Shared and distinct neural signatures of feature and spatial attention

Anmin Yang[a], Jinhua Tian[a], Wenbo Wang[a], Jia Liu[b], Liqin Zhou[a*], Ke Zhou[a*]

[a]Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education (Beijing Normal University), Faculty of Psychology, Beijing Normal University, Beijing, China

[b]Department of Psychology & Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing, China

*Co-corresponding authors:

Liqin Zhou

Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education (Beijing Normal University),

Faculty of Psychology, Beijing Normal University

No. 19, Xinjiekouwai Street, Haidian District, Beijing 100875, China

E-mail: zhouliqin@bnu.edu.cn

Ke Zhou

Beijing Key Laboratory of Applied Experimental Psychology, National Demonstration Center for Experimental Psychology Education (Beijing Normal University),

Faculty of Psychology, Beijing Normal University

No. 19, Xinjiekouwai Street, Haidian District, Beijing 100875, China

E-mail: kzhou@bnu.edu.cn

Abstract

The debate on whether feature attention (FA) and spatial attention (SA) share a common neural mechanism remains unresolved. Previous neuroimaging studies have identified fronto-parietal-temporal attention-related regions that exhibited consistent activation during various visual attention tasks. However, these studies have been limited by small sample sizes and methodological constraints inherent in univariate analysis. Here, we utilized a between-subject whole-brain machine learning approach with a large sample size (N=235) to investigate the neural signatures of FA (FAS) and SA (SAS). Both FAS and SAS showed cross-task predictive capabilities, though inter-task prediction was weaker than intra-task prediction, suggesting both shared and distinct mechanisms. Specifically, the frontoparietal network exhibited the highest predictive performance for FA, while the visual network excelled in predicting SA, highlighting their respective prominence in the two attention processes. Moreover, both signatures demonstrated distributed representations across large-scale brain networks, as each cluster within the signatues was sufficient for predicting FA and SA, but none of them were deemed necessary for either FA or SA. Our study challenges traditional network-centric models of attention, emphasizing distributed brain functioning in attention, and provides comprehensive evidence for shared and distinct neural mechanisms underlying FA and SA.

Keywords: feature attention, spatial attention, between-subject whole-brain machine learning approach, brain signature

Introduction

The ability of visual attention to selectively focus on relevant or salient information from the environment while ignoring other stimuli is crucial for the successful functioning of various cognitive processes, including perception, memory, learning, and decision-making. Attention can be modulated across different perceptual domains, such as space[1,2], features[3–5], objects[6–8], time periods[9–11], and value[12,13]. These various forms of attentional modulation enable individuals to allocate their cognitive resources more effectively and efficiently by prioritizing information that is most relevant to their current goals, intentions, or interests.

In spite of the importance of understanding the mechanisms of attention for improving attention and related cognitive abilities, the long-standing debate over whether different forms of visual attention have common mechanisms has remained unsolved[5,7,14–16]. Here we focused on two main types of visual attention: feature attention (FA) and spatial attention (SA). FA improves perception of attended features of a stimulus, while SA enhances perception at attended spatial locations. Previous research suggested that these two forms of attentional modulation have similar neural mechanisms. Both FA and SA could strengthen the responses of the same neuron in area V4[17] and middle temporal area (MT)[18]. Patzwahl & Treue (2009) also found a positive correlation between the strength of feature and spatial attentional modulation at the single neuron level, implying that spatial location may not be unique, but only one of the features of a stimulus [18]. Moreover, the location and color attention-directing cues have been shown to produce largely overlapped activation in the frontoparietal network, reflecting that FA and SA shared common top-down neural control systems[19–21].

However, further investigations have suggested that FA and SA are distinct. These two types of attentional modulation can independently and additively affect an observer's behavioral

performance[22], the amplitudes of steady-state visual evoked potentials elicited in early visual areas[13], and the responses of single V4 neurons[23]. Electrophysiological recordings of neurons in area V4 revealed that SA modulated neuronal mean response gain but did not change tuning, whereas FA shifted neuronal tuning[24]. SA often modulated spatially localized subgroups of neurons in V4, while FA could modulate neurons far apart[25]. The time course of attention also varies between FA and SA. The initial part of the visual response modulated by SA was weak, while the later part was strong, reflecting a combination of bottom-up and top-down processes. In contrast, the modulation by FA remained constant over time, reflecting a purely top-down process[26]. Neurophysiological evidence from recordings in area MT implied that although spatial and feature attention may share a common top-down attention signal, they differ in their sensory normalization mechanism, as the normalization mechanism is only tuned by space, but not by feature[27]. Consistent with the findings from single-neuron recordings in non-human primates, a recent study based on multivariate decoding of human MEG showed that the two forms of attention induced different representational patterns of enhancement in the occipital cortex[28]. In addition to differences in sensory modulation areas, some research also proposed that FA and SA differed in top-down control systems. A reorienting for FA resulted in a more ventral activation pattern, despite large overlap in activity between FA and SA[19]. Bichot and collegues (2015) identified a subregion of prefrontal cortex called the ventral prearcuate (VPA) that played a critical role in FA, because neurons in the VPA showed earlier times of feature selection than those in the frontal eye field (FEF) or inferior temporal (IT) area, and inactivation of the VPA resulted in impairment of the animal's ability to find targets based on features while it did not affect the SA effect in FEF[3]. Further evidence supporting a distinct mechanism between the SA

and FA was that SA was more sensitive to individual perceptual load than to working memory capacity, while the opposite was true for FA[29].

While previous research has explored the similarities and differences between FA and SA, the limitations of existing methods have hindered a comprehensive understanding of the neural mechanisms underlying these attentional processes. For instance, electrocorticographic (ECoG) recordings could not simultaneously take into account all the signals from the whole brain, with most studies focusing primarily on sensory areas. On the other hand, univariate analyses in functional magnetic resonance imaging (fMRI) studies have identified overlapping brain activity associated with different cognitive processes but have not provided precise representational mechanisms underlying these processes. To overcome these limitations, we employed a between-subject whole-brain machine learning-based multivariate analysis[30–32] on a large sample to identify the neural representations associated with FA and SA. The approach allowed us to capture the complex and distributed patterns of neural activity across the entire brain, enabling us to uncover new algorithmic mechanisms underlying these attentional processes. Importantly, the use of machine learning algorithms enabled us to go beyond traditional univariate methods and identify fine-grained differences in the neural representations associated with different attentional processes, providing us with new insights into the functional organization of the brain regarding visual attention that were not available previously with ECoG or univariate fMRI analyses.

In this study, we employed the aforementioned machine learning approach to uncover the whole-brain multivariate patterns which could predict the conditions of FA or SA. Then we tested whether the brain signature predictive for FA was specific to FA, or it also carried information predictive for SA, and vice versa. In addition to evaluating the brain signatures at the

whole-brain level, we also retrained the prediction models based on cortical networks (network level) and brain clusters (cluster level). Single-cluster and 'virtual lesion' analysis were applied at the cluster level to investigate the necessity and sufficiency of each cluster for FA and SA separately. Patterns of FA and SA were further compared at the voxel level to validate the distributed property of the neural underpinning.

Results

Participant's behavioral performance in FA and SA tasks. 235 participants underwent both FA and SA tasks with fMRI scanning (Fig. 1, see Methods for details of experiment paradigms). Fig. 2a showed the sensitivity (d') and response time (RT) in the FA and SA experiments (details see Supplementary Table 1). In the FA experiment, both sensitivity and RT were significantly different between feature condition and conjunction condition (two-sided paired sample t-test: d': $t(234) = 55.18, p < 0.001, Cohen's\ d = 4.92$; RT: $t(234) = -36.29, p < 0.001, Cohen's\ d = 2.75$), supporting conjunction condition being more difficult. Likewise, in the SA experiment, the peripheral condition was more challenging than the central condition, yet with weaker effect size (two-sided paired sample t-test: d: $t(234) = 16.74, p < 0.001, Cohen's\ d = 1.33$; RT: $t(234) = -2.78, p = 0.008, Cohen's\ d = 0.17$). The less sensitivity and slower RT in difficult conditions was consistent with the behavioral performance discovered in Wojciulik & Kanwisher (1999)[21].

The FAS and SAS were capable of mutually predicting each other's behaviors. We applied the Support Vector Classifier (SVC) with a linear kernel to unveil the brain signatures of FA (FAS) and SA (SAS) separately (for details see Methods). Each attention type's brain signature consisted of the voxel weights that reflected their respective contribution to predicting the attention condition. To evaluate the robustness and mutual predictability of these brain

signatures, we conducted tests on a hold-out dataset ($n = 78$) including both the intra-task scenario (e.g., using the FAS to predict FA conditions) and inter-task scenario (e.g., using the FAS to predict SA conditions). The prediction performance was quantified using the Area Under the Curve (AUC) metric, as illustrated in Fig. 2b (for other model performance metrics see in Supplementary Table 2). First, both brain signatures exhibited prediction performance above chance in intra-task scenario (with bootstrap 10,000 times, FA intra-task: $mean\ AUC = 0.99$, $95\%\ CI = [0.97, 1.00]$, one-sided one-sample t-test: $t_{9999} = 7308.93,\ p < 0.001$; SA intra-task: $mean\ AUC = 0.67$, $95\%\ CI = [0.62, 0.72]$, one-sided one-sample t-test: $t_{9999} = 655.11$, $p < 0.001$). We also observed that the SAS exhibited comparatively lower prediction performance within intra-task scenario when compared to the FAS. One possible explanation for this discrepancy could be the larger variation in task difficulty between conditions in the FA experiment.

More interestingly, both FAS and SAS exhibited prediction performance above chance level when applied to inter-task scenarios (with bootstrap 10,000 times, FA inter-task: $mean\ AUC = 0.66$, $95\%\ CI = [0.58, 0.73]$, one-sided one-sample t-test: $t_{9999} = 422.89, p < 0.001$; SA inter-task: $mean\ AUC = 0.60$, $95\%\ CI = [0.54, 0.66]$, one-sided one-sample t-test: $t_{9999} = 319.56,\ p < 0.001$). These findings suggested the presence of shared neural components within the FAS and SAS.

However, it was also observed that the SAS outperformed the FAS when tested on the SA dataset (two-sided independent-sample t-test, $t_{9999} = 26.84,\ p < 0.001$, $Cohen's\ d = -0.38$), indicating a higher sensitivity of the SAS in predicting SA conditions compared to the FAS. Conversely, the FAS demonstrated better predictive capabilities on the FA dataset than the SAS (two-sided independent-sample t-test, $t_{9999} = 1210.75,\ p < 0.001$, $Cohen's\ d = -17.12$). In

addition, both FAS and SAS exhibited a significant drop in performance from the intra-task scenario to the inter-task scenario, suggesting the existence of distinct components between the FAS and SAS (one-sided independent-sample t-test, FA: $t_{9999} = 872.49,\ p < 0.001, Cohen's\ d = -12.34$, SA: $t_{9999} = 171.50,\ p < 0.001,\ Cohen's\ d = -2.43$).

To verify the robustness of the signatures, we additionally employed a PCA-Logistic algorithm and compared the signatures identified by both approaches (for details see Methods). Remarkably, the signatures derived from both approaches exhibited a high degree of concordance, as evidenced by a significant correlation in weight (For FA, $r = 0.76$, 95% $CI = [0.75, 0.76]$, $p < 0.001$; For SA, $r = 1.00$, 95% $CI = [1.00, 1.00]$, $p < 0.001$, see Supplementary Fig.1). These results provided compelling evidence for the robustness of the identified signatures.

FAS and SAS exhibited both shared and distinct components. Using a linear kernel in the SVC models, we identified the FAS and SAS (Fig. 3a). To ensure reliable voxel predictability, we utilized bootstrap resampling and employed FDR correction with a threshold of $q < 0.05$ for multi-comparison correction. The FAS and SAS manifested both shared and distinct components. As illustrated in Fig. 3a, the brain regions within these signatures could be categorized into three categories: those specific to FA, those specific to SA, and the conjunction regions. The conjunction regions were widely distributed across frontoparietal areas and visual areas, including bilateral IPTO (junction of intraparietal and transverse occipital sulci), left aSMG (anterior supramarginal gyrus), right ACG (anterior cingulate gyrus), bilateral PcC (precuneus cortex), right OFG (occipital fusiform gyrus), right PrG (precentral gyrus) and right OcP (occipital pole), etc. Regarding the regions specific to FA, we observed activations in the right AG (angular gyrus), bilateral SFG (superior frontal lobe), bilateral COpC (central opercular

cortex), and left FOpC (frontal opercular cortex), etc. In contrast, we did not find any regions specific to SA.

In addition, we conducted a univariate analysis similar to a previous study by Wojciulik & Kanwisher (1999), who identified overlapping regions such as IPTO and aIPS (anterior intraparietal sulcus). Consistent with their findings, we also found overlapping regions at IPTO, FEF (frontal eye field) and SPL (superior parietal lobe) (see Supplementary Fig. 2). These results provided robust validation of our experimental design and data.

However, it is crucial to acknowledge that while univariate analysis can reveal brain regions that co-activate during FA and SA, assuming that the whole overlapping region represents a homogeneous common neural substrate underlying different attentional tasks might oversimplify the complexity of the underlying neural mechanisms. By employing multivariate machine-learning analysis, we have uncovered that even within the brain regions identified through conjunction analysis in the univariate approach, distinct patterns could exist. This highlighted the importance of adopting a multivariate approach to capture the nuanced variations and heterogeneity within these shared brain regions. Specifically, we found that the right OFG ($r = -0.49$, $p < 0.001$) and left PrG ($r = -0.39$, $p < 0.001$) exhibited the opposite patterns, whereases the right OcP showed a similar pattern ($r = 0.24$, $p = 0.02$)(Fig. 3a). On the other hand, other conjunction regions, such as bilateral IPTO (left: $r = -0.11$, $p = 0.73$; right: $r = 0.23$, $p = 0.33$), did not exhibit correlated patterns (also refer to Supplementary Table 3).

To assess the distribution of the two brain signatures across established intrinsic cortical networks, we computed voxel weight sums for each network using Yeo's atlas[33]. As depicted in Fig. 3b, the magnitude of a specific brain network was determined by the proportion of weight sums from significant voxels within that network, relative to the network with the highest

intensity. This analysis revealed distinct involvement patterns of the FAS and SAS in brain networks defined by Yeo's atlas. This quantification enabled us to assess the relative contribution of each network in terms of the weight distribution of the significant voxels. The most prominent regions associated with the FAS were the dorsal and ventral attention networks, whereas the SAS was predominantly distributed within the visual network.

In addition, conducting the prediction analysis with stepwise controlled voxel number further confirmed the dissociation among cortical networks (Fig. 3c). As the number of significant voxels in SA was considerably smaller than in FA, increasing the voxel count in the prediction models could lead to the incorporation of additional noise. This phenomenon likely explained the observed decline in performance within each network, which typically occurred at around 100 voxels.

Single-Cluster and 'Virtual Lesion' analyses revealed similar and distinct representation patterns in FA and SA, along with their signatures' distributed nature. The multivariate machine-learning analysis and the univariate analysis have revealed the distributed activations throughout the brain elicited by both FAS and SAS. Then how sufficient and necessary each cluster of the signatures was to predict FA and SA? To address the question, we conducted single-cluster analysis and 'virtual lesion' analysis. Using the watershed algorithm with the minimum size at 10 voxels (for details see methods), We first defined the clusters within the FAS and SAS . The conjunction clusters were primarily found in fronto/parietal, temporal/occipital, and limbic areas. Their prediction performance in both single-cluster and 'virtual lesion' analysis was presented in Table 1. Supplementary Table 4 showed the performance of all the clusters in FAS.

The single-cluster analysis enabled us to assess each cluster's predictability in both intra- and inter-task scenarios, and to ascertain whether a single cluster alone was sufficient for prediction. As shown in Table 1, most clusters within the FAS and SAS, except for left aSmG in SA, showed higher than chance prediction performance. This indicated that each individual cluster was sufficient for predicting outcomes in both intra-task and inter-task scenarios, and thus they exhibited somewhat similar representational patterns between FA and SA. The only exception was that four of these clusters, situated in the temporal/parietal regions, namely left Lingual Gyrus (LG), right OcP, right OFG, and left Temporal Occipital Fusiform Cortex (TOF), which displayed distinct representational patterns between FA and SA, because they showed predictive capabilities exclusively within the intra-task scenario.

In addition, the 'virtual lesion' analysis allowed us to investigate the necessity of specific brain clusters by temporarily removing the cluster from the brain signatures. If the predictability of the remaining areas within the signature dropped to chance level after removing the cluster, it indicated that the cluster was necessary for making predictions. As shown in Table 1, for both FAS and SAS, both intra-task and inter-task classification accuracy remained significantly higher than chance after removing each individual cluster. However, when comparing the prediction accuracy after removing each single cluster with that of the entire brain signature, most clusters showed significant decrease in accuracy, except for the inter-task scenario of the FAS. These findings supported that no single cluster was necessary for predicting FA or SA.

To account for the potential confounding factor of varying voxel numbers among clusters, we conducted stepwise single-cluster analysis by randomly selecting the same significant voxel number for each cluster (Fig. 4a). Regarding FA, the whole-brain signature consistently outperformed all individual clusters in both intra- and inter-task scenarios (for the complete

trajectory of the whole-brain signature, see Supplementary Fig. 3). As for SA, within the intra-task scenario, the most accurate prediction models in the single-cluster analysis were found in occipital regions, including right OcP, right OFG and right iLoC. The superior prediction performance compared to whole-brain single-cluster analysis might be attributed to the sparseness of significant voxels in the SAS, as the whole-brain SAS exhibited a prediction advantage when voxel numbers increased or in the 'virtual lesion' analysis.

In short, the single-cluster analysis and the “virtual lesion” analysis revealed both shared and distinct representation patterns in FA and SA, along with the engagement of a distributed brain network in processing both forms of attention. While most individual conjunctive clusters could partially represent both FA and SA, none of them individually matched the effectiveness of the whole-brain signature in handling attention processing. Therefore, all the clusters were sufficient but not necessary to differentiate the two conditions in both forms of attention.

To further characterize the distributed nature of FAS and SAS, we performed a voxel-level analysis, plotting the voxels of both signatures together on a normalized 2-D plane (Fig. 4b). The plane was evenly divided into octants, with voxels in each octants having the same attentional selectivity. Conjunction voxels were color-coded, FA-specific voxels were represented in light gray, and SA-specific voxels were shown in dark gray. Next, we conducted a comparative analysis of the cortical extent covered by conjunction voxels within cortical networks across the octants. The results, presented in radar plots (Fig. 4b), revealed an intriguing discovery: each octant exhibited coverage of multiple networks, once again supporting the idea of a distributed representation of attention.

Discussions

In this study, we utilized a fine-grained between-subject whole-brain multivariate machine-learning approach to explore the neural mechanisms underlying FA and SA. Our findings revealed robust brain signatures for both attentional processes respectively, which showed predictive capabilities not only within their respective tasks but also across tasks. This mutual predictivity suggested shared components between the FAS and SAS, while reduced inter-task prediction performance indicated the presence of distinct components within each signature. Moreover, we observed overlapping regions, consistent with previous univariate analyses, along with task-specific regions within the two brain signatures. Through single-cluster analysis and "virtual lesion" analysis on the overlapping clusters, we further revealed both shared and distinct representation patterns in FA and SA. This was evidenced by inter-task predictability in certain conjunctive clusters and its absence in others. Moreover, within the intra-task scenario, although each single cluster was sufficient for processing SA and FA, none of the clusters were necessary for either form of attention, supporting the distributed nature of these two brain signatures.

Unlike traditional brain mapping methods that focus on isolated core regions' activation strengths or representational patterns, such as univariate activation analyses and single-subject level multivariate machine-learning approaches (e.g., MVPA, representational similarity pattern analyses), our study employed between-subject whole-brain multivariate machine-learning approaches to construct predictive models that maps whole-brain activity patterns to specific mental states or behaviors (in this study: FA conditions, or SA conditions). Despite having higher complexity and lower interpretability, these approaches integrate neural information across multiple distributed brain networks, resulting in sensitive brain signatures with greater predictive power for mind/behavior states [30–32,34,35]. To mitigate overfitting risk and enabling the

use of between-subject whole-brain multivariate machine-learning approaches, we included a substantial sample of 235 participants for fMRI scanning during FA and SA tasks, and validated the training models on a separate hold-out test dataset. We developed sensitive and reliable brain signatures of FA and SA respectively using the SVC algorithm, and validated them using the PCA-Logistic algorithm. The findings further offered a comprehensive understanding of the neural representation of FA and SA and their relation.

Previous fMRI studies using univariate activation analyses have identified overlapping activations in fronto-parietal-temporal attention-related reegions for both FA and SA [19–21], along with common attention-dependent activity modulation in visual areas [36]. However, these findings relied on Boolean maps with p-value thresholds, considering conjunction areas as shared processing components. In addition, although single-subject level multivariate machine-learning approaches could identify searchlight or region of interests (ROIs) with different activity patterns across conditions, it lacked voxel-specific preferences. Our high spatial resolution brain signature, incorporating voxel-specific preferences, revealed common regions in dorsal and ventral attention networks and visual areas related to both attentional processes, including IPTO and SmG, consistently with previous research. Notably, uncorrelated weight patterns between FA and SA common clusters and reduced inter-task prediction accuracies, both indicat partial overlapping yet distinct neural representations, aligning with electrophysiological research [3,17,24–28]. Our study offered a comprehensive whole-brain perspective on FA and SA, surpassing local neural population or neuron-level analysis.

An additional advancement of multivariate machine-learning approaches is the development of quantitative statistical models, providing advantages over qualitative models in univariate analyses. In detail, we assessed intra- and inter-task prediction performance through AUC scores

at three levels: whole-brain, network, and cluster, enabling a robust evaluation of the similarity and distinctiveness of the two brain signatures.

At the whole-brain level, the FAS and SAS exhibited mutual predictability, albeit with reduced capacity in the inter-task scenario, suggesting the existence of both shared and distinct components. Furthermore, we evaluated network-level prediction performance using individual established cortical networks [33]. All seven networks demonstrated above-chance performance in the intra-task scenario, indicating the involvement of multiple distributed networks in both FA and SA. Specifically, the frontoparietal network exhibited the highest predictive performance for FA, while the visual network excelled in predicting SA, highlighting their respective prominence in the two attention processes.

At the finer cluster level, we employed single-cluster and 'virtual lesion' techniques to investigate the sufficiency and necessity of individual clusters for FA and SA. The single-cluster analysis revealed that all frontal/parietal clusters were sufficient in both intra- and inter-task scenarios, while four temporal/occipital clusters exhibited task-specificity. This disassociation of sufficiency implied similar top-down modulation from the frontoparietal regions, but with different modulation effects on sensory areas [17,18,18,24–28]. Additionaly, the 'virtual lesion' analysis enabled us to iteratively remove individual clusters from the whole-brain prediction model and evaluate their irreplaceable contributions. Surprisingly, no cluster was found to be necessary for FA or SA, as lesioning any cluster did not result in chance-level performance. These findings suggested a distributed brain functioning in both FA and SA. It was further supported by the voxel-level observation showing that the voxels with similar attentional selectivity spanned multiple networks.

The perspective of distributed representation of FA and SA, as revealed by machine-learning models with high predictability yet lower interpretability, challenges the traditional network-centric attention models that emphasize a few core regions, such as IPS, FEF, temporo-parietal junction (TPJ), and inferior frontal junction (IFJ), as primary centers for selective attention processing [37–39]. This idea aligns with the emerging connectome-based predictive models, which have demonstrated effective individualized predictions of attentional and other mental/behavioral states through broadly distributed across-network connections [40–43]. Both perspectives emphasize the significance of large-scale, multivariate neural networks in contributing to cognition and behavior, supporting the notion that attentional processes may rely on distributed and interconnected brain networks [44–47]. The shift from classic univariate (core region(s) corresponding to one specific behavior) to large-scale, multivariate neural network frameworks allows for advancements in linking brain elements to different cognition and behaviors [44]. Further, it facilitates the development of more efficient neural modulation techniques compared to traditional modulation techniques focused on single core regions [48,49]. These developments hold promise for enhancing or restoring attention abilities in individuals with attention-related disorders, such as autism spectrum disorder (ASD) and Attention-Deficit/Hyperactivity Disorder (ADHD).

In addition, by leveraging publicly available brain-wide gene expression atlases, such as the Allen Human Brain Atlas (AHBA) [50,51], the FAS and SAS can serve as valuable neuroimaging phenotypes for identifying attention-related gene expression. This approach opens up new possibilities to investigate the interplay between genetic expression profiles and the distributed FAS and SAS, exploring their associations with attentional states and potentially developing neurogenetic signatures of FA and SA [52,53].

In summary, we successfully identified robust whole-brain neural representations of FA and SA (i.e., FAS and SAS). These two brain signatures exhibited both partial overlap and distinct representational characteristics. Moreover, they both demonstrated distributed representations across large-scale brain networks: all frontal/parietal clusters of the signatures were sufficient for both intra- and inter-task scenarios, while most temporal/occipital clusters showed task-specificity; none were deemed necessary for either FA or SA. These comprehensive understanding and direct comparison of the neural correlates in FA and SA, shed light on the coexistence of both shared and distinct mechanisms.

Materials and Methods

Participants in the study. This study is part of an ongoing project (gene, environment, brain and behavior)[54–59]. 264 healthy participants (247 right-handed) were recruited from Beijing Normal University in this study. All had normal or corrected to normal vision. The study was approved by the Ethical Committee of the Beijing Normal University, and all participants were provided with written informed consent prior to the study. Twenty-nine participants with intensive head motion (2 mm in translation or 2mm in rotation from the first volume in any axis) were excluded from further analysis, leading to a final sample of 235 participants (136 female, $mean\ age = 21.47, SD = 1.01$).

Feature Attention Experiment. The experiment was block-designed with 6 blocks in total (Fig. 1a), three for feature task and three for conjunction task. The total block time was 32 s. The block ended with a fixation of 15 s. The order of the blocks was counterbalanced, with either feature task as the leading block and intercepted by conjunction task (F C F C F C, F: feature task, C: conjunction task) or conjunction task as the leading block (C F C F C F).

Each block began with an instruction lasting 4.5 s indicating the target for the block. The sequence of letters (with size of $7.5^{\circ} * 8.5^{\circ}$ of visual angle) was displayed at the center of the screen against a gray background. Each trial lasted 1 s with 6 letters in repaid serial visual presentation (RSVP) and with no interstimulus interval (ISI). The difference between the target and distractors could be their color, shape, or pattern. All the stimuli were either white or black. The shape of the stimuli consisted seven letters: E, H, L, O, S, T, and X, while the patterns of the stimuli came with four types: left-stripped, right-stripped, checked and wall-pattern. We designed two counterbalanced versions of RSVP. In the first version, all the distractors were black; the target in feature task was “white checked O” and the target in conjunction task was “black checked S”. In the second version, all the distractors were white; the target in feature task was “black checked O”, and the target in conjunction task was “white checked S”. This design enabled us to balance out the confounding factor of contrast while keeping both targets and distractors constant across versions. The target occurred on half of the trials with two constraints: (1) No consecutive stimuli in a RSVP were the same letter. (2) No more than three consecutive trials contained a target.

Spatial Attention Experiment. This experiment consisted of 18 blocks (Fig. 1b). The first and the last block were the fixation blocks which lasted 15 s. The remaining 16 blocks comprised spatial attention tasks: 8 blocks for central task and 8 blocks for peripheral task, depending on whether the target was the central dot or one of the peripheral dots. The 8 peripheral task blocks were further divided into 4 blocks for the left visual field condition and 4 blocks for the right visual field condition. Subjects first completed the central tasks then the peripheral tasks. The order of the left and right visual field blocks was counterbalanced. Each block began with a 3-s

instruction, indicating whether the current block was a central task or a peripheral task, then followed by 20 trials. The total duration of each block was 19 s.

One dot was displayed at the center of the screen and eight peripheral dots were placed evenly on a hypothetical circle whose diameter was 15°. Against a dark gray background, the presented dots were light gray. Among them, a colored cross was located on the central dot, which had one gray arm and the other arm was half green and half red. The colored cross rotated every 4 s counterclockwise. In each block, the cross rotated 4 times under each visual field condition, where for the left visual field blocks, the red part of the colored cross always pointed at the dots in left visual field, and vice versa.

Except for the cross, the entire display blinked on and off every 400 ms as one trial. In each display that blinked on, there was always a smaller dot on the display, either at the center or one of the four dots defined by the visual field condition. In one of the five blinks, the dot at the attended area, which was indicated by the red part of the colored cross in the peripheral task blocks or the central dot in the central task block, became smaller and reverted back to its original size on the next blink. The diameter of a normal dot was 1.3°. The diameter of the smaller dot was 0.8° when it was the central dot and 1.0° when it was at the peripheral area. The task for participants was to press the key when the dot in the attended area became smaller.

fMRI Data Acquisition. fMRI Scanning was performed on a 3T Siemens Trio scanner (Erlangen, Germany) at the BNU Imaging Center for Brain Research, Beijing, China. Functional images were acquired using a standard 12-channel head matrix coil, with a T2*-weighted gradient-echo based echo-planar imaging sequence (30 contiguous interleaved 4.8 mm axial slices, $repetition\ time\ (TR) = 2000\ ms$, $echo\ time\ (TE) = 30\ ms$, $voxel\ size = 3.125 \times 3.125 \times 4.8\ mm^3$, $field\ of\ view\ (FOV) = 64 \times 64\ mm^2$, $flip\ angle = 90^{\circ}$,

$inter - slice\ gap\ =\ 0.8\,mm$). High-resolution structural images were acquired using a T1-weighted 3D magnetization-prepared rapid gradient-echo (3D MP-RAGE) sequence (176 slices, $TR\ =\ 2530\,ms, TE\ =\ 3.39\,ms, inversion\ time\ =\ 1100\,ms, voxel\ size\ =\ 1 \times 1 \times 1.33\,mm^3, FOV\ =\ 256\ \times\ 256\,mm^2, flip\ angle\ =\ 7^{\circ}$). Head motion was constrained using foam pillows and extendable padded head clamps, while scanner noise was reduced by wearing earplugs.

fMRI Data Preprocessing. All images were preprocessed using SPM12 (Welcome Trust Centre for Neuroimaging, http://www.fil.ion.ucl.ac.uk/spm/software/spm12/), FSL (FMRIB's Software Library, http://www.fmrib.ox.ac.uk/fsl) and custom MATLAB functions. Preprocessing consisted of slice timing, motion correction, spatial smoothing with a Gaussian kernel (6 mm full-width at half-maximum), and high-pass temporal filtering with a $128s$ cutoff. For the first-level analysis, the first four volumes of each run were discarde.

First-level general linear model (GLM) analyses were conducted using SPM12. For the FA experiment, the two experimental conditions (feature condition, conjunction condition) were implemented as box-car regressors with an epoch length of 32 s, convolved with the canonical hemodynamic response function. Additionally, six motion parameters (three translation parameters and three rotation parameters) were included in the GLM as confounding variables to account for residual head motion artifacts. The GLM for the SA experiment closely mirrored that of the FA, except that the two experimental conditions (central condition, peripheral condition) were implemented as box-car regressors with an epoch length of 16 s.

Between-subject Whole-brain Multivariate Machine-learning Analysis. We conducted a between-subject whole-brain multivariate machine-learning analysis to investigate the neural signatures of FA and SA using individuals' beta maps derived from the first-level GLM analysis.

Each individual had four beta maps, one for each attention condition (i.e., 2 conditions for FA, 2 conditions for SA). To ensure that every voxel in the prediction model possesses a numerical value, we constructed a comprehensive union mask comprising solely those voxels exhibiting beta weights across all four beta maps. This union mask was then employed to mask the beta maps, rendering them suitable for subsequent multivariate pattern analysis.

For the training and testing data split, we randomly divided $1/3$ of the participants ($n = 78$) as the testing dataset, using only the remaining $2/3$ of the participants ($n = 157$) as the training dataset. We used support vector classification algorithm (SVC) with the linear kernel ($c = 1$) implemented by *scikit-learn* [60] that utilizes beta maps as the training features and experiment condition as prediction targets (For FA, feature condition is denoted as 0 and conjunction condition is denoted as 1; for SA, central condition is denoted as 0 and peripheral condition is denoted as 1). Employing a linear kernel provides specific benefits, among which is its interpretability in deriving beta weights for each voxel, facilitating the prediction of the task being investigated. In addition, we also employed a PCA-Logistic algorithm [30] to validate the reliability of the results of the SVC analysis. In the PCA-Logistic method, we first used principal component analysis (PCA) to reduce the dimension of features to the number of samples (i.e., reduced to 314 features as 157×2 ($number\ of\ participants \times\ number\ of\ task\ type$)). Then we used logistic regression for binary classification. The following analysis of the study was based on the beta maps from SVC analysis.

To mitigate the potential bias associated with train-validation splits, we implemented a meticulous repeated 10×10-fold procedure. Within each training iteration, the complete training dataset was partitioned into 10 subsets, with 9 subsets allocated for training purposes while the remaining subset served as the validation set. This process was iterated 10 times, with

the training data being re-segmented for each repetition. Notably, this meticulous procedure was repeated 10 times, ensuring robustness in our analyses. For the two attention experiments conducted, we acquired two distinct beta maps, each elucidating the predictive capacity of every voxel.

Model Evaluation. To assess the performance of our model, we employed a diverse range of metrics, including accuracy, precision, recall, f-1 measure, and the Area under the Receiver Operating Characteristic Curve (AUC). The AUC's scale-invariant and classification-threshold-invariant nature makes it the most suitable metric for quantifying and evaluating the effectiveness of our model. Thus, we placed particular importance on AUC as the primary evaluation measure for the model's performance among all the metrics. Importantly, we measured the model's performance on the hold-out test data ($n = 78$), which had been allocated prior to the training phase.

Determining Predictive Voxels. In order to identify the voxels that consistently demonstrate predictive power for attentional tasks, we employed a subject-level bootstrap resampling approach. This involved generating 1,000 new training datasets by sampling with replacement. The same training procedure was then applied to each new training dataset. Consequently, for each voxel, we obtained a distribution comprising 1,000 samples. To determine the significance of voxel predictability, we conducted a two-sided one-sample t-test with the null hypothesis set at 0. To account for multiple comparisons, we applied a false discovery rate (FDR) correction with a threshold of $q < 0.05$. This rigorous statistical procedure enabled us to pinpoint the voxels that exhibit consistent and reliable predictive capabilities for attentional tasks.

Univariate Analysis. To provide a benchmark for our multivariate pattern analysis, we performed univariate analysis. For each type of attentional experiment (FA and SA), we conducted two-sided matched-pair t-test on the entire sample of subjects ($n = 235$), and employed FDR correction $q < 0.05$ to control for multiple comparisons. We calculated Cohen's d as an effect size measure, which served as the criteria for selecting predictive voxels ($Cohen's\ d > 0.5$, denoting moderate or greater effect size).

Network Level Analysis. To discern the network-level cortical mapping distinctions between the two attentional tasks, we employed Yeo's Atlas[33] which leveraged resting-state functional connectivity MRI to identify seven networks comprising functionally coupled regions across the cerebral cortex. Following FDR correction, all significant voxels were classified into one of these seven distinct networks as defined by Yeo's 1-mm liberal mask. To quantify the intensity of predictability associated with that particular network, we summed up the absolute value of all the significant voxels that fell within the network. For comparative analysis across different attentional tasks, we standardized the intensity values by expressing them as a percentage relative to the network with the highest intensity for the specific task under investigation. This approach resulted in an intensity value at 1 for one network in the respective task, with other networks as the relative intensity to the target network, which ensured consistency and enabled meaningful comparisons of intensity values across diverse attentional tasks.

To further assess the predictability of each network in both intra- and inter-task scenarios, we implemented a stepwise prediction approach that effectively controlled for the confounding influence stemming from voxel count disparities among networks. Starting with 10 voxels, we progressively increased the number of voxels included in the prediction model until all voxels

within the network of interest were incorporated. At each step, a fixed number of voxels within the network were randomly selected as features for training SVCs to predict the conditions of either FA or SA. The performance of the models was then evaluated on the hold-out test dataset for both intra-task and inter-task scenarios. The training and testing protocols followed the same methodology as employed in the whole-brain analysis. To ensure robustness of our findings, we repeated the training and testing procedures 100 times at each step.

Cluster-level Analysis. We used the watershed image segmentation algorithm[58,61,62] to partition the significant voxels, after FDR correction, into separate clusters within the beta maps. By isolating these clusters, we were able to delve into their individual contributions to attentional processes. Specifically, we utilized single-cluster prediction analysis and 'virtual lesion' analysis to examine the predictive sufficiency and necessity of individual clusters in attentional tasks, respectively.

In the single-cluster prediction analysis, we focused on isolating the cluster of interest and applied the same model training procedures employed in the whole-brain analysis. Subsequently, we assessed the model's performance in both intra- and inter-task scenarios. If the cluster of interest is indeed sufficient in predicting attentional tasks, we would anticipate an AUC score exceeding chance level. This analysis provides valuable insights into the independent predictive contribution of the isolated cluster in attentional processing.

In the 'virtual lesion' study, we performed an analysis wherein we excluded the cluster of interest from the whole-brain signature and proceeded to retrain the classifier. If the performance of the newly trained prediction model drops to chance level, it provides compelling evidence that the lesioned cluster plays a necessary role in the specific attentional task.

Again, we implemented a gradual single-cluster prediction and ‘virtual lesion’ analysis with a step size of 10 voxels, to exclude the potential confounding effect of voxel number variations across clusters when comparing their influence on the change in model predictability. In each step, we randomly selected the same number of voxels from each cluster, and this process was repeated 1000 times to retrain the prediction models. This procedure allowed us to ensure comparable model complexity across different clusters, effectively controlling for any voxel number variations. The stepwise study also incorporated the analysis of whole-brain signatures.

Voxel-level Analysis. To investigate the joint normalized distribution of the neural signatures of FA and SA, we utilized a visualization technique to represent them in a 2-D space. To ensure comparability among voxels within neural signatures, we first normalized the weights within each signature independently. Thus, for each voxel, the x-axis represented its normalized weight for FA, and the y-axis represented its normalized weight for SA. Next, we divided the 2-D space into octants, with each two octants corresponding to a distinct conceptual category: FA dominant, SA dominant, same pattern, and opposite pattern.

To incorporate the cortical relationship among these octants, we calculated the coverage ratio of significant voxels falling into their respective networks. These ratios were then normalized by expressing them as proportions relative to the highest coverage ratio observed within the network. This normalization procedure allowed us to account for variations in the coverage across different networks and facilitated fair comparisons among them. By employing this approach, we gained insights into the distribution patterns and cortical relationships among the various categories within the joint FA and SA signatures.

Acknowledgments. This work was supported by the STI2030-Major Projects (2021ZD0203803), National Natural Science Foundation of China (32200840), National Key R&D Program of China (2019YFA0709503), China Postdoctoral Science Foundation (2022T150061, 2022M710435), and Fundamental Research Funds for the Central Universities.

Code and data availability. Code and data used in the analysis are available on OSF: https://osf.io/s8wvb/ .

Conflicts of Interest. The authors report no biomedical financial interests or potential conflicts of interest.

Reference:

1. Luck, S. J., Chelazzi, L., Hillyard, S. A. & Desimone, R. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* **77**, 24–42 (1997).
2. McAdams, C. J. & Maunsell, J. H. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J. Neurosci.* **19**, 431–441 (1999).
3. Bichot, N. P., Heard, M. T., DeGennaro, E. M. & Desimone, R. A source for feature-based attention in the prefrontal cortex. *Neuron* **88**, 832–844 (2015).
4. Liu, T. Feature-based attention: effects and control. *Curr. Opin. Psychol.* **29**, 187–192 (2019).
5. Maunsell, J. H. & Treue, S. Feature-based attention in visual cortex. *Trends Neurosci.* **29**, 317–322 (2006).
6. Baldauf, D. & Desimone, R. Neural mechanisms of object-based attention. *Science* **344**, 424–427 (2014).
7. Chen, Z. Object-based attention: A tutorial review. *Atten. Percept. Psychophys.* **74**, 784–802 (2012).
8. Roelfsema, P. R., Lamme, V. A. & Spekreijse, H. Object-based attention in the primary visual cortex of the macaque monkey. *Nature* **395**, 376–381 (1998).
9. Correa, Á., Sanabria, D., Spence, C., Tudela, P. & Lupiáñez, J. Selective temporal attention enhances the temporal resolution of visual perception: Evidence from a temporal order judgment task. *Brain Res.* **1070**, 202–205 (2006).

10. Coull, J. T. & Nobre, A. C. Where and when to pay attention: the neural systems for directing attention to spatial locations and to time intervals as revealed by both PET and fMRI. *J. Neurosci.* **18**, 7426–7435 (1998).
11. Denison, R. N., Carrasco, M. & Heeger, D. J. A dynamic normalization model of temporal attention. *Nat. Hum. Behav.* **5**, 1674–1685 (2021).
12. Anderson, B. A. Neurobiology of value-driven attention. *Curr. Opin. Psychol.* **29**, 27–33 (2019).
13. Anderson, B. A., Laurent, P. A. & Yantis, S. Value-driven attentional capture. *Proc. Natl. Acad. Sci.* **108**, 10367–10371 (2011).
14. Moore, T. & Zirnsak, M. The what and where of visual attention. *Neuron* **88**, 626–628 (2015).
15. Park, G. Y. *et al.* Neural correlates of spatial and nonspatial attention determined using intracranial electroencephalographic signals in humans. *Hum. Brain Mapp.* **37**, 3041–3054 (2016).
16. Yantis, S. & Serences, J. T. Cortical mechanisms of space-based and object-based attentional control. *Curr. Opin. Neurobiol.* **13**, 187–193 (2003).
17. McAdams, C. J. & Maunsell, J. H. R. Attention to Both Space and Feature Modulates Neuronal Responses in Macaque Area V4. *J. Neurophysiol.* **83**, 1751–1755 (2000).
18. Patzwahl, D. R. & Treue, S. Combining spatial and feature-based attention within the receptive field of MT neurons. *Vision Res.* **49**, 1188–1193 (2009).
19. Galashan, D. & Siemann, J. Differences and Similarities for Spatial and Feature-Based Selective Attentional Orienting. *Front. Neurosci.* **11**, 283 (2017).

20. Giesbrecht, B., Woldorff, M. G., Song, A. W. & Mangun, G. R. Neural mechanisms of top-down control during spatial and feature attention. *NeuroImage* **19**, 496–512 (2003).
21. Wojciulik, E. & Kanwisher, N. The Generality of Parietal Involvement in Visual Attention. *Neuron* **23**, 747–764 (1999).
22. White, A. L., Rolfs, M. & Carrasco, M. Stimulus competition mediates the joint effects of spatial and feature-based attention. *J. Vis.* **15**, 7–7 (2015).
23. Hayden, B. Y. & Gallant, J. L. Combined effects of spatial and feature-based attention on responses of V4 neurons. *Vision Res.* **49**, 1182–1187 (2009).
24. David, S. V., Hayden, B. Y., Mazer, J. A. & Gallant, J. L. Attention to Stimulus Features Shifts Spectral Tuning of V4 Neurons during Natural Vision. *Neuron* **59**, 509–521 (2008).
25. Cohen, M. R. & Maunsell, J. H. R. Using Neuronal Populations to Study the Mechanisms Underlying Spatial and Feature Attention. *Neuron* **70**, 1192–1204 (2011).
26. Hayden, B. Y. & Gallant, J. L. Time Course of Attention Reveals Different Mechanisms for Spatial and Feature-Based Attention in Area V4. *Neuron* **47**, 637–643 (2005).
27. Ni, A. M. & Maunsell, J. H. R. Neuronal Effects of Spatial and Feature Attention Differ Due to Normalization. *J. Neurosci.* **39**, 5493–5505 (2019).
28. Goddard, E., Carlson, T. A. & Woolgar, A. Spatial and Feature-selective Attention Have Distinct, Interacting Effects on Population-level Tuning. *J. Cogn. Neurosci.* **34**, 290–312 (2022).
29. Bengson, J. J. & Mangun, G. R. Spatial attention and feature-based attention are differentially sensitive to individual working memory capacity and perceptual load. *Vis. Cogn.* **26**, 545–551 (2018).

30. Chang, L. J., Gianaros, P. J., Manuck, S. B., Krishnan, A. & Wager, T. D. A Sensitive and Specific Neural Signature for Picture-Induced Negative Affect. *PLOS Biol.* **13**, e1002180 (2015).

31. Krishnan, A. *et al.* Somatic and vicarious pain are represented by dissociable multivariate brain patterns. *eLife* **5**, e15166 (2016).

32. Zhou, F. *et al.* A distributed fMRI-based signature for the subjective experience of fear. *Nat. Commun.* **12**, 6643 (2021).

33. Thomas Yeo, B. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165 (2011).

34. Kragel, P. A., Koban, L., Barrett, L. F. & Wager, T. D. Representation, Pattern Information, and Brain Signatures: From Neurons to Neuroimaging. *Neuron* **99**, 257–273 (2018).

35. Wager, T. D. *et al.* An fMRI-Based Neurologic Signature of Physical Pain. *N. Engl. J. Med.* **368**, 1388–1397 (2013).

36. Van Es, D. M., Theeuwes, J. & Knapen, T. Spatial sampling in human visual cortex is modulated by both spatial and feature-based attention. *eLife* **7**, e36928 (2018).

37. Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* **3**, 201–215 (2002).

38. Corbetta, M. & Shulman, G. L. Spatial Neglect and Attention Networks. *Annu. Rev. Neurosci.* **34**, 569–599 (2011).

39. Petersen, S. E. & Posner, M. I. The Attention System of the Human Brain: 20 Years After. *Annu. Rev. Neurosci.* **35**, 73–89 (2012).

40. Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664–1671 (2015).

41. Gao, M. *et al.* Connectome-based models can predict processing speed in older adults. *NeuroImage* **223**, 117290 (2020).

42. Rosenberg, M. D. *et al.* A neuromarker of sustained attention from whole-brain functional connectivity. *Nat. Neurosci.* **19**, 165–171 (2016).

43. Wu, E. X. W. *et al.* Overlapping attentional networks yield divergent behavioral predictions across tasks: Neuromarkers for diffuse and focused attention? *NeuroImage* **209**, 116535 (2020).

44. Bassett, D. S. & Sporns, O. Network neuroscience. *Nat. Neurosci.* **20**, 353–364 (2017).

45. Mišić, B. & Sporns, O. From regions to connections and networks: new bridges between brain and behavior. *Curr. Opin. Neurobiol.* **40**, 1–7 (2016).

46. Steinmetz, N. A., Zatka-Haas, P., Carandini, M. & Harris, K. D. Distributed coding of choice, action and engagement across the mouse brain. *Nature* **576**, 266–273 (2019).

47. Urai, A. E., Doiron, B., Leifer, A. M. & Churchland, A. K. Large-scale neural recordings call for new insights to link brain and behavior. *Nat. Neurosci.* **25**, 11–19 (2022).

48. Ching, S., Brown, E. N. & Kramer, M. A. Distributed control in a mean-field cortical network model: Implications for seizure suppression. *Phys. Rev. E* **86**, 021920 (2012).

49. Gu, S. *et al.* Controllability of structural brain networks. *Nat. Commun.* **6**, 8414 (2015).

50. Hawrylycz, M. J. *et al.* An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012).

51. Shen, E. H., Overly, C. C. & Jones, A. R. The Allen Human Brain Atlas. *Trends Neurosci.* **35**, 711–714 (2012).

52. Bueichekú, E. *et al.* Central neurogenetic signatures of the visuomotor integration system. *Proc. Natl. Acad. Sci.* **117**, 6836–6843 (2020).

53. Burt, J. B. *et al.* Hierarchy of transcriptomic specialization across human cortex captured by structural neuroimaging topography. *Nat. Neurosci.* **21**, 1251–1259 (2018).

54. Huang, L. *et al.* Individual differences in cortical face selectivity predict behavioral performance in face recognition. *Front. Hum. Neurosci.* **8**, 483 (2014).

55. Wang, X. *et al.* The hierarchical structure of the face network revealed by its functional connectivity pattern. *J. Neurosci.* **36**, 890–900 (2016).

56. Wang, X., Zhu, Q., Song, Y. & Liu, J. Developmental reorganization of the core and extended face networks revealed by global functional connectivity. *Cereb. Cortex* **28**, 3521–3530 (2018).

57. Wang, Y., Song, Y., Li, X., Zhang, L. & Liu, J. Influence of parental care on offspring hippocampal volume in young adults varies as a function of overprotection. *Sci. Rep.* **7**, 46429 (2017).

58. Zhen, Z. *et al.* Quantifying interindividual variability and asymmetry of face-selective regions: A probabilistic functional atlas. *NeuroImage* **113**, 13–25 (2015).

59. Zhou, L., Zhen, Z., Liu, J. & Zhou, K. Brain structure and functional connectivity associated with individual differences in the attentional blink. *Cereb. Cortex* **30**, 6224–6237 (2020).

60. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

61. Fedorenko, E., Hsieh, P.-J., Nieto-Castañón, A., Whitfield-Gabrieli, S. & Kanwisher, N. New Method for fMRI Investigations of Language: Defining ROIs Functionally in Individual Subjects. *J. Neurophysiol.* **104**, 1177–1194 (2010).

62. Meyer, F. Topographic distance and watershed lines. *Signal Process.* **38**, 113–125 (1994).

Figures

Table 1

FA

Clusters	Number of Voxels	MNI Peak Coordinates			z-score	Single-cluster		Virtual Lesion	
		x	y	z	Mean(SD)	Intra-task	Inter-task	Intra-task	Inter-task
Whole Brain	149270					0.99[0.97, 1.00]; (***)(†††)	0.66[0.58, 0.73]; (***)(†††)		
Fronto/Parietal Regions									
left IPTO	3343	-20	-74	48	1.86(0.77)	0.89[0.84, 0.93]; (***)(†††)	0.59[0.53, 0.65]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.66[0.59, 0.73]; (***)(†††); (‡‡‡)(NP: 0.05)
right IPTO	1711	26	-58	46	1.64(0.50)	0.83[0.78, 0.88]; (***)(†††)	0.60[0.54, 0.65]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.66[0.59, 0.73]; (***)(†††); (NS: 1.00)(NP: -0.04)
left PrG	1646	-36	-6	50	1.61(0.54)	0.88[0.84, 0.93]; (***)(†††)	0.59[0.53, 0.64]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(NP: 0.18)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.27)
right PcC	874	12	-60	46	1.50(0.38)	0.82[0.77, 0.87]; (***)(†††)	0.62[0.57, 0.67]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.67[0.60, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.35)
left MFG	689	-42	28	30	1.33(0.40)	0.80[0.74, 0.86]; (***)(†††)	0.58[0.52, 0.64]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.22)
right FP	316	28	40	42	1.28(0.31)	0.78[0.72, 0.85]; (***)(†††)	0.55[0.51, 0.60]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.30)
left POpC	244	-46	-28	16	1.74(0.42)	0.63[0.56, 0.69]; (***)(†††)	0.58[0.54, 0.63]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.23)
right COpC	202	40	-14	22	1.47(0.39)	0.70[0.64, 0.77]; (***)(†††)	0.57[0.53, 0.61]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.27)
left aSmG	122	-60	-34	40	1.59(0.30)	0.80[0.74, 0.85]; (***)(†††)	0.53[0.49, 0.58]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.23)
left PcC	122	-8	-62	44	1.39(0.20)	0.71[0.65, 0.77]; (***)(†††)	0.54[0.50, 0.58]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.67[0.60, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.30)
Temporal/Occipital Regions									
left LG	807	-26	-54	-8	1.60(0.41)	0.75[0.70, 0.80]; (***)(†††)	0.48[0.44, 0.52]; (NS: 1.00)(NP: -1.21)	0.99[0.98, 1.00]; (***)(†††); (‡‡)(NP: 0.04)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.34)
right OcP	336	28	-94	12	1.52(0.35)	0.65[0.62, 0.69]; (***)(†††)	0.45[0.41, 0.49]; (NS: 1.00)(NP: -3.13)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.24)
right pSTG	297	50	-30	4	1.45(0.23)	0.60[0.53, 0.67]; (***)(†††)	0.54[0.50, 0.58]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.23)
right iLOC	249	32	-84	-2	1.92(0.33)	0.61[0.57, 0.65]; (***)(†††)	0.52[0.47, 0.56]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§§)	0.66[0.59, 0.73]; (***)(†††); (NS: 1.00)(NP: -0.07)
right OFG	193	20	-86	-6	1.72(0.31)	0.72[0.66, 0.77]; (***)(†††)	0.47[0.41, 0.53]; (NS: 1.00)(NP: -1.22)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§§)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.28)
left TOF	11	-40	-58	-14	1.35(0.14)	0.68[0.62, 0.74]; (***)(†††)	0.43[0.39, 0.48]; (NS: 1.00)(NP: -4.17)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.24)
Limbic Regions									
right Ins	1183	36	-20	0	1.60(0.47)	0.89[0.85, 0.93]; (***)(†††)	0.61[0.55, 0.67]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§§)	0.66[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.19)
left PcG	664	-2	14	52	1.75(0.48)	0.84[0.79, 0.89]; (***)(†††)	0.57[0.52, 0.62]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.26)
right aCG	225	0	14	22	1.73(0.50)	0.80[0.74, 0.86]; (***)(†††)	0.57[0.53, 0.61]; (***)(†††)	0.99[0.98, 1.00]; (***)(†††); (‡‡‡)(§)	0.67[0.59, 0.74]; (***)(†††); (NS: 1.00)(NP: -0.26)

SA

Clusters	Number of Voxels	MNI Peak Coordinates			z-score	Single-cluster		Virtual Lesion	
		x	y	z	Mean(SD)	Intra-task	Inter-task	Intra-task	Inter-task
Whole Brain	149270					0.99[0.97, 1.00]; (***)(†††)	0.66[0.58, 0.73]; (***)(†††)		
Fronto/Parietal Regions									
left aSmG	58	-60	-32	40	3.30(0.63)	0.50[0.46, 0.54]; (NS: 0.09)(NP: 0.02)	0.63[0.56, 0.70]; (***)(†††)	0.64[0.59, 0.70]; (***)(†††); (‡‡‡)(§§§)	0.60[0.55, 0.65]; (***)(†††); (NS: 1.00)(NP: -0.08)
left PcC	41	-2	-68	60	2.39(0.36)	0.63[0.56, 0.71]; (***)(†††)	0.55[0.48, 0.63]; (***)(†††)	0.64[0.59, 0.70]; (***)(†††); (‡‡‡)(§§§)	0.58[0.52, 0.63]; (***)(†††); (‡‡‡)(§§§)
left PcG	37	-8	20	32	2.68(0.34)	0.59[0.54, 0.63]; (***)(†††)	0.57[0.49, 0.64]; (***)(†††)	0.65[0.59, 0.70]; (***)(†††); (‡‡‡)(§§§)	0.59[0.53, 0.65]; (***)(†††); (‡‡‡)(§)
right FP	21	34	40	38	2.42(0.12)	0.53[0.48, 0.58]; (***)(†††)	0.61[0.54, 0.69]; (***)(†††)	0.65[0.60, 0.71]; (***)(†††); (‡‡‡)(§§§)	0.58[0.52, 0.63]; (***)(†††); (‡‡‡)(§§§)
right IPTO	20	24	-78	36	2.05(0.21)	0.61[0.55, 0.66]; (***)(†††)	0.55[0.49, 0.61]; (***)(†††)	0.65[0.59, 0.70]; (***)(†††); (‡‡‡)(§§§)	0.57[0.51, 0.63]; (***)(†††); (‡‡‡)(§§§)
left POpC	19	-48	-22	12	2.34(0.22)	0.57[0.52, 0.62]; (***)(†††)	0.61[0.53, 0.68]; (***)(†††)	0.66[0.60, 0.71]; (***)(†††); (‡‡‡)(§§)	0.57[0.52, 0.63]; (***)(†††); (‡‡‡)(§§§)
left MFG	16	-38	2	42	2.47(0.19)	0.55[0.50, 0.61]; (***)(†††)	0.68[0.62, 0.74]; (***)(†††)	0.65[0.59, 0.71]; (***)(†††); (‡‡‡)(§§§)	0.58[0.52, 0.63]; (***)(†††); (‡‡‡)(§§§)
right COpC	16	48	-16	18	2.53(0.16)	0.60[0.54, 0.66]; (***)(†††)	0.72[0.66, 0.78]; (***)(†††)	0.65[0.59, 0.71]; (***)(†††); (‡‡‡)(§§§)	0.57[0.51, 0.62]; (***)(†††); (‡‡‡)(§§§)
left IPTO	13	-22	-68	32	2.54(0.12)	0.57[0.53, 0.60]; (***)(†††)	0.61[0.56, 0.66]; (***)(†††)	0.66[0.60, 0.71]; (***)(†††); (‡‡‡)(§§)	0.59[0.53, 0.64]; (***)(†††); (‡‡‡)(§§)
right PcC	11	12	-54	32	1.97(0.23)	0.58[0.53, 0.63]; (***)(†††)	0.70[0.63, 0.77]; (***)(†††)	0.65[0.59, 0.71]; (***)(†††); (‡‡‡)(§§§)	0.58[0.52, 0.63]; (***)(†††); (‡‡‡)(§§§)
Temporal/Occipital Regions									
right OcP	362	24	-92	4	3.46(1.00)	0.70[0.63, 0.77]; (***)(†††)	0.45[0.42, 0.49]; (NS: 1.00)(NP: -3.29)	0.64[0.59, 0.70]; (***)(†††); (‡‡‡)(§§§)	0.60[0.54, 0.66]; (***)(†††); (NS: 1.00)(NP: -0.10)
right iLOC	276	30	-86	2	2.56(0.47)	0.68[0.61, 0.74]; (***)(†††)	0.51[0.48, 0.55]; (***)(†††)	0.66[0.60, 0.72]; (***)(†††); (‡‡‡)(§§)	0.57[0.52, 0.63]; (***)(†††); (‡‡‡)(§§§)
right OFG	197	22	-90	-8	3.33(0.86)	0.71[0.64, 0.78]; (***)(†††)	0.47[0.42, 0.52]; (NS: 1.00)(NP: -1.90)	0.65[0.59, 0.70]; (***)(†††); (‡‡‡)(§§§)	0.59[0.53, 0.65]; (***)(†††); (‡‡‡)(§)

right pSTG	34	52	-18	-2	3.16(0.55)	0.54[0.50, 0.58]; (***)(†††)	0.64[0.57, 0.71]; (***)(†††)	0.64[0.59, 0.69]; (***)(†††); (‡‡‡)(§§§)	0.54[0.48, 0.59]; (***)(†††); (‡‡‡)(§§§)
left TOF	18	-40	-56	-10	2.15(0.14)	0.54[0.49, 0.59]; (***)(†††)	0.39[0.33, 0.45]; (NS: 1.00)(NP: -5.11)	0.65[0.59, 0.70]; (***)(†††); (‡‡‡)(§§§)	0.58[0.52, 0.63]; (***)(†††); (‡‡‡)(§§§)
left LG	15	-10	-76	0	1.87(0.16)	0.54[0.50, 0.58]; (***)(†††)	0.26[0.20, 0.32]; (NS: 1.00)(NP: -11.49)	0.65[0.59, 0.71]; (***)(†††); (‡‡‡)(§§§)	0.58[0.52, 0.63]; (***)(†††); (‡‡‡)(§§§)
Limbic Regions									
left PcG	37	-8	20	32	2.68(0.34)	0.59[0.54, 0.63]; (***)(†††)	0.57[0.49, 0.64]; (***)(†††)	0.65[0.59, 0.70]; (***)(†††); (‡‡‡)(§§§)	0.59[0.53, 0.65]; (***)(†††); (‡‡‡)(§)
right aCG	34	6	12	26	3.38(0.53)	0.57[0.53, 0.62]; (***)(†††)	0.60[0.52, 0.67]; (***)(†††)	0.65[0.60, 0.71]; (***)(†††); (‡‡‡)(§§)	0.57[0.52, 0.63]; (***)(†††); (‡‡‡)(§§§)
right Ins	11	38	-10	2	2.92(0.28)	0.58[0.53, 0.63]; (***)(†††)	0.66[0.59, 0.73]; (***)(†††)	0.65[0.59, 0.70]; (***)(†††); (‡‡‡)(§§§)	0.57[0.51, 0.63]; (***)(†††); (‡‡‡)(§§§)

Table 1. Single-cluster and 'Virtual Lesion' Analysis. We used the watershed algorithm to parcel clusters that exceeded the size of 100 voxels for FA and 10 voxels for SA. 'Intra-task' refers to testing the model on the same task it was trained for. 'Inter-task' refers to testing the model on a different task that it has not been trained on. 'Whole Brain' is the uncorrected beta map acquired directly from SCV training. Model performance was measured by AUC. We constructed 95% CI through bootstrap over 10,000 times. For single-cluster analysis, model performance was evaluated with one-sample one-tailed t-test with null hypothesis at 0.5. For 'virtual lesion' analysis, the impact of the virtual lesion was assessed through both model performance and the reduction in model performance compared to the whole-brain condition.

Asterisks indicate the p-value of statistical inference. *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$; NS, not significant, $p > 0.05$, with the specific p-value indicated. The null hypothesis assumes the model performance in AUC is at 0.5.

Daggers indicate the effect size of statistical inference. †, $Cohen's\ d\ < 0.5$, ††, $Cohen's\ d < 0.8$, †††, $Cohen's\ d\ \geq 0.8$; NP, not practical, $Cohen's\ d\ < 0.2$, with the value of Cohen's d indicated.

Double daggers indicate the p-value of statistical inference in 'virtual lesion' analysis. ‡, $p < 0.05$; ‡‡, $p < 0.01$; ‡‡‡, $p < 0.001$; NS, not significant, $p > 0.05$, with the specific p-value indicated. The null hypothesis posits that there is no model performance drop (ΔAUC=0).

Section markers indicate the effect size of the performance drop in 'virtual lesion' analysis. §, $Cohen's\ d\ < 0.5$, §§, $Cohen's\ d < 0.8$, §§§, $Cohen's\ d\ \geq 0.8$; NP, not practical, $Cohen's\ d\ < 0.2$, with the value of Cohen's d indicated.

Fig. 1

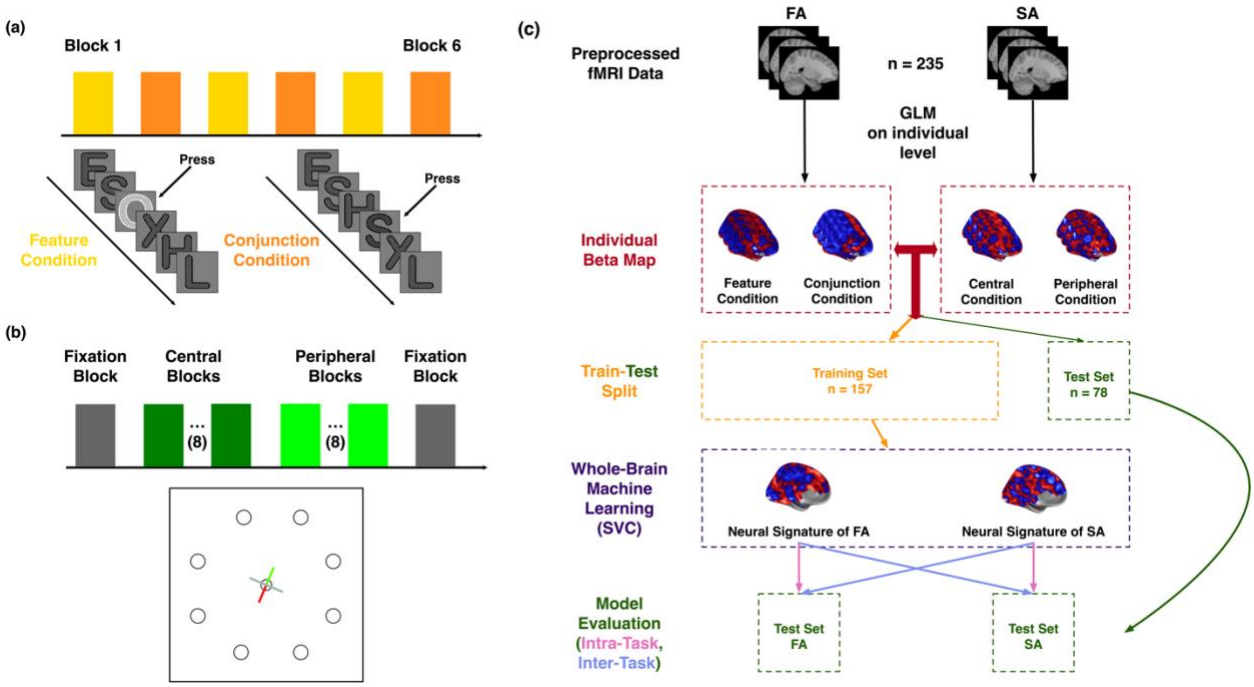

Fig. 1 Experimental Paradigms and Analysis Pipeline. The participants took part in both FA and SA experiments. Both experiments were block designed with two conditions. **(a)** In the FA experiment, serials of letters were rapidly presented. In the feature condition, participants were instructed to press the key when a white letter was presented, while ignoring other letters that were black. In the conjunction condition, participants were required to attend to both the shape and the pattern of the letter, pressing the key only when 'black checked S' was presented. **(b)** In the SA experiment, participants were instructed to press a key when the dot within the attended area decreased in size while maintaining fixation at the center of the screen. The attended area was manipulated between central and peripheral blocks. During central blocks, participants directed their attention to the central dot, whereas in peripheral blocks, their attention was focused on the dot indicated by the red arm. **(c)** illustrates the analysis pipeline utilized in this study. Individual beta maps ($n = 235$, 4 per participant) obtained from the first-level analysis were first split into training and test sets ($2/3$ as the training set). We employed a whole-brain multivariate machine-learning analysis using SVC to identify the distinct neural signatures of FA and SA. Model performance was evaluated using the hold-out test dataset, encompassing both intra-task and inter-task scenarios.

Fig. 2

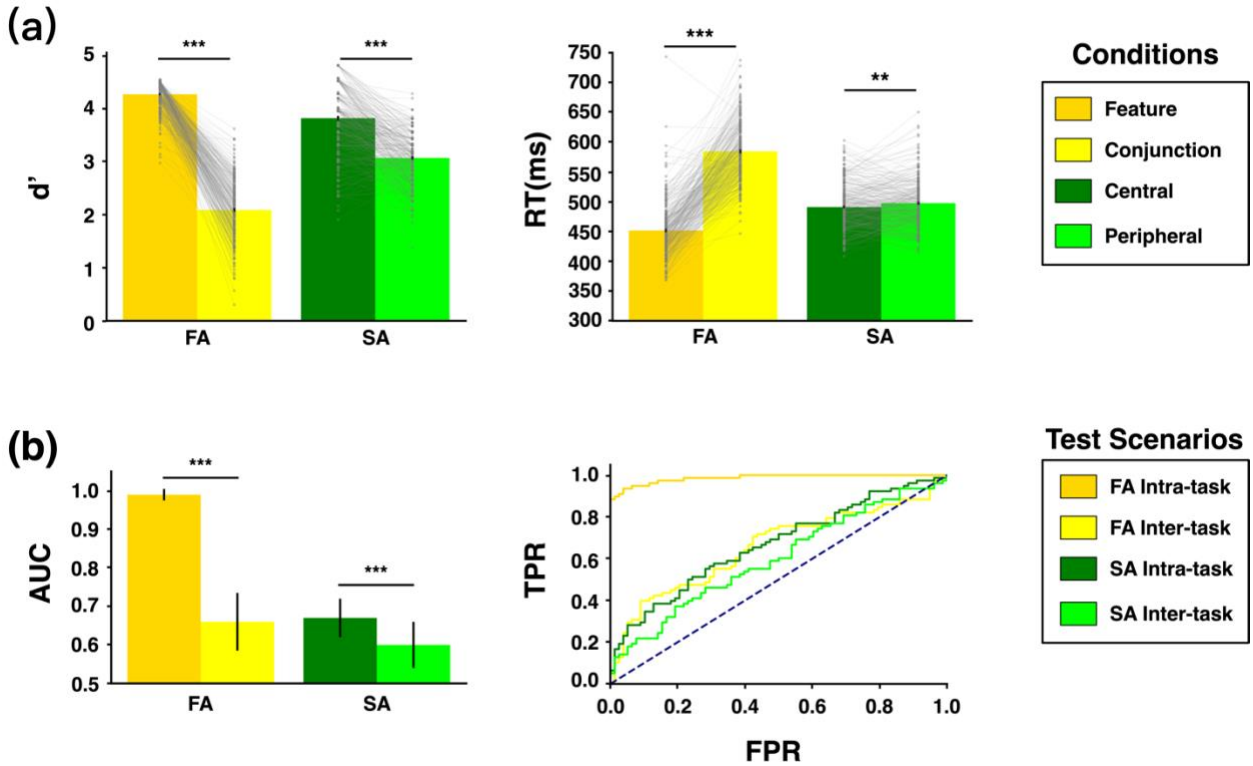

Fig. 2 Behavioral Results and Model Performance. (a) presents the behavioral results of the FA and SA experiments. The d' index quantifies participants' discriminability of signals from distractors, while Reaction Time (RT) measures the time elapsed from signal appearance to the subject's key press response, serving as metrics of task difficulty. The error bar represents the standard error (SE) of measurement. **(b)** Model performance of the neural signatures of FA and SA was assessed using the AUC metric, with error bars representing the 95% confidence interval (CI) obtained through 10,000 bootstrap iterations. The receiver operating characteristic curve (ROC) is depicted on the right, showing true positive rate (TPR) and false positive rate (FPR). Asterisks indicate the p-value obtained from matched-sample two-tailed t-test: ***, $p < 0.001$; **, $p < 0.01$.

Fig. 3

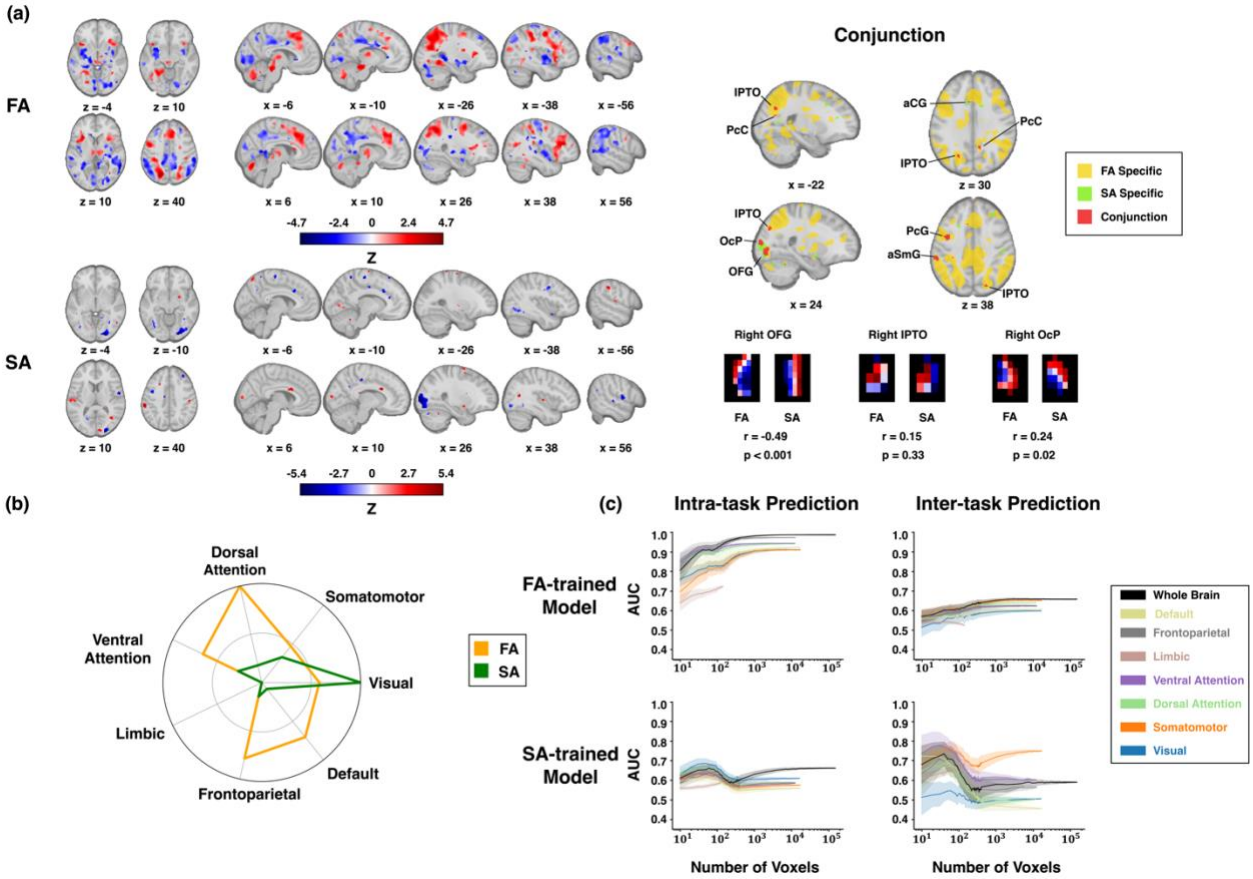

Fig. 3 Neural Signatures of FA and SA. (a) depicts the brain signatures of FA and SA with predictive voxels determined by bootstrap over 1,000 times and corrected by FDR at $q < 0.05$. Conjunction plots illustrate the voxels falling into three categories: FA Specific, SA Specific, and the overlapping regions between FA and SA. IPTO: intraparietal/transverse occipital sulci, PcC: Precuneus Cortex, aCG: anterior Cingulate Gyrus, OcP: Occipital Pole, OFG: Occipital Fusiform Gyrus, aSmG: anterior Supramarginal Gyrus. **(b)** The intensity of predictive voxels in FAS and SAS is presented across 7 cortical resting-state networks. **(c)** The predictivity of each network across the two tasks is depicted, with an equal number of voxels in each step.

Fig. 4

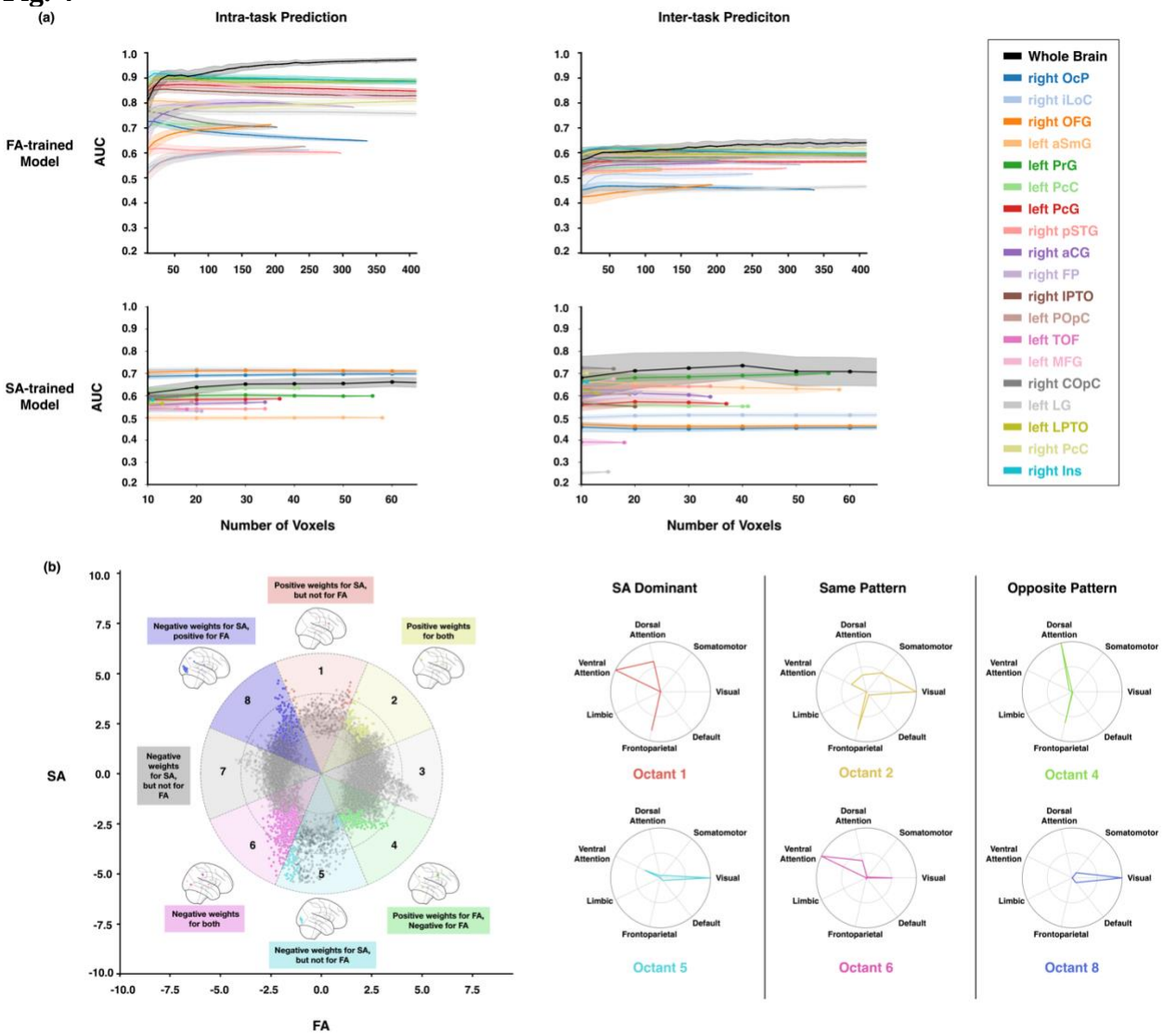

Fig. 4 Single-cluster Analysis and Voxel-level Analysis. (a) illustrates the outcome of stepwise single-cluster analysis, where each cluster was tested on both intra-task and inter-task datasets in steps of 10 voxels. For each step, we randomly sampled voxels within the cluster 1,000 times to retrain classifiers and measured model performance using the AUC score. Shaded regions indicate the standard deviation. **(b)** depicts the whole-brain neural signatures of FA and SA on a 2-D plane with normalized absolute voxel values within each signature. The plane is divided into octants, categorizing voxel patterns into FA dominant (Octants 3 and 7), SA dominant (Octants 1 and 5), Same Pattern (Octants 2 and 6), and Opposite Pattern (Octants 4 and 8). Light gray dots represent voxels uniquely predictive for FA, while dark gray dots represent those uniquely predictive for SA. Colored dots indicate voxels predictive for both FA and SA. Glass brains display the brain regions of conjunction voxels. Radar plots show the normalized cortical coverage of significant voxels across cortical networks in each octant.