感与认知功能的高度整合和成熟中起到了重要作用。此外，我们还建立了神经元族群差异性共激活与组构性编码机制之间的联系，作为一种潜在的机制来支持跨情感与认知领域的整合性发展。神经泛化过程是儿童跨情感与认知领域发展的基础架构，支持发育过程中多种情感与认知功能的增强和完善，或许也作为通用性的机制支持着学习泛化能力的产生与发展。”（1125 行-1132 行）

此外，我们重新阅读和认真推敲了整篇论文，对类似问题也都进行了修正，以避免过度推论。

**意见 34:** 请作者在讨论部分补充本研究的局限。

**回应:** 我们已在修改稿中补充了本研究的局限性。（1143 行-1154 行）

次要意见

**意见 35:** 表 1，年龄分组没有同下列的统计学信息对齐。

**回应:** 我们已在修改稿中将表 1 的年龄分组与统计学信息进行了对齐。

**意见 36:** 图 5 和图 6，请给出英文缩写对应的脑区名称，以及请作者确认图 6 中的“**IPL**”是否应改为“**IPS**”。

**回应:** 我们已在修改稿中统一改正为 **IPS**。

**意见 37:** “有趣的是，我们在两个儿童组的多元需求系统和控制分析的 **ROI** 中均未发现显著的皮层区域 ( $p \geq 0.854$ )。”这里的 **p** 是多个的话或许要表达为 **ps**?

**回应:** 我们已针对多个 **p** 的情况改为 **ps**。

意见 38：前言第 4 段，“多元需求系统在参与和控制各种认知任务时起着关键的“Hub”作用”，“Hub”一词能否赋予准确的中文名称？比如，枢纽？

回应：感谢审稿人为我们思考出了一个准确的翻译，我们已经在修改稿中统一翻译为了“枢纽”。

意见 39：一些文字表达比较难懂或者不够直接，建议稍微优化表述：

a. 前言第 2 段，“新情感认知功能的出现并不一定与被认为之前不成熟的“沉默”神经区域成熟时变得活跃有关”。

b. 前言第 4 段，“会随着意识注意点的变化而迅速重组以及分离连续的任务到不同阶段、值得注意的不仅仅要探索神经系统如何控制孤立的思想或行为步骤”。

c. 前言倒数第 2 段，“位于额顶网络的前额叶皮层对许多情感与认知功能都及其重要；

d. 方法 2.5 第 2 段，空间距离是通过 1 减去不同情感认知域下子条件相关的空间激活模式向量两两配对之间的皮尔逊相关系数（ $1-r$ ）”。

e. 结论，“并为研发儿童脑智异常发育（如自闭症、ADHD 等）的客观生物标记提供实验依据、促进提高基础教育质量水平和启发基于脑智发育的人工智能新型算法和理论模型”。

回应：感谢审稿人指出的文字表达问题，我们已针对这些文字进行了阅读并重新思考优化了表述：

a. “新情感与认知功能的出现，并不一定是由于神经发育过程中原本不成熟的“沉默”区域发展为成熟的“活跃”区域而形成的。”（782 行-784 行）

b. “专注于当前认知操作的具体内容，会随着注意力的转移而快速重新组织并有效分离连续的任务步骤。”（817 行-818 行）、“不仅仅是要探索神经系统如何控制孤立的思想或行为步骤”（823 行）

c. “位于额顶皮层的前额叶参与了许多情感与认知功能，不同的神经表征模式可以发挥各种功能，例如通过向大脑的其他部分提供自上而下的指令，引导着认知过程正常进行(Fuster, 2015; Miller & Cohen, 2001; Stuss & Knight, 2013), 它在多种任务表征中还支持着规则的理解和运用等(Cole et al., 2011; Sakai, 2008; Tschentscher et al., 2017; Wallis et al., 2001), 此外，还有外侧前额叶皮层中复杂认知表征对组结构性编码的有效支持(Cole et al., 2011; Reverberi et al., 2012)。”（840 行-845 行）

d. “是通过皮尔逊相关系数（ $r$ ）计算所有被试的情感与认知过程成对匹配的空间距离之后，再用 1 减去该值得到的（ $1-r$ ）。”（1004 行-1005 行）

e. “并帮助教育者了解儿童脑智发育规律以促进提高基础教育质量水平和启发基于脑智发育的人工智能新型算法和理论模型。”（1141 行-1143 行）

---

## 第二轮

**审稿人 1 意见：**研究从跨不同情感与认知域的角度探讨了交互式特异化理论的构想，考察了儿童多种情感认知能力发展的可泛化神经表征脑发育机制，研究方法具有创新性，有助于更好的理解儿童青少年的脑智发育规律。作者对相关问题做了详细的回答和补充说明，提升了文章的科学性和严谨性，建议进一步修改文字提升可读性后发表。

**回应：**感谢审稿人认可我们对其提出问题做的回答和补充，根据审稿人的建议，在第二轮修改中，我们对整篇论文的文字表述进行了仔细阅读和改进，以提升其可读性。

**审稿人 2 意见：**感谢作者对我提出的问题做了细致的回答，我也阅读了其它两位审稿人的意见及作者的回复。第三个审稿人的审稿意见非常细，作者也做了详细回复。修改后的手稿得到了很大提高。

**回应：**感谢审稿人的积极评价和宝贵意见，使我们稿件的科学性和专业性得到了极大的提升。

.....

**审稿人 3 意见：**感谢作者的专业回复，很好地解答了我的大部分疑问。我个人特别喜欢这项研究，也特别喜欢《学报》，因此读得比较细致，正好借这个审稿的机会跟作者做更深入的交流学习，或许可以给其他读者以参考；另一方面我也把看到的或许还可以改进的地方说出来，供作者和编辑参考。在作者对以下意见进行澄清或者修改之前，我暂时还不能推荐发表。意见中的图，请见附件。

**回应：**我们感受到了审稿人对本研究的热情，花费了大量的时间精力阅读修改稿，提出宝贵的意见，以帮助我们改进论文水平，特别是方法部分的专业性，以及在此轮修改中又为我们指出了多处失误。针对审稿人的意见，我们进行了第二轮认真仔细的修改，希望本次修改后的稿件能够达到审稿人认可的发表水平。

**意见 1：**图 R1 和表 R1 很好地补充了被试缺失的情况，这部分很重要，建议作者放入正文

或者补充材料。另外，第 424-425 行，作者回复“这个缺失比例与前人文献中类似的儿童影像大样本研究是类似的(Zhang, 2019)”，建议作者在正文中明确写出本研究的缺失比例和 Zhang (2019) 的缺失比例。

**回应：**被试筛选流程确实是非常重要的部分，可以为未来同类型研究和结果可重复性验证的研究提供参考。根据审稿人的建议，我们将图 R1 和表 R1 放入了补充材料，并在正文中增加了相关表述，供感兴趣的读者参考。此外，我们也将本研究的样本缺失比例和 Zhang (2019) 的缺失比例在正文中进行了补充。修改的内容为：“相对于总体样本，最终纳入分析的样本量剔除比例较高，其原因是由于儿科神经影像学检查通常极具挑战性，例如，扫描过程中头部运动过多、扫描任务意外中止、以及需要四个任务同时满足剔除标准等。我们目前的研究（剔除比例 66%）以及国际上类似的儿童研究（剔除比例 49%）均会遇到这样的问题(Zhang et al., 2019)，头部运动对 fMRI 数据收集的质量和后续数据分析的可靠性有着至关重要的影响，因此我们选择了相对严格的纳入标准，具体的被试筛选流程见补充材料（图 S1）。此外，为了排除未被筛除被试和筛除被试在认知或情感上特质差异对实验结果的影响，我们对比了两组被试各个任务的绩效，未发现显著性的任务绩效差异，结果报告见补充材料（表 S1）。”（1079 行-1086 行）

**意见 2：**作者似乎遗漏了我在上一稿的意见 16 中的问题：“图 5A 中的重叠脑区，是指这六个条件中都激活了的区域吗？方法部分好像没有足够的交代。”我现在仍然保留这个好奇，并建议作者可以在正文中写得更清楚。

**回应：**审稿人的理解是准确的，图 5A 中的重叠脑区，是指被试在六个条件中均被激活的脑区的重叠部分，即各条件的激活交集。论文提交时第一稿的图 5A 是三组人（低年龄儿童、高年龄儿童、成人）合并的结果，根据审稿人 3 第一轮意见 3 中的建议，我们对图 5 进行了修改，将此结果分三组被试进行呈现，现在的图 5A/B/C 表示三组被试在六个条件中均被激活脑区的重叠部分。我们在此次修改中增加了相关的表述，内容为：“如图 5 所示，红色区域表示被试在六个条件中均被激活的脑区的重叠部分，即各条件的激活交集。”（1241 行-1242 行）

**意见 3：**请问作者是把 6 个条件的激活图在空间上简单地做了重叠吗？我注意到代码里有个“Generalizable-Emotional-Cognitive-Tasks-main/Generalization/AnalyScripts/BayesFactorOverlap.m”文件，但作者在正文中完全没提到贝叶斯因子，请问是作者在正文中漏写了，还是误

提供了无关多余的代码？

回应：本研究中的分析代码，是基于 Kragel 博士发表的论文（Kragel et al., 2018）中代码进行的修改变式，并未使用其上传的所有脚本。如审稿人所说，我们确实上传了部分无关多余的代码，“BayesFactorOverlap.m”文件是本研究未使用的分析代码。在本轮修改中，我们对代码重新进行了整理和上传，只保留了分析中使用到的代码。

意见 4：作者的代码写得很好、也有不小的工作量。不过，有些脚本我没看明白用处。比如，“Generalizable-Emotional-Cognitive-Tasksmain/Generalization/AnalyScripts/BICAnalysisVogt.m”文件，从注释来看，似乎是前人研究的脚本，不是本研究要用到的。更重要的是，这么多脚本和函数，我作为读者不知道从哪开始读这些文件，因为缺少数据，也没办法运行。如果较大的原始数据不便分享，建议作者提供部分数据，或者部分重要的过程数据，以便读者能自己跑下来。比如，提供 GLM 后的 beta 图，一名被试 6 张 beta 图，共 269 名被试，占用的存储其实也没有太夸张。再次请作者梳理代码和数据：删除无关的文件和代码、确保注释的准确性、提供让代码能正常运行的（部分）数据、给出主要脚本的运行/阅读顺序（比如，主要脚本命名为 A1\_xxx、A2\_xxx 来表示先后）等。

回应：结合意见 3 和意见 4，我们在上传代码时，提供了部分未在本研究中用到的脚本，也未提供相关的数据，可能会给审稿人和读者带来困惑。因此，基于审稿人的建议，我们重新对代码进行了梳理，提供了相关数据。同时也删除了无关代码、确保注释的准确性、以及给出了主要脚本的运行顺序。纯代码版本仍可访问“<https://github.com/drhaol/Generalizable-Emotional-Cognitive-Tasks>”获取，包含数据版本可访问“[https://pan.baidu.com/s/1rKQIg3dINKE-zg38\\_YmTnQ?pwd=ycn1](https://pan.baidu.com/s/1rKQIg3dINKE-zg38_YmTnQ?pwd=ycn1)”获取。

意见 5：第 1042-1045 行，“在三组被试中，我们发现双侧顶内沟、双侧额眼区域均出现了跨情感与认知领域的神经共激活模式，另外，右侧腹侧额叶皮层（ventral frontal cortex, VFC）和左侧前辅助运动区（pre-supplementary motor area, pSMA）也表现出了部分组别的共激活现象，但未在所有组别出现（如图 5 所示）”，这部分文字因为缺少精确的表格呈现，使得这部分结果大部分读者估计都读不懂。建议作者提供与图 5 配套的表格，交代清楚脑区名称、团块大小、peak 点脑区坐标、peak 点的值（这个值是 t 值吗？作者好像也没明确在文章中说这个全脑分析做的是 t 检验）。因为不重叠的脑区特别多，作者可以只报告重叠脑区（也就是图 5 标注为红色的脑区），这样读者可以获得更多的信息。



**回应:** 感谢审稿人提出文字表述需要增加精确表格呈现的建议, 我们现已针对图 5 在正文补充了表格报告的表述 (1245 行-1246 行), 提供了各组别额顶系统在不同情感与认知条件下激活的重叠脑区表格 (表 S2), 包括了脑区名称、团块大小、peak 点脑区坐标, 由于图 5 是使用经过阈值校正后的 Tmap 做成 Mask 之后取的交集, 所以无法提供 T 值的报告。

**意见 6:** 第 1027 行, 作者对于 Bootstrapping 的描述不够清晰, 不知道是在每组内抽样, 还是把 269 名被试混在一起抽样, 我看过代码后才知道其实是前者, 建议作者在正文中写清楚。

**回应:** 感谢审稿人提出的建议, 我们在修改稿中针对此问题进行了表述上的改进, 改进的内容为: “在每次重建中, 通过对每组被试在不同情感与认知条件下进行组内抽样, 再使用神经激活表征模式来构建了一个新的 RDM 并进行相应 GLM 的参数估计。”(1224 行-1226 行)

**意见 7:** 第 1030-1031 行, 作者写道 “每次我们使用独立样本 T 检验从 5000 次迭代中随机选择 100 个值进行比较并获得 p 值, 这个过程会重复 1000 次”。感谢作者对这部分的优化, 我明白这个操作了。但是, 表述上, 不是 “100 个值”, 而是 “100 组值, 每组有 3 个值, 分别指代儿童低年龄组、儿童高年龄组、成人组”。不然读者仍然会困惑, 这 100 个值怎么做独立样本 T 检验。

**回应:** 感谢审稿人提出的改进建议, 我们针对此问题进行了改进, 改进的表述为: “每次我们使用独立样本 T 检验从 5000 次迭代中针对这三组每组随机选择 100 个值进行组间的两两比较并获得 p 值。”(1228 行-1229 行)

**意见 8:** 第 1031-1032 行, “相应的显著性定义为差异显著结果 (我们使用了  $p = 0.001$  的阈值决定是否显著) 的总数量除以 1000”。这里应该是 “差异不显著结果的总数除以 1000” 才对吧? 不然, 绝大部分次数算出来都是差异显著, 你除以 1000 后, 得到的 p 值反而还挺大, 这不合理。

**回应:** 感谢审稿人为我们指出的表述错误, 确实应该按照审稿人的理解, 显著性的定义应该是差异不显著结果的总数除以 1000, 我们的代码也是按照该定义进行的计算的。修改稿中已针对此问题进行了修改, 修改的内容为: “相应的显著性定义为差异不显著结果 (我们使用了  $p = 0.001$  的阈值决定是否显著) 的总数量除以 1000。”(1230 行-1231 行)

**意见 9:** 我暂且把前面两点提到的这个计算组间差异的显著性的方法 (第 1030-1035 行) 称

作“5000-100-0.001-1000 方法”或者“方法 A”，请问该方法中的 100、0.001 是如何确定的？目前的设置有没有前人文献的依据？如果改成 200、0.05，是不是就更容易显著了？或者改成 20，0.0001，是不是就更不容易显著了？这两个地方很容易有 p hacking 的嫌疑。作者在第 1031-1035 行解释道：“……这么做的原因在于我们使用的方法在 5000 次迭代中本身就具有一定的随机性，每次迭代会对每个情感与认知条件成对匹配的神经相异性进行随机化，因此所产生的参数估计值 ( $\beta$ ) 也有所不同，这就构成了 5000 次结果的样本分布因此，进行组间差异比较时，要随机抽取 100 个值进行 T 检验。”这部分解释在我看来并不清楚（或者说不具有说服力）最后那个“因此”前后的逻辑关系似乎并不成立。在最后那个“因此”之前，作者应该要加上句号或者逗号。

**意见 10:** 我把第 1024-1029 行中计算每组组内神经泛化水平显著性的算法称作“5000-0 方法”或者“方法 B”。如果在上一条意见中我转述的作者的解释成立，那么这里的方法 B 也应该改成方法 A（5000-100-0.001-1000 方法）才对，即在 5000 次 bootstrapping 中抽取 100 组 beta 值，做单样本 t 检验，重复 1000 次，（不）显著的次数占 1000 次的比例作为 p 值。为什么不统一使用方法 A 或者方法 B？在这一版审改稿中，作者用方法 B 在多元需求额项系统里做了组间差异的显著性分析，感谢作者，同时我也注意到作者没有呈现方法 B 在控制脑区里的结果，我对这部分很好奇。如果方法 A 和方法 B 都有前人文献的依据，我仍然建议作者统一使用其中一种方法，以避免 p hacking 的嫌疑。从目前的信息来看，我个人从逻辑上更倾向于方法 B，但如果作者可以说清楚他们的其他考虑，我也很乐意改变我的观点。

**回应:** 可以看出审稿人对于神经影像学研究的统计方法有专业且深度的理解，针对审稿人提出的统计显著性问题，我们将合并意见 9 和意见 10 来统一回复。审稿人疑问为什么不统一使用方法 A 或者方法 B，在本研究中，使用方法 A 还是方法 B，是根据研究问题而定的。对于我们使用的“方法 A”，可以认为是置换检验的一种形式，置换检验在各学科数据统计中的应用非常广泛。针对本研究中的问题，使用方法 A 是为了检查组间的差异。该方法中的 (100、0.001)、(200、0.05)、(20, 0.0001) 在本研究中则是主观选取的标准，就比如我们在报告 T-test 的结果时，可以报告  $p < 0.05$ 、 $p < 0.01$  或  $p < 0.001$ ，当然越严格越好，我们选取了较为折中的标准。对于审稿人提出的 (200、0.05) 和 (20, 0.0001) 两个标准，我们也补充了相关的差异显著性分析(审稿人也可以在我们提供的含数据代码中尝试其他不同的标准)。结果的趋势如审稿人所料，越严格的标准则显著的结果少。重新分析的结果显示，(200、0.05) 的结果仍然与 (100、0.001) 结果一致；(20, 0.0001) 的结果中，成人组与儿童组的

差异在各 ROI 中仍然显著，只有在 FEF 脑区中，成人与高年龄组儿童差异不显著，基本上可以说明结果的鲁棒性较强。

对于我们使用的“方法 B”，就是拔靴法（Bootstrapping）的应用了，针对本研究中的问题，使用方法 B 是为了检查神经泛化水平的显著性，该方法是我们直接参考了 Kragel 博士的论文中使用的显著性统计方法（Kragel et al., 2018），用以检查不同脑区是否存在神经泛化的现象。该显著性检验是在组内进行的，没有进行组间的比较。方法 B 在控制脑区中是进行了显著性检验的，手稿中也有相关报告，内容为：“但是，在控制分析的 ROI（包括双侧 dACC、双侧 DLPFC 和双侧 AI,  $p_s \geq 0.991$ ）中却未发现统计上显著的皮层区域（图 6）”。方法 A 是没有在控制脑区中进行检验的，原因是由于控制脑区中没有神经泛化显著的脑区，并且大部分神经泛化指标的值均为负值（图 6），负值是没有意义的，只有正值才能代表神经泛化的水平，这一点我们与 Kragel 博士也进行过交流，在该方法中，他建议只关注正值即可。

此外，最后那个“因此”属于笔误，现已删除。

**意见 11:** 第 1066 行，“†，显著性  $p < 0.001$ ”，但正文的第 1056-1058 行（“成人在多元需求额顶系统（双侧 IPS + 双侧 FEF:  $p = 0.0018$ ）中表现出了显著高于随机水平的跨情感与认知的神经泛化水平，特别是在双侧 IPS 中（ $p = 0.004$ ）”）里的 p 值却是大于 0.001 的。是我理解错了，还是作者笔误？

**回应:** 感谢审稿人指出该问题，这里的“†，显著性  $p < 0.001$ ”是由于笔误造成的，实际上应该为“†，显著性  $p < 0.01$ ”，我们在修改稿中对此问题进行了修正。（1266 行）

**意见 12:** 另外，作者所说的“随机水平”是指的 0 吗？我怀疑此处有待商榷，你这里其实是图 4 的 beta 28、beta29、beta30 跟 0 比较。然而，计算该 beta 的自变量的编码方式，以 beta 28 为例，应该是我下图（图 C1）中圈红的地方为 1 其余地方为 0，对吗？是的话，那么这个 beta28 值相当于是圈红的地方的表征相异性（Y）减去其余地方的表征相异性（Y），如果差值小于 0，那么完全可能是其余地方表征相异性更大的原因。我理解的是，这个地方没有体现“随机水平”。如果我理解错了，请作者澄清。

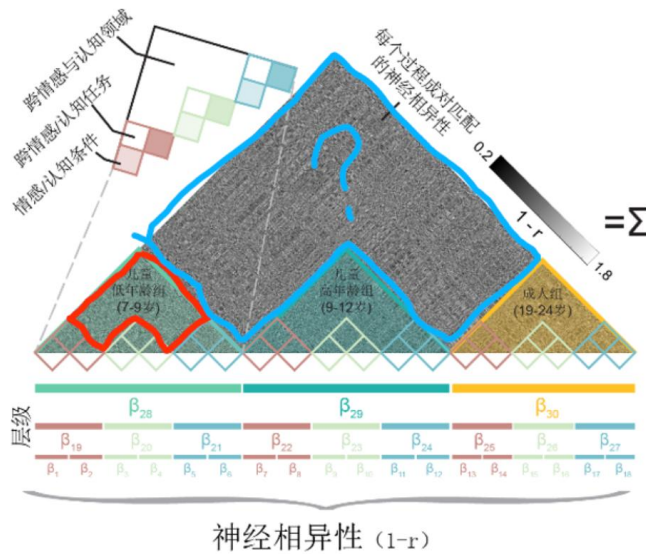


图 C1 对模型构建的编码方式的疑问

回应：在神经泛化水平显著性检验时，“随机水平”是指的 0，显著性的定义为参数估计值（也就是 beta 值）中小于 0 的样本的比率，审稿人的理解是正确的，确实是 beta 28、beta29、beta30 跟 0 比较。对于 beta 值的估计，我们再进行澄清一下。首先，如图 4 所示，整个一般线性模型（GLM）是包含了左右两部分，左侧为计算神经相异性（代码可参考 A4\_BootstrapROI 中的 122-124 行）的真实值，也就是一般线性模型的 Y，是从 ROI 中提取的激活模式，转化为向量后用以计算神经表征相异性（1-r）；右侧为计算理论相异性（代码可参考 A4\_BootstrapROI 中的 59-75 行）的模拟值（0 和 1），也就是一般线性模型的 X1，X2，… …，Xn，每个 X 对应的 $\beta$  值就代表着神经泛化水平。以 beta28 为例，对于理论相异性的计算，模拟值应该是如下“图 C1-R1”中，圈红的地方为 1，圈蓝的地方为 0，这个 beta28 值并不是圈红的地方的表征相异性（Y）减去其余地方的表征相异性（Y）得到的，而是通过估计 GLM 右侧的 $\beta$  28 得到的（图 4 “=” 右侧）。

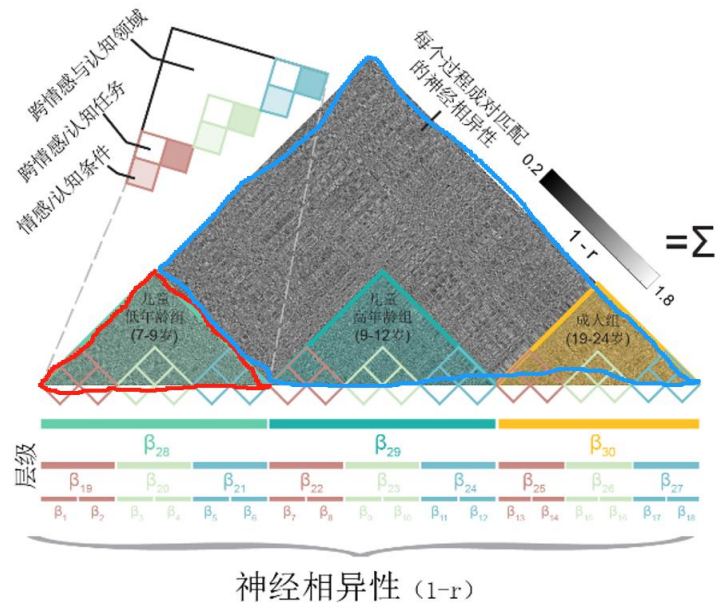


图 C1-R1

意见 13: 我也很好奇, 在图 C1 中, 我框出的蓝色区域 (也就是儿童低年龄组、儿童高年龄组、成人组三组之间的相异性, 这个相异性的值肯定就很大了, 儿童组中减去它得到负值也是“情有可原”了), 作者是否把它也纳入计算作为我前面所说的“其余地方”? 从图上看, 作者像是纳入了 (不然应该在图上扣掉该区域), 但我没找到相应的代码位置 (请作者指出)。另外, 我也在思考如果不纳入会否在计算量上更加高效、得到的 beta 28、beta29、beta30 大概率也是正值, 而且更能说明问题: 只纳入情感认知条件 vs 跨情感/认知任务 vs 跨情感与认知领域, 不纳入“跨被试组跨情感与认知领域”。这只是我的思考, 期待和作者交流。这里并不是在建议作者在本研究中要去做这样的分析

回应: 在图 C1 中, 审稿人框出的蓝色区域并未纳入计算, 如果纳入计算, 还会存在一个 beta31, 这就是跨组别的神经泛化水平了, 与本研究的核心问题关系不大。而本研究构建的模型, 也正与审稿人所理解的是一致的, 只纳入情感认知条件 vs 跨情感/认知任务 vs 跨情感与认知领域, 并未在模型中纳入“跨被试组跨情感与认知领域”的分析。

意见 14: 图 4 的右半部分, 最下面这层 (第一层), N 分别等于 91、91、89, 对吗? 如果具体写出来, 可能会更好帮助读者理解。因为确实有门槛。此外, 第二层作者写的是 A1、A2、A3、A4。我没理解错的话, 应该只有 A1 和 A2 吧? 因为注意任务下, 只有警觉和定向两个条件。同样, 工作记忆任务和情绪匹配任务也是? 如果我理解错了, 欢迎纠正我, 以

及考虑是否有优化的地方以避免读者错误理解。

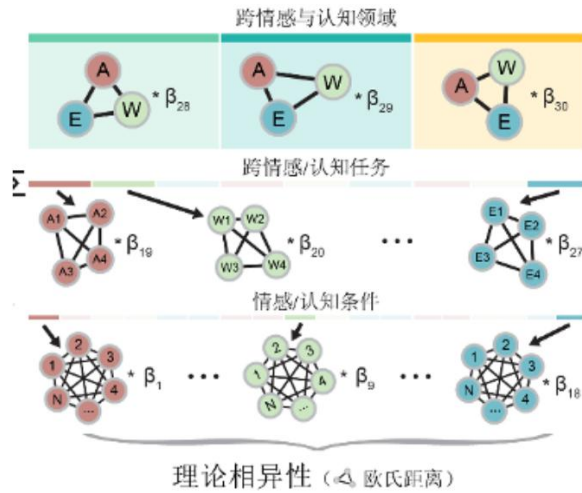


图 C2 转述图 4 右半部分

回应：感谢审稿人提出的优化建议和错误纠正。第一层级的 N 分别等于 91、91、89，我们已经在更新的图 4 中对此进行了标注。第二层级对应到每组分别是两个条件，我们也在图 4 中进行了纠正。由此可以看出，审稿人在短时间内也对该方法有了深入的理解。

意见 15：神经泛化水平这个指标，似乎作者并没有明确地去界定/描述它。我刚开始以为，这个指标越大意味着一个脑区更有能力用差不多相同的激活模式来完成多种不同的任务，泛化指的是该脑区有能力“以不变应万变”。但仔细看才发现，作者算的是神经相异性（而不是神经相似性），我更正后的理解是，这个指标越大意味着一个脑区更有能力用更加独特的激活模式来完成多种不同的任务，泛化指的是该脑区有能力“以‘万变’应万变”。我更正后的理解正确吗？另外，因为这个概念/指标是本研究的核心，作者如果给出更清晰的界定或者描述可能会更好。

回应：感谢审稿人提出关于神经泛化指标计算原理和概念界定的建议。首先，关于计算过程中使用的神经相异性而不是神经相似性的疑问的解释，类似于意见 12 的回复，如正文中图 4 所示，对于 beta 值的估计，整个一般线性模型是包含了左右两部分，左侧为计算神经相异性的真实值，右侧为计算理论相异性。由于右侧的理论相异性是基于欧氏距离计算的，因此值越大，代表着空间距离越大，理论相异性越大；而左侧要保持方向上的一致，就应该使用神经相异性，值越大，代表距离越大，这样估计出的 beta 值才能够代表神经泛化水平。如果想要使用神经相似性也是可以的，那右侧的就应该变式为理论相似性了，而非理论相异性，这个是本方法的核心基础一般线性模型决定的。

此外，根据审稿人的建议，我们也在修改后的正文中补充了神经泛化的界定描述，补充的内容为：“将神经泛化指标应用到脑智发育领域，可以理解为某一特定神经系统通过组构性的编码机制，在发育中逐渐成熟从而支持多种情感与认知发展的过程。”（1041 行-1043 行）

**意见 16：**请问控制脑区也是作者基于“domain general”一词搜索出来的激活图，卡阈值后，得到吗？那么，控制脑区岂不也是“domain general”脑区了？也属于多元需求网络？另外，第 987-989 行，在脑区后面紧跟参考文献，容易让人误以为是根据这些文献选取的 ROI，让读者感觉信息矛盾。

**回应：**控制脑区是基于“domain general”一词搜索出来的激活图卡阈值后得到的，这么做的原因有两个方面：一是考虑统一阈值，如果另外单独搜索控制脑区的激活图也能够搜索到，但是每个脑区一张图，会存在阈值无法统一的问题，而且每个脑区基于的研究数量也不一致，又增加了误差因素；二是考虑本研究关注的核心在于多于需求额顶系统，虽然我们选取的控制脑区也属于多元需求系统，但是研究结果在控制脑区中并未表现出神经泛化的现象，这更加说明了多元需求额顶系统在脑智发育过程中表现出神经泛化的唯一性和关键性。关于在脑区后面紧跟参考文献的问题，我们修改了相关内容的表述，修改后的内容为：“我们还选取了支持注意、工作记忆和情绪功能的相关脑区作为本研究的控制分析 ROI，这些 ROI 在前人文献中也充分的参与着这些功能的执行，是具有代表性的脑区(Barbey et al., 2013; Petersen & Posner, 2012; Uddin et al., 2017)。”（1183 行-1185 行）

**意见 17：**图 R4 的行为结果很好。建议放入正文或者补充材料。

**回应：**感谢审稿人的建议，我们现已将图 R4 的行为结果放入了补充材料当中（见补充材料图 S2）。

**意见 18：**第 887 行，“最终满足入组标准的被试包括 89 名被试的成人组（19-24 岁， $M \pm SD = 21.74 \pm 1.64$ ）和 182 名儿童，儿童按人数平均分为儿童低年龄组（7-9 岁， $M \pm SD = 8.22 \pm 0.67$ ）和儿童高年龄组（9-12 岁， $M \pm SD = 10.24 \pm 0.74$ ）。”请问作者是怎么平均分的？这两组里都有 9 岁的儿童。假设你们有 40 名 9 岁儿童，需要 10 名归低年龄组、30 名归高年龄组，你们是随机分配吗？还是说你们也记录了月龄，把他们再按月龄排序，最小的前 10 名分配到低年龄组，剩下 20 名月龄较大的分配到高年龄组。请作者写得更加清楚。

**回应：**感谢审稿人指出的表述问题，关于儿童分组，我们是按照出生天数平均分到儿童低年

龄组和儿童高年龄组的。我们在改稿中补充了相关内容，表述清楚了分组的规则。(1078 行)

**意见 19:** 第 966 行,“对于 WM 任务,将两个条件(1-Back 和 2-Back)建模为单独的回归变量”,请问在 GLM 建模的时候 0-back 有作为 event 放进去吗?如果有,请明确写出来,以避免读者误解。尽管作者不关心它,我也同意它不太需要工作记忆,但从数据分析的角度,一般是要把所有 events 都放进去的,只是在做 contrast 的时候,不关心它,不提取它的 beta 值。作者是这么做的吗?请澄清。

**回应:** 感谢审稿人提出的建议,我们在建模时,0-back 是有作为 event 放进 GLM 当中的,我们在改稿中也对此信息进行了补充(1163 行-1164 行)。0-back 在某些分析当中也会作为基线用于计算 1-0back 和 2-0back,所以常规分析中我们均会将 0-back 纳入到模型中。

**意见 20:** 作者对上一稿的意见 30、31 的回复中,作者提到“目前还无法使用神经泛化指标进行个体水平的计算”,并且解释道:“实际上我们一开始阅读到 Kragel 博士这篇论文(Kragel et al., 2018)时也是这样想的,但是在了解神经表征泛化的方法原理之后,发现并不能进行个体水平的计算,我们也联系了该方法的作者 Kragel 博士讨论在个体水平上实现的可能性,但是并没有讨论出可行的解决思路。”经过作者的耐心细致的讲解(感谢作者),我感觉我也懂了神经表征泛化的方法原理(也可能还不是真的懂了),但我认为或许有办法进行个体水平的计算。Kragel et al (2018) 的研究是考察以往研究之间的相异性,确实不能算出每个研究内的神经泛化水平。但是你这里的是一个被试,一个被试的 6 张图,实际上可以看作是多张图叠加平均而来的(比如,每个 trial 一张图,或者每个 block 一张图;或者把 trial 或者 Block 分成组,每组一张图,假设分 N 组,你就能得到 N 张图)。只要 N 大于 2,那么在每个被试内,就能用这 6N 张图能构建如图 C3 的矩阵,进而用类似的方法去每个被试的计算神经泛化水平。当然,作者在这方面深耕多年,或许已经考虑过这个方案了,我很期待知道作者的想法。如果我有理解不正确的地方,或者没考虑周全的地方,欢迎纠正我。当然,我这里也并不是建议作者在本研究中就做这个分析,因为目前的方法已经可以说明问题了,我只是在探讨这种方法的可能性。



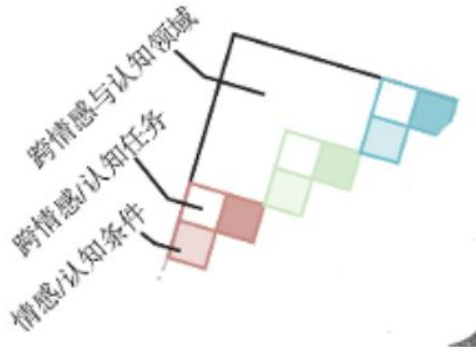


图 C3 截取自图 4 右上角

**回应：**关于神经泛化指标计算的问题，实际上我们也处于初步探索当中，并且也非常希望在未来的研究中更新和变式其应用，以解决更多的神经科学问题，遇到能够与我们深入讨论该方法的审稿人，我们是极其高兴的，审稿人 3 也给我们提供了许多灵感，我们由衷的表示感谢！

关于个体水平神经泛化指标的计算，审稿人提到“只要  $N$  大于 2，那么在每个被试内，就能用这  $6N$  张图能构建如图 C3 的矩阵，进而用类似的方法去每个被试的计算神经泛化水平。”这确实是一个全新的分析角度，可以计算出每个被试  $N$  次神经泛化指标，最后再求平均值，这样就能得到个体水平的神经泛化指标。但是，要这样做，并且得到稳定的结果，可能需要以下两个条件：首先，每个任务这里的  $N$  需要保持一致，这样不同任务构建出的模型才能保持一致，这就需要重新设计任务程序和采集影像数据；其次，在重新设计任务的时候，可能需要考虑全部使用基于 block 设计的实验范式，以保证每张图结果的稳定性，event 设计的结果稳定性稍差，因为单个 trial 的任务持续时间太短了。我们在未来的研究中，会考虑在小样本实验中尝试该个体水平的神经泛化指标的计算方法，这样就能实现获取个体水平生物标记的目标了。最后，如果有机会的话，我们也非常乐意与审稿人 3 在未来继续深入讨论该方法在未来的应用，共同推进脑智发育领域的研究！

---

### 第三轮

**编委意见：**本论文经过多名审稿人建设性评价，以及作者的认真修改，目前已基本达到了学报的发表要求。建议作者进一步对文中的有关文字表述方式(如英文式中文表达)进行改进，提高语言的简明性(摘要和结论部分)和可读性后即可发表。

**主编意见：**本研究采用层级化神经表征建模方法，结合多个经典的情感与认知任务范式，对学龄儿童可泛化神经表征模式进行了考察。本论文的选题具有较强创新性，数据处理过程科学规范。参考外审专家意见进行修改后，本论文已经达到《心理学报》发表文章的相关要求，建议发表。