The distinct role of human PIT in attention control

Siyuan Huang[1,2], Lan Wang[1], Sheng He[1,2].

1State Key Laboratory of Brain and Cognitive Science, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China

2University of Chinese Academy of Sciences, Beijing 100049, China

* Sheng He.

Email: hes@ibp.ac.cn

Author Contributions: S. Huang, L. Wang, and S. He designed research; S. Huang performed research; S. Huang analyzed data; and S. Huang, L. Wang, and S. He wrote the paper.

Competing Interest Statement: The authors declare no competing interest.

Classification: Psychological and Cognitive Sciences

Keywords: attention; cognition; fMRI; PITd; priority map

This Microsoft Word file includes:

Main Text
Figures 1 to 7

Abstract

Attentional distribution depends on both endogenous and exogenous processes, but how they interact in attention allocation remains unclear. The attentional priority map, jointly determined by stimulus saliency and task relevance, provides a framework for investigating their interplay. We propose that the human posterior inferotemporal cortex (hPIT), located near object-processing cortical areas, serves as an attentional priority map. Using fMRI with behavioral tasks, we show that hPIT shows stronger attentional modulation than classical attention regions across motion, color, and shape tasks. hPIT shows lateralized attentional enhancement even in the absence of visual input, with further elevated modulation in the presence of stimuli, indicating its integrated role in priority control. Furthermore, its modulation is invariant to stimulus category but sensitive to attentional demands, and the region is functionally connected to both dorsal and ventral attentional networks. These findings highlight the hPIT as an integrator in attentional control and provide critical insights into the brain's strategy for optimizing responses to the environment.

Main Text

Introduction

Attention serves as a pivotal mechanism within the brain, facilitating the allocation of its finite resource towards the most pertinent stimuli in our environment [1]. Attention can be systematically categorized according to its origin: bottom-up attention (exogenous), triggered by highly salient stimuli, and top-down attention (endogenous), initiated through voluntary selection [2–4]. The neural networks for bottom-up and top-down attention have been extensively investigated, with ventral and dorsal attention networks (VAN and DAN) specialized to support exogenous and endogenous attention respectively [5–7]. How these two networks flexibly interact to achieve integrated attentional selection, under situations that both salient stimulus and intentional goal are present, remains inadequately understood [8].

Critical to the integration of exogenous and endogenous attention is the concept of "priority map", defined as a map that reflects both the low-level salience and top-down influences [9–11]. An important question arises concerning the location of the priority map in the brain. For a brain region to qualify as the neural substrate supporting a priority map, it must satisfy the following criteria [12,13]: 1) having spatially restricted receptive fields, 2) exhibiting robust attentional modulation to predict attentional focus regardless of stimulus attributes but sensitive to attentional load, 3) displaying little feature specificity and high visual responsivity, and 4) exhibiting integration of bottom-up and top-down inputs. While the saliency map - where objects compete for greater representation and attentional allocation [14] across all feature dimensions - plays an important role in the mechanism of bottom-up attention, the priority map (jointly detemined by saliency and task guidance) is crucial for the cooperation of bottom-up and top-down attention, making it a key aspect to understand for the implementation of attention.

Previous research has indicated that the prefrontal cortex (PFC) and posterior parietal cortex (PPC) are involved in both top-down and bottom-up attention, rendering them potential candidates for supporting a priority map guiding attentional selection [3,15–18]. In addition, the superior colliculus (SC), especially its intermediate layers, receives signals from the frontal eye fields (FEF) and lateral intraparietal area (LIP) / intraparietal sulcus (IPS)[19–21], could also support a priority representation [22]. While visual areas, driven by stimulus features, were considered unlikely to exhibit priority representation [9], a recent study in monkeys revealed that the posterior inferotemporal cortex (PITd) could encode the locus of attention independent of features [12]. The structural connectivity between PITd and parieto-frontal attentional areas (specifically LIP and FEF) further supports the idea that PITd is part of attentional networks [23]. In the context of human brains, recent results have identified an area in the posterior inferotemporal cortex (hPIT) as a retinotopic and functional homologue of the macaque PITd, which was proposed to be a putative node for the human endogenous attentional control network [24]. However, since only endogenous attention was investigated in the previous experiments on monkey PITd as well as human PIT, it remains unclear whether hPIT is important for the integration of endogenous and exogenous attention.

The current study investigated if attention can modulate activation in hPIT independent of stimulus properties and cognitive demands. Instead of relying solely on the attentive motion task [24], hPIT was localized and validated using three distinct spatial attention tasks. Furthermore, we aim to discern the role of hPIT in both top-down and bottom-up attention, by manipulating the presence or absence of visual stimuli. Distinct from nodes of classical endogenous attention network, hPIT was more strongly modulated by attention in the presence of visual input. Furthermore, attention load, rather than object category, significantly impacts the modulation in hPIT. Additionally, hPIT showed strong functional connection with nodes in both ventral and dorsal attention network. Our results strongly suggest that hPIT, different from its adjacent object-processing areas and other parietal and frontal nodes of attention network, functions as an attention priority map.

Results

Localization and Validation of hPIT from Task-Invariant Activation to Spatial Attention

To identify brain area(s) that encode the location of attention invariant to stimulus type and cognitive demand [12], we employed three different spatial attention tasks in Exp. 1. In all three tasks, participants were instructed to fixate on a center dot and pay attention to circular aperture on one side, as directed by the cue. First task was about motion discrimination, and the second and third task required the discrimination of color proportion and shape proportion.

To isolate cortical areas modulated by endogenous spatial attention, contrasts between the conditions [attend contralateral – attend ipsilateral] were calculated for each of the three task blocks. Significant attention-modulated voxel maps were generated for individual participant (in most cases $p < 0.05$, cluster size > 20, but for activated voxel number > 10000, using $p < 0.01$). Then we took the intersection of these three maps and projected it to the cortex in every participant (example shown in Fig 1A), as our goal was to locate hPIT in inferotemporal cortex of human brain which should exhibit the properties of priority map, specifically in this context, insensitive to stimulus and cognition dimension. On the intersection of three maps, we found bilateral brain areas significantly modulated by spatial attention in intra-parietal sulcus (exclude the right hemisphere of S13) and inferotemporal cortex in all participants, and in most participants also the middle temporal cortex (exclude S07). Only three participants (S03, S04, S11) showed intersection area in the prefrontal cortex.

As illustrated in Fig 1B, the hPIT was manually identified bilaterally on the intersection maps within the posterior and ventral part of mid-fusiform sulcus as a contiguous cluster of voxels that are non-overlapping with fusiform face area (FFA) nor parahippocampal place area (PPA). In all participants, the hPIT is ventromedial to the occipital face area (OFA) and posterior to the FFA. The locations of hPIT, OFA and FFA identified in our study are shown in Fig 1C on the surface of standard brain

MNI152_2009c. The MNI coordinates delineating the center of mass (COM) for the ROIs are specified as follows: hPIT (-34, -72, -14), (33, -73, -13); FFA (-43, -55, -19), (41, -54, -17); OFA (-46, -79, -3), (46, -74, -8). The average cluster size of the hPIT is 210 voxels in left hemisphere and 191 voxels in right hemisphere, with 2x2x2 mm^3 voxels. Detailed location and cluster size of hPIT in every participant's cortical surface are reported in Supplementary Fig. 1.

Enhanced activation to attended than unattended moving dots in experiment 1, constrained by cerebral atlas, allowed us to also locate cortical areas that are critical nodes of attention network [5,6] bilaterally, including medial temporal visual area (MT), IPS (classified into IPS_P, the posterior part, and IPS_A, the anterior part), FEF, temporal parietal junction (TPJ), and ventral frontal cortex (VFC) (see Methods). Additionally, V1 was included as a ROI, so that we can see the consequences of attentional modulation in the primary visual cortex. Furthermore, the posterior part of lateral–occipital cortex (LOp) and FFA) were localized using cortical parcellation and functional contrast as controls, since they were adjacent to hPIT physically.

Attentional modulation of hPIT With and Without Bottom-up input

To investigate the modulation in hPIT and other brain areas applied by top-down attention (without stimulus) and top-down combined with bottom-up attention (with stimulus), we manipulated the presence or absence of visual stimuli in Exp. 2. Using event-related design, participants were required to direct their attention to the left or right visual field based on central cues. A dot was then presented on the attended side on half of the trials (50% probability), requiring participants to report, by pressing keys, its location relative to two reference points that were constantly presented (Fig. 2).

ROI-based analysis was performed to BOLD activation signals from hPIT and other ROIs identified in Exp. 1. Fig. 3A shows the averaged beta of contrast [attended - unattended (attend contralateral - attend ipsilateral)] from bilateral ROIs for each participant, illustrating activation patterns of our ROIs under both blank and dot conditions. A Two-Way ANOVA (n=15), considering blank/dot and ROI as factors, revealed significant main effects for both: ROI [$F(2.246, 31.44) = 19.20$, $P<0.0001$] and condition blank/dot [$F(1.000, 14.00) = 20.41$, $P=0.0005$]. A significant interaction between these two factors was also observed [$F(4.061, 56.86) = 10.98$, $P<0.0001$].

The condition wherein participants were required to attend to one side in the absence of any stimulus except the constant background, served to elucidate the impact of top-down attentional modulation. As illustrated by the blue bars, top-down modulation was apparent in multiple ROIs (V1, hPIT, MT, IPS_P, IPS_A, FEF, VFC), but it was strongest in hPIT. To establish quantitative metrics, a post-hoc multiple comparison test (Dunnett, one-sided) was utilized to compare the strength of attentional modulation between hPIT and other ROIs. The attentional modulation in hPIT was significantly stronger than all the other ROIs : V1 [$q(14) = 2.648$, adjusted $P= 0.0451$], MT [$q(14) = 2.856$, adjusted $P= 0.031$], IPS_P [q

(14) =2.680, adjusted P= 0.0426], IPS_A [q (14) =3.883, adjusted P= 0.005], FEF [q (14) =3.823, adjusted P= 0.005], TPJ [q (14) =4.261, adjusted P= 0.002], VFC [q (14) =3.161, adjusted P= 0.018].

With the introduction of a small dot as a target stimulus, we were able to examine the combined influence of top-down and bottom-up attention. As indicated by the pink bars, hPIT again exhibited the strongest attentional modulation effect, however, here the modulation effect reflects the combined effect of top-down and bottom-up attention. Post-hoc comparison (Dunnett, one-sided) revealed again that hPIT had significantly higher combined attentional modulation effect than all the other ROIs: V1[q (14) =5.717, adjusted P< 0.001], MT [q (14) =5.443, adjusted P< 0.001], IPS_P [q (14) =6.742, adjusted P< 0.001], IPS_A [q (14) =7.804, adjusted P< 0.001], FEF [q (14) =7.939, adjusted P< 0.001], TPJ [q (14) =7.944, adjusted P< 0.001], and VFC [q (14) =7.554, adjusted P< 0.001]. Further, among all the ROIs, hPIT and MT showed significantly stronger elevation in attentional modulation in the dot than the blank condition [Bonferroni, hPIT: t (14) =5.321, adjusted P< 0.001; MT: t (14) =4.326, adjusted P =0.003]. Comparatively, the combined attentional effect was stronger in hPIT than in MT [paired t test, one-tailed, t (14) =1.969, p=0.035].

As a supplement, comparison of attentional modulation across ROIs adjacent to hPIT was conducted (Fig.3B). Results of Two-way ANOVA showed significant main effect of condition blank/dot [df=1, F (1.000, 14.00) = 30.23, P<0.0001] and ROIs [df=2, F (1.921, 26.90) = 9.788, P=0.0007], as well as the interaction [df=2, F (1.311, 18.36) = 5.691, P=0.0210]. Multiple comparisons (Tukey's) demonstrated that the modulation in hPIT was significantly stronger than LOp and FFA in both conditions [Blank: q (14) =3.822, P=0.043 for LOp, q (14) =4.581, P=0.015 for FFA; Dot: q (14) =7.054, P=0.001 for LOp, q (14) =4.137, P=0.028 for FFA]. These findings highlight hPIT's distinctive pattern of attentional engagement compared to the other two areas.

Generally, while most ROIs showed significant modulation by attention in both the blank and dot conditions (see supplementary Table 1), hPIT is unique in that it showed the strongest attentional modulation in each condition as well as showing the largest elevation effect from blank to dot condition, suggesting that hPIT is deeply engaged in both bottom-up attention and top-down attention.

Image Category-invariant response but load-sensitive attentional effect in hPIT

To effectively guide spatial attention, hPIT should be broadly responsive to different stimulus features (i.e., lack of feature selectivity) but show high sensitivity to locations with salient and task-relevant features. In addition, such spatial sensitivity should be modulated by attentional load. In order to explore stimulus feature and attentional load sensitivity of hPIT, in experiment 3, images of three different categories (face, scene, scramble) were presented, participants were required to pay attention to the image on the left or right side according to the cue and report its moving direction. To modulate the attentional load, in each run, participants were instructed to make more demanding judgements (higher load) about the content of one pre-specified category of the attended images. For the pre-specified

category, the additional responses were female or male for faces, indoor or outdoor for scenes, overlay grating tilted clockwise counterclockwise for the scrambled images (Fig. 4).

Fig.5A shows the response of hPIT to different categories of images in different conditions, indicated by the beta values. To test if hPIT showed any effect of category preference, spatial attention and attention load sensitivity, we first performed a three-way ANOVA (n=15, spatial attention x stimulus category x task load). Results showed that the main effect of attention was significant [F (0.7628, 10.68) = 119.8, P<0.0001], the interaction between attention and category was significant [F (2.000, 28.00) = 4.068, P=0.0281], and the interaction between attention and task load was significant [F (0.6964, 9.749) = 8.307, P=0.0230]. Other main effects and interactions did not reach significance.

Since there is attention x category interaction, we further explored stimulus category sensitivity separately in the attend and un-attend conditions. Bayesian repeated measures ANOVA were adopted to compare the responses to face, scene and scramble stimuli under attend and un-attend conditions: attend [BF_{10}=0.93], unattend [BF_{10}=0.202]. The small Bayesian Factors (<1) under both conditions provide support that hPIT response was insensitive to these three categories, consistent with the criteria for priority map.

In addition, as the main effect of attention is significant, to gain a better understanding of the attention effect related to stimulus categories, we calculated the attentional modulation of hPIT for different image categories and task load (Fig.5B). A Two-Way ANOVA (n=15, stimulus category x task load) was performed. Significant main effects of category [F (1.852, 25.93) = 4.069, P= 0.0317] and task load [F (1.000, 14.00) = 8.307, P=0.0121] were found while their interaction was non-significant [F (1.527, 21.37) = 2.714, P=0.1004].

In general, unlike its adjacent object-processing areas, the response of hPIT is not sensitive to stimulus category. There is evidence that attentional modulation in hPIT is slightly stronger to faces than scenes and scrambled images, possibly reflecting the intrinsic saliency difference between these image categories. The observation that the attentional modulation in hPIT is sensitive to task load further supports hPIT's role as an attentional priority map.

Functional connectivity of hPIT to attentional networks

To unravel the functional connectivity of hPIT with the whole brain (especially the nodes of attention network), resting-state data were analyzed, with ROIs defined by task-based data (experiment 1 and 3, Fig. 6A) and cortical parcellation atlas [25]. Firstly, to visualize the patterns of functional connectivity, we calculated the average time course of hPIT, FFA and LOp using small sphere-shaped ROIs (2 mm radius) to avoid signal contamination for better visualization, and obtained the correlation coefficient between the seeds and all other voxels of brain to generate functional connectivity maps from the three spatially restricted seeds (Fig.6B). It is apparent that hPIT had strong functional connections with nodes of dorsal (FEF, IPS, MT) and ventral (VFC, TPJ) attention networks. Then the functional correlation using whole

ROIs (hPIT, FFA, and LOp, example shown in Fig.6C) with bilateral attention network nodes were calculated on each participant. The connection strength of bilateral hPIT, FFA, and LOp with DAN and VAN were compared using Two-Way ANOVA [n=15, attention networks x seed ROIs], results showed significant main effects of seed ROIs [F (1.732, 24.25) = 8.653, P=0.0021] and attentional networks [F (1.000, 14.00) = 60.56, P<0.0001] (Fig.6D). Specifically, the correlation coefficients of hPIT with attention networks were stronger than the other two seed areas (Dunnett, one-sided): PIT vs. FFA [q (14) = 2.135, adjusted P= 0.0452], PIT vs. LOp [q (14) = 4.725, adjusted P=0.0003]. The stronger connection of hPIT with each attentional nodes could be visualized in a circular connectivity plot (Fig.6E). The strong functional connections of hPIT with attention networks, especially the dorsal attention network, supports the important role it plays in attention control.

Discussion

It is traditionally accepted that the networks of frontal and parietal cortex play important roles in the control and engagement of endogenous and exogenous attention, while regions of occipito-temporal cortex specialize in the processing of colors and shapes, leading to the representation of scenes and objects, such as faces, bodies, words [12,26–33]. Our results provide new distinct insights about the role of the human posterior inferotemporal cortex in the control and implementation of attention. We identified a specific area within the posterior inferotemporal cortex, hPIT, where its activation exhibited little category selectivity, but was strongly modulated by attention across tasks, and reflected a combined effect of top-down attention and bottom-up attention. Further, the attentional modulation in hPIT was significantly affected by attention load as well as image category, consistent with the nature of attention. In addition, functional connectivity analysis revealed that hPIT, compared to its neighbor areas FFA and Lop, exhibited stronger connectivity with nodes of the attention network.

The hPIT is in close proximity to other object-processing regions in the inferotemporal cortex, including the FFA, the visual word form area (VWFA), and lateral occipital complex (LOC). The hPIT ((-34, -72, -14), (33, -73, -13)) is posterior to the FFA (40, -55, -10) [34] and VWFA (−43, −56, −16) [35,36]. LOC is localized by more activation when viewing objects than scramble, consisting of two subregions: LO (in our paper, LOp) and Loa/pFs [37]. The hPIT is inferior to the LOp [bound by (-41, -77, 3) and (-36, -71, -13)], posterior to the Loa/pFs (−38, −50, −17) [38].

Why does a brain region in the inferotemporal cortex exhibit the properties of an attention priority map, integrating the top-down and bottom-up attention? Previous studies have suggested that LIP/IPS, FEF, and SC are closely linked to attention guidance and eye movement and controls: LIP as a representation of attentional priority that remaps across saccades, FEF as an eye movement controller receiving LIP responses, and SC as reflecting the final saccade [9,16,39–45]. On the other hand, both top-down and bottom-

up attention are frequently associated with objects, ideally an area serving as attentional priority map should be positioned where it is easy to assess object information and also has broad connections with other regions of the attentional networks. The hPIT's location in the inferotemporal cortex facilitates the integration of key information from adjacent areas specialized in object-processing. As shown in our functional connectivity results as well as in previous studies [23,24,46,47], hPIT is connected with key areas of attention and eye movement controls, such as LIP/IPS, FEF, and SC. The placement of an area serving attentional priority map in the inferotemporal cortex, i.e., hPIT, has strategic advantages for both bottom-up information transmission and top-down attentional modulation.

In addition to the strong connection between hPIT and the dorsal attention network, consistent with a recent study [48], our functional connectivity analysis also revealed positive connections with TPJ and VFC, as shown in individual connectivity maps, though these connections were weaker compared to those with dorsal attention nodes. The connectivity with both the dorsal and ventral attention networks further supports hPIT's unique role in bridging the bottom-up and top-down attention.

Our study is limited in that the effects of feature-based attention were not specifically examined in our experiments. In addition, our current study lacks the temporal dynamic information of processing in hPIT and its downstream and upstream brain areas due to the use of fMRI measurements. Future studies would benefit from utilizing imaging methods with higher time resolution, with designs that address both spatial and feature-based attention.

In summary, the hPIT showed strong attentional modulation across stimuli and tasks, with sensitivity to attentional load, and more robust attentional modulation in the presence compared to the absence of visual input. The hPIT also demonstrated function connectivity to both dorsal and ventral attention networks. Together, our findings demonstrate the distinct role of hPIT in attention control, namely as an attentional priority map that integrates endogenous and exogenous attentional processes.

Materials and Methods

1. Participants

Fifteen health volunteers (8 males and 7 females, age ranged from 22 to 28 years old) with normal or corrected to normal vision participated in this study. None of the participants reported history of neurological or psychiatric symptoms. Written informed consent was obtained from all participants, and the study protocol was approved by the Institutional Review Board of the Institute of Biophysics, Chinese Academy of Sciences. Each participant completed three separate experimental sessions, conducted on three different days.

2 Stimuli and procedures

Visual stimuli for the fMRI experiment were programmed using Matlab (MathWorks, Natick, MA, USA) with Psychtoolbox (http://psychtoolbox.org/) and displayed via an MRI-compatible projector (1024 × 768@60 Hz) onto a screen positioned at the rear of the MRI scanner. Participants viewed the stimuli through a mirror attached to the head coil.

Because of the complexity of the task, each participant was trained on 2 or 3 separate days, before the fMRI sessions. This training was designed to familiarize participants with the tasks to relieve some possible tension during scanning and minimize the effects of learning.

2.1 Stimuli and procedures for experiment 1

In Experiment 1 (Day 1), participants completed three distinct spatial attention tasks across three separate experimental blocks (Experiments 1a, 1b, 1c). Throughout each task, participants were instructed to maintain fixation on a central point.

Experiment 1a (Fig.7A): Participants focused on a unilateral moving dot display and was required to discriminate the direction of the coherent motion. As shown on Fig.7A, each trial began with an initial cue indicating the target side, represented by a bar (0.6° visual angle) attached to the fixation point. Random dot stimuli were then presented within a circular aperture (radius: 3° visual angle) centered 9° from the fixation point. The dot display consisted of dots (size: 0.15° visual angle; density: 5 dots per degree of visual angle) in motion for 2.5 seconds. During this sequence, a 0.5-second interval of coherent motion (coherence: 50%, velocity: 6°/s) occurred randomly between 0.5 and 1.55 seconds after motion onset, while the remaining duration involved random dot movement. Following the motion sequence, the circular aperture disappeared, and an arrow appeared at the center of the screen for 1.5 seconds. Participants were required to press a key to indicate whether the direction of coherent motion matched the direction indicated by the arrow. Each block consisted of three trials with the same attended side, followed by a 6-second rest period.

Experiment 1b (Fig.7B): Participants were instructed to attend to the proportion of red and green dots presented within a unilateral circular aperture. The red and green dots were adjusted individually by each participant to achieve iso-luminance, ensuring perceptual equality between the colors. The positioning of the circular aperture were consistent with Experiment 1a, though the dot size was increased to 0.3° of visual angle in this task. At the start of each trial, within the first second, dots were displayed in random proportions of red (20%, 40%, 60%, 80%) with the remainder being green (pattern 1). Following this, the position of the dots remained static during the 2nd second, but their color could potentially change (pattern2). The proportion of red dots remain to be one of the pre-set ratios. In the final, third second of each trial, the bilateral dots disappeared. Participants were then required to press a key to indicate

whether the color ratio of the dots had changed from pattern 1 to pattern 2. Each block in this experiment consisted of four trials, with participants attending to the same side throughout, followed by a 6-second rest period.

Experiment 1c (Fig.7C): In this experiment, participants were instructed to attend to a unilateral geometrical pattern display, focusing on the proportion of different shapes. Each trial began with the first pattern presented for 2 seconds, followed by a gradual transition in which the second pattern emerged over the next 2 seconds and then remained visible. At the end of this sequence, participants were required to press a key indicating whether the shape proportion between the two patterns had changed. Each block consisted of two trials with attention directed to the same side, followed by a 6-second rest period.

There were 16 blocks in one run, 8 attending left and 8 attending right for experiment 1a, 1b and 1c. On the first day of scanning, participants completed 6 runs of Experiment 1a, 4 runs of Experiment 1b, and 4 runs of Experiment 1c.

2.2 Stimuli and procedures for experiment 2

In experiment 2 (day 2), participants first underwent a 488-second resting-state functional scan, during which they passively viewed a grey screen. For the subsequent attention tasks (event-related design), participants were instructed to maintain fixation on a central dot while directing their attention to one side.

The procedure (Fig.2) began with the central dot (10 pixels in diameter) flashing three times over 2 seconds, serving as an alert for the upcoming cue and target. Following this, a cue was presented for 0.8s. After a variable delay, randomly selected from intervals of 0.1, 0.15, 0.2, or 0.25 seconds, a target dot (8 pixels in diameter, RGB: [0.7,0.7,0.7]) either appeared on the cued side (dot condition) or did not appear (blank condition). In the dot condition, participants were required to report the location of the target by pressing a key. In the blank condition, participants were instructed to refrain from any response.

Each run consisted of 16 trials, with individual trial durations of 14, 16, or 18 seconds. On this second day of scanning, participants completed one run of the resting-state scan, followed by 8 task runs.

2.3 Stimuli and procedures for experiment 3

In experiment 3 (day 3), participants were asked to attend unilateral images presented within a circular aperture, consistent with the method used in Experiment 1. The images were drawn from three distinct categories:

Faces: Asian faces with an equal gender distribution (50% female).

Scenes: Equally divided between indoor and outdoor settings (50% outdoor).

Scrambles: Phase-scrambled images superimposed with a grating, with 50% of the gratings spanning the 1st and 3rd quadrants.

To ensure visual consistency across images, they were standardized for luminance histograms and spatial frequency using the SHINE toolbox (as referenced from Willenbockel et al., 2010).

At the onset of each run, participants were instructed to pay extra attention to a designated category throughout that run. The procedure (Fig.4) began with the center dot flashing three times over two seconds, signaling the forthcoming presentation of the cue and target. Following this, both the cue and bilateral images from the same category were presented simultaneously. This images briefly shifted in one direction before returning to their original position (0.25s+0.25s). During the next 1.5 seconds, participants first pressed a key to indicate whether the direction of the images' motion matched that of the arrow at the center. A green dot then appeared at the center, reminding participants to categorize the image in their attended location. If the image matched the category to which they had been instructed to pay attention, participants were required to identify specific content about the image (e.g., determining whether a face was female or male) and responded by pressing either key '1' or '2'. If the image did not belong to the designated category, participants simply pressed '3'.

Each run consisted of 32 trials: 50% of the trials featured images from the emphasized category, and the remaining 50% were evenly distributed between the other two categories (25% each). Each trial lasted either 8, 10, or 12 seconds. Across Experiment 3, there were a total of nine runs, with each category being the focus of three runs. The accuracy of extra judgements is 82.9% for face, 81.8% for scene and 84.3% for scramble [$F(1.373, 19.23) = 1.096$, $P=0.3308$].

3 MRI Data Acquisition

MRI scanning was conducted using a 3T Siemens Prisma scanner at the Beijing MRI Center for Brain Research (BMCBR), utilizing a standard 20-channel head coil. High-resolution T1-weighted anatomical images were acquiredat the start of each session (TR = 3000 ms; TE = 3.02 ms; 176 slices; slice thickness = 1 mm; no inter-slice gap; field of view = 256 mm; flip angle = 8°; image matrix: 256×256).

Functional data were collected with gradient-echo EPI sequences (TR = 2000 ms; TE = 30.0 ms; 52 slices; slice thickness = 2 mm; no inter-slice gap; voxel resolution 2.0 x 2.0 x 2.0 mm, field of view = 192 mm; flip angle = 80°; image matrix: 96×96).

4 MRI Data Analysis

fMRI data were preprocessed and analyzed using FreeSurfer and AFNI and custom Python code. The preprocessing steps included de-spiking, slice timing correction, EPI distortion correction (PE blip-up), rigid body motion correction, spatial smoothing (4 mm FWHM Gaussian kernel, for the task runs), and per run scaling (as percent signal change).

For task runs, general linear models were used to estimate BOLD signal change from baseline for each stimulus condition. For each individual, bilateral FFA, OFA and PPA were defined based on the functional contrast between faces and scenes from experiment 3. Bilateral V1 were defined using function contrast [attend contra moving dot - baseline] ($P < 0.001$, uncorrected), and IPS, FEF and MT were defined using

the same contrast (P < 0.01, uncorrected) and referring the cerebral atlas (Glasser, 2016). Due to the lack of standardized anatomical definitions, bilateral VFC and TPJ were identified using the same functional contrast (P < 0.01, uncorrected) [5]. Bilateral hPIT was localized using data from all three tasks in Experiment 1, by intersecting activation maps on the inferotemporal surface, while avoiding the locations of FFA and PPA.

The resting-state data was preprocessed similarly to the functional data, with additional steps for removing white matter and cerebrospinal fluid signals. Subsequently, confound regression, spatial smoothing (4 mm), and bandpass filtering (0.01–0.1 Hz) were performed using AFNI's 3dTproject.

Acknowledgments

This work was supported by STI2030-Major Projects (Grant Nos. 2021ZD0204200 and 2021ZD0203800); and Key Research Program of Frontier Sciences, Chinese Academy of Science (Grant No. KJZD-SW-L08).

References

1. Desimone, R. & Duncan, J. Neural Mechanisms of Selective Visual Attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
2. Connor, C. E., Egeth, H. E. & Yantis, S. Visual attention: bottom-up versus top-down. *Current Biology* **14**, R850–R852 (2004).
3. Katsuki, F. & Constantinidis, C. Bottom-Up and Top-Down Attention: Different Processes and Overlapping Neural Systems. *Neuroscientist* **20**, 509–521 (2014).
4. Buschman, T. J. & Miller, E. K. Top-Down Versus Bottom-Up Control of Attention in the Prefrontal and Posterior Parietal Cortices. *Science* **315**, 1860–1862 (2007).
5. Vossel, S., Geng, J. J. & Fink, G. R. Dorsal and Ventral Attention Systems: Distinct Neural Circuits but Collaborative Roles. *Neuroscientist* **20**, 150–159 (2014).
6. Corbetta, M. & Shulman, G. L. Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* **3**, 201–215 (2002).
7. Corbetta, M., Patel, G. & Shulman, G. L. The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron* **58**, 306–324 (2008).

8. Noudoost, B., Chang, M. H., Steinmetz, N. A. & Moore, T. Top-down control of visual attention. *Current Opinion in Neurobiology* **20**, 183–190 (2010).

9. Bisley, J. W. & Mirpour, K. The neural instantiation of a priority map. *Current Opinion in Psychology* **29**, 108–112 (2019).

10. Bisley, J. W. & Goldberg, M. E. Attention, Intention, and Priority in the Parietal Lobe. *Annu. Rev. Neurosci.* **33**, 1–21 (2010).

11. Treue, S. Visual attention: the where, what, how and why of saliency. *Current Opinion in Neurobiology* **13**, 428–432 (2003).

12. Stemmann, H. & Freiwald, W. A. Evidence for an attentional priority map in inferotemporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 23797–23805 (2019).

13. Ptak, R. The Frontoparietal Attention Network of the Human Brain: Action, Saliency, and a Priority Map of the Environment. *Neuroscientist* **18**, 502–515 (2012).

14. Fecteau, J. & Munoz, D. Salience, relevance, and firing: a priority map for target selection. *Trends in Cognitive Sciences* **10**, 382–390 (2006).

15. Katsuki, F. & Constantinidis, C. Early involvement of prefrontal cortex in visual bottom-up attention. *Nat Neurosci* **15**, 1160–1166 (2012).

16. Arcizet, F., Mirpour, K. & Bisley, J. W. A Pure Salience Response in Posterior Parietal Cortex. *Cerebral Cortex* **21**, 2498–2506 (2011).

17. Herrington, T. M. & Assad, J. A. Neural Activity in the Middle Temporal Area and Lateral Intraparietal Area during Endogenously Cued Shifts of Attention. *J. Neurosci.* **29**, 14160–14176 (2009).

18. Ninomiya, T., Sawamura, H., Inoue, K. & Takada, M. Segregated Pathways Carrying Frontally Derived Top-Down Signals to Visual Areas MT and V4 in Macaques. *J. Neurosci.* **32**, 6851–6858 (2012).

19. Schall, J. D., Purcell, B. A., Heitz, R. P., Logan, G. D. & Palmeri, T. J. Neural mechanisms of saccade target selection: gated accumulator model of the visual–motor cascade. *Eur J of Neuroscience* **33**, 1991–2002 (2011).

20. Benarroch, E. What Are the Functions of the Superior Colliculus and Its Involvement in Neurologic Disorders? *Neurology* **100**, 784–790 (2023).
21. Rushworth, M. F. S., Behrens, T. E. J. & Johansen-Berg, H. Connection Patterns Distinguish 3 Regions of Human Parietal Cortex. *Cerebral Cortex* **16**, 1418–1430 (2006).
22. White, B. J. *et al.* Superior colliculus neurons encode a visual saliency map during free viewing of natural dynamic video. *Nat Commun* **8**, 14263 (2017).
23. Sani, I., McPherson, B. C., Stemmann, H., Pestilli, F. & Freiwald, W. A. Functionally defined white matter of the macaque monkey brain reveals a dorso-ventral attention network. *eLife* **8**, e40520 (2019).
24. Sani, I. *et al.* The human endogenous attentional control network includes a ventro-temporal cortical node. *Nat Commun* **12**, 360 (2021).
25. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
26. Ungerleider, S. K. A. L. G. Mechanisms of Visual Attention in the Human Cortex. *Annu. Rev. Neurosci.* **23**, 315–341 (2000).
27. Moore, T., Armstrong, K. M. & Fallah, M. Visuomotor Origins of Covert Spatial Attention. *Neuron* **40**, 671–683 (2003).
28. Moore, T. & Zirnsak, M. Neural Mechanisms of Selective Visual Attention. *Annu. Rev. Psychol.* **68**, 47–72 (2017).
29. Baldauf, D. & Desimone, R. Neural Mechanisms of Object-Based Attention. *Science* **344**, 424–427 (2014).
30. Lafer-Sousa, R. & Conway, B. R. Parallel, multi-stage processing of colors, faces and shapes in macaque inferior temporal cortex. *Nat Neurosci* **16**, 1870–1878 (2013).
31. Peelen, M. V. & Downing, P. E. Selectivity for the Human Body in the Fusiform Gyrus. *Journal of Neurophysiology* **93**, 603–608 (2005).
32. McCandliss, B. D., Cohen, L. & Dehaene, S. The visual word form area: expertise for reading in the fusiform gyrus. *Trends in Cognitive Sciences* **7**, 293–299 (2003).

33. Grill-Spector, K. & Weiner, K. S. The functional architecture of the ventral temporal cortex and its role in categorization. *Nat Rev Neurosci* **15**, 536–548 (2014).

34. Kanwisher, N., McDermott, J. & Chun, M. M. The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *J. Neurosci.* **17**, 4302–4311 (1997).

35. Rauschecker, A. M., Bowen, R. F., Parvizi, J. & Wandell, B. A. Position sensitivity in the visual word form area. *Proc. Natl. Acad. Sci. U.S.A.* **109**, (2012).

36. Zhang, B., He, S. & Weng, X. Localization and Functional Characterization of an Occipital Visual Word form Sensitive Area. *Sci Rep* **8**, 6723 (2018).

37. Grill-Spector, K., Kourtzi, Z. & Kanwisher, N. The lateral occipital complex and its role in object recognition. *Vision Research* **41**, 1409–1422 (2001).

38. Grill-Spector, K. *et al.* Differential Processing of Objects under Various Viewing Conditions in the Human Lateral Occipital Complex. *Neuron* **24**, 187–203 (1999).

39. Thompson, K. G., Hanes, D. P., Bichot, N. P. & Schall, J. D. Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *Journal of Neurophysiology* **76**, 4040–4055 (1996).

40. McPeek, R. M. & Keller, E. L. Saccade Target Selection in the Superior Colliculus During a Visual Search Task. *Journal of Neurophysiology* **88**, 2019–2034 (2002).

41. Buschman, T. J. & Miller, E. K. Serial, Covert Shifts of Attention during Visual Search Are Reflected by the Frontal Eye Fields and Correlated with Population Oscillations. *Neuron* **63**, 386–396 (2009).

42. Zhou, H. & Desimone, R. Feature-Based Attention in the Frontal Eye Field and Area V4 during Visual Search. *Neuron* **70**, 1205–1217 (2011).

43. Shen, K. & Paré, M. Neuronal activity in superior colliculus signals both stimulus identity and saccade goals during visual conjunction search. *Journal of Vision* **7**, 15 (2007).

44. Foley, N. C., Kelly, S. P., Mhatre, H., Lopes, M. & Gottlieb, J. Parietal neurons encode expected gains in instrumental information. *Proc. Natl. Acad. Sci. U.S.A.* **114**, (2017).

45. Andersen, R. A., Brotchie, P. R. & Mazzoni, P. Evidence for the lateral intraparietal area as the parietal eye field. *Current Opinion in Neurobiology* **2**, 840–846 (1992).

46. Boshra, R. & Kastner, S. Attention control in the primate brain. *Current Opinion in Neurobiology* **76**, 102605 (2022).

47. Bogadhi, A. R., Bollimunta, A., Leopold, D. A. & Krauzlis, R. J. Spatial Attention Deficits Are Causally Linked to an Area in Macaque Temporal Cortex. *Current Biology* **29**, 726-736.e4 (2019).

48. Meng, Z., Huang, Y., Wang, W., Zhou, L. & Zhou, K. Orienting role of the putative human posterior infero-temporal area in visual attention. *Cortex* **175**, 54–65 (2024).

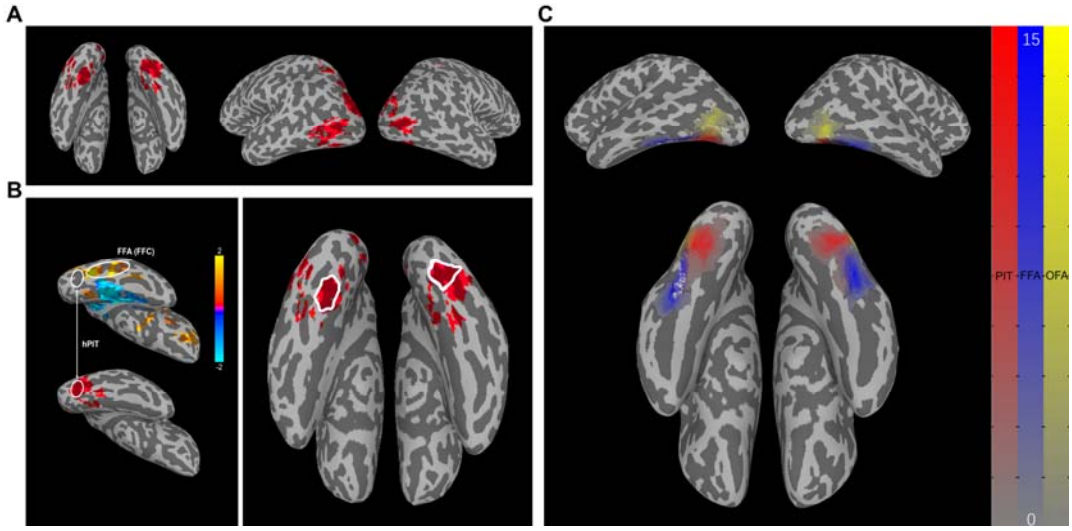

Figure 1. Functional localization of hPIT. (A): The intersection of three maps of three block tasks on one typical participant (S02). The cortical areas which are attached in red has shown significant activation in all three attention tasks. (B): The exact position of hPIT. Left: Positions of FFA and hPIT in statistical parametric maps of the contrast [attend face – attend scene] (p=0.05) (top) and hPIT in intersection map on the right hemisphere of S02. Right: The location of hPIT on both hemispheres of S02 circled by white line. The cortical areas which are attached in red has shown significant activation in all three attention tasks. (C): The positions of hPIT, OFA and FFA of 15 participants overlapped on surface of MNI152_2009c, with larger numerical values (manifested as deeper colors) indicating a higher degree of overlap among participants. The color scale ranging from grey to red delineates the spatial distribution of the hPIT, while the scale from grey to blue represents the FFA, and grey to yellow signifies the OFA.

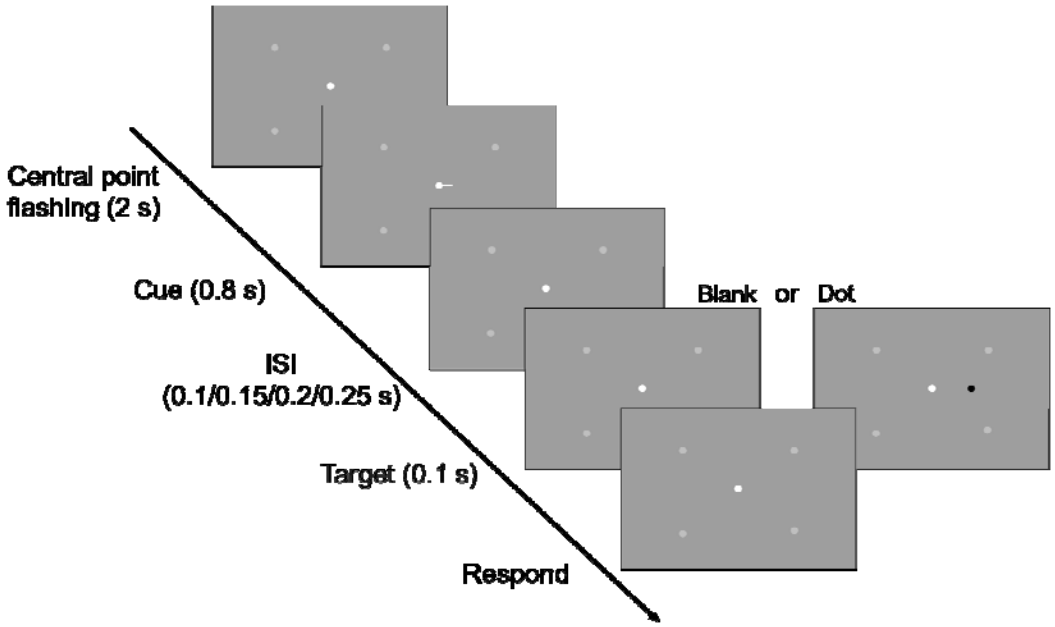

Figure 2. Schematic depiction of the experimental design for Experiment 2. Following the flashing of central point, a cue is presented. In the dot condition, a dot appears on the attended side, prompting participants to report its relative position. In the blank condition, participants are instructed not to press any keys. (Light grey dots: serve as indicators, showing participants the potential target locations during the experiment.).

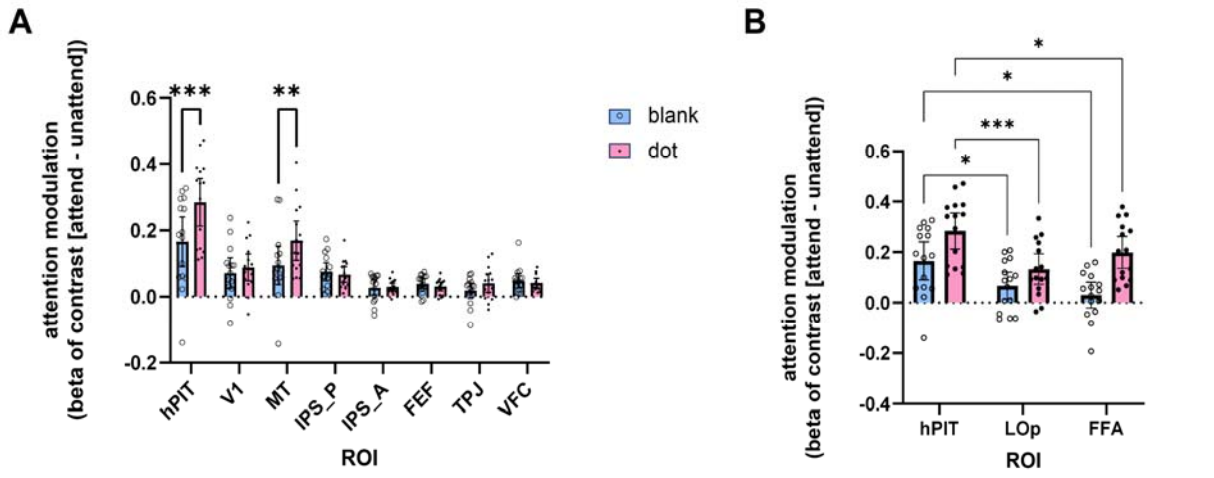

Figure 3. Attention modulation in different brain regions. (A): The modulation pattern of V1, hPIT, MT, IPS, FEF, TPJ and VFC in condition blank (blue bar) and condition dot (pink bar), using beta of contrast: [attended -unattended (attend contralateral - attend ipsilateral)]. The modulation difference of attention between condition blank and condition dot reached significant level in PITd and MT. (B): The modulation pattern of hPIT, LOp, and FFA in condition blank (blue bar) and condition dot (pink bar). Error bars indicate 95% confidence interval. ∗∗∗ indicates the paired t test with significance of $P < 0.001$. ∗∗ indicates the paired t test with significance of $P < 0.01$. ∗ indicates the paired t test with significance of $P < 0.05$.

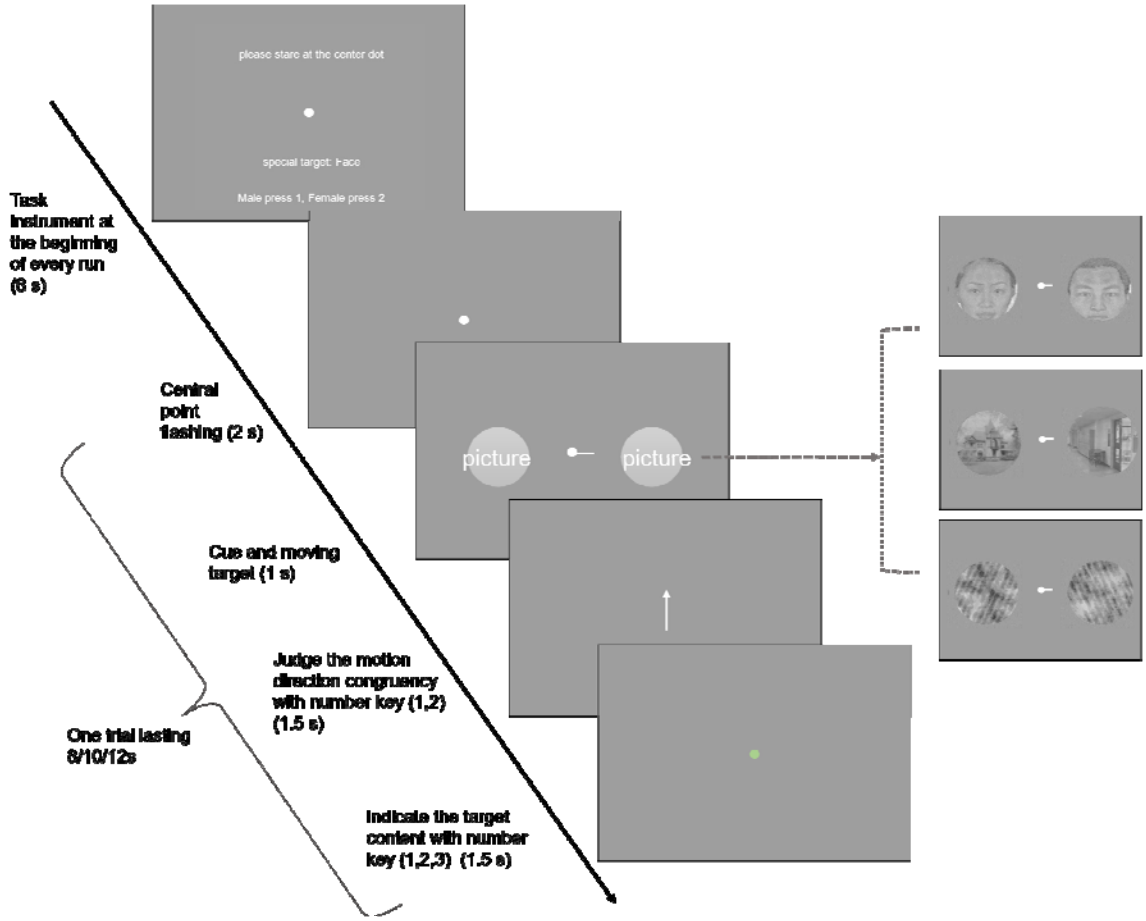

Figure 4. Schematic overview of the experimental paradigm for the experiment 3. At the start of the "face run", participants were instructed to pay heightened attention to the face images throughout that specific run. Following the central point's flashing sequence, a cue and bilateral images was presented. Participants first judged if the movement direction of the attended image aligned with the central arrow. Upon the central point changing to green, participants then responded based on the content of the attended image.

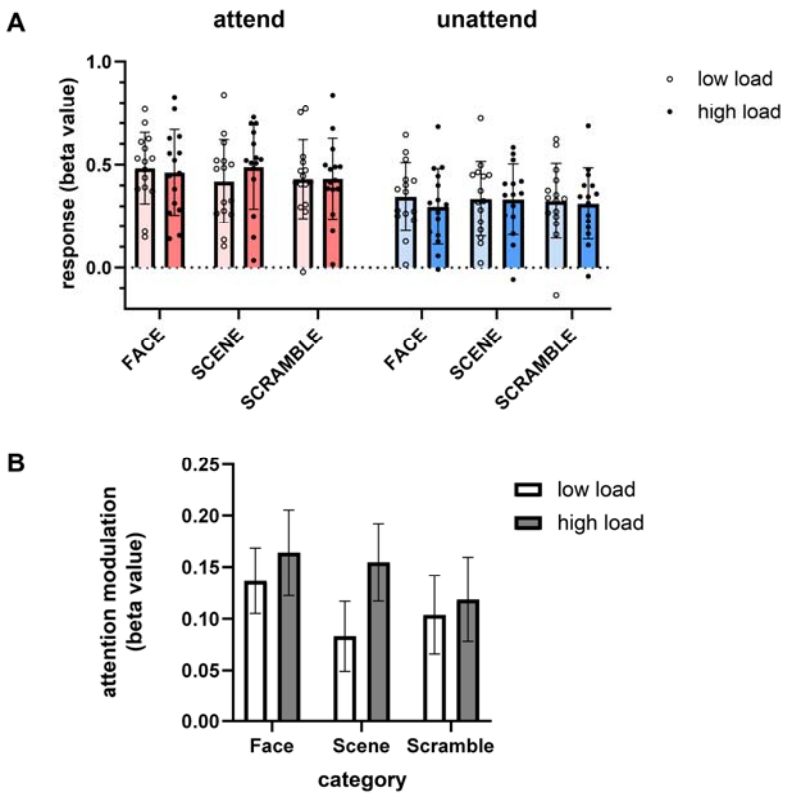

Figure 5. Response and attention modulation of hPIT to images of different categories. (A): The activation level of hPIT when attending to or not attending to images of different categories. Red bars indicate condition with attention in the receptive field (attended), while blue bars indicate condition with attention in the other side (unattended). Bars with higher chroma represent high load condition, lower chroma represent low load condition. (B): The modulation pattern of hPIT when presenting different categories of images with different attention load. White bars represent condition with low-load attention; Grey bars represent condition high-load attention. Error bars indicate 95% confidence interval.

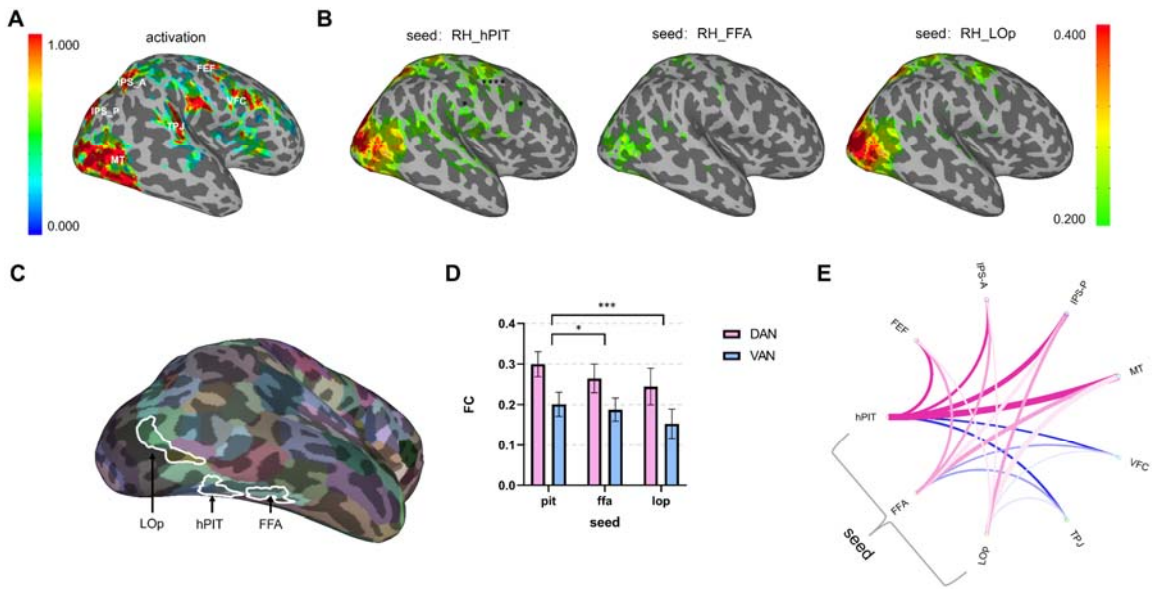

Figure 6. Functional connectivity analysis of hPIT and its neighboring areas. (A): Activation map showing beta of contrast [attend contra moving dot - baseline] ($p < 0.01$, uncorrected) and the location of critical nodes in dorsal and ventral attention network on one typical subject (S03). (B): Thresholded map showing functional connectivity of seed sphere right-hemi hPIT, right-hemi FFA, and right-hemi LOp (Spearman's rank correlation coefficient > 0.2, $p<0.05$, uncorrected), averaged across subjects and projected onto the surface of standard brain MNI152_2009c. Color bar attached indicate the intensity of activation (A) and correlation (B). (C): The relative location of LOp, hPIT and FFA on the inflated cortical surface of parcellation map (Glasser's atlas). (D): Strength of functional connectivity of seed hPIT, FFA and LOp with DAN and VAN. Error bars indicate 95% confidence interval. $***$ indicates the significance of $P < 0.001$. $*$ indicates the significance of $P < 0.05$. (E): Circular plot for functional connectivity of seed hPIT, FFA and LOp with nodes of attention network of right hemisphere, with pink lines indicating connection with nodes of DAN, blue lines indicating nodes of VAN. Connections to left hemisphere nodes show similar but weaker trends. Opacity of each line connecting seed and nodes represents the rank of its connectivity strength (the strongest 100%, the middle 44%, the weakest 11%). Width of each line is scaled to its cubic numerical intensities of connectivity.

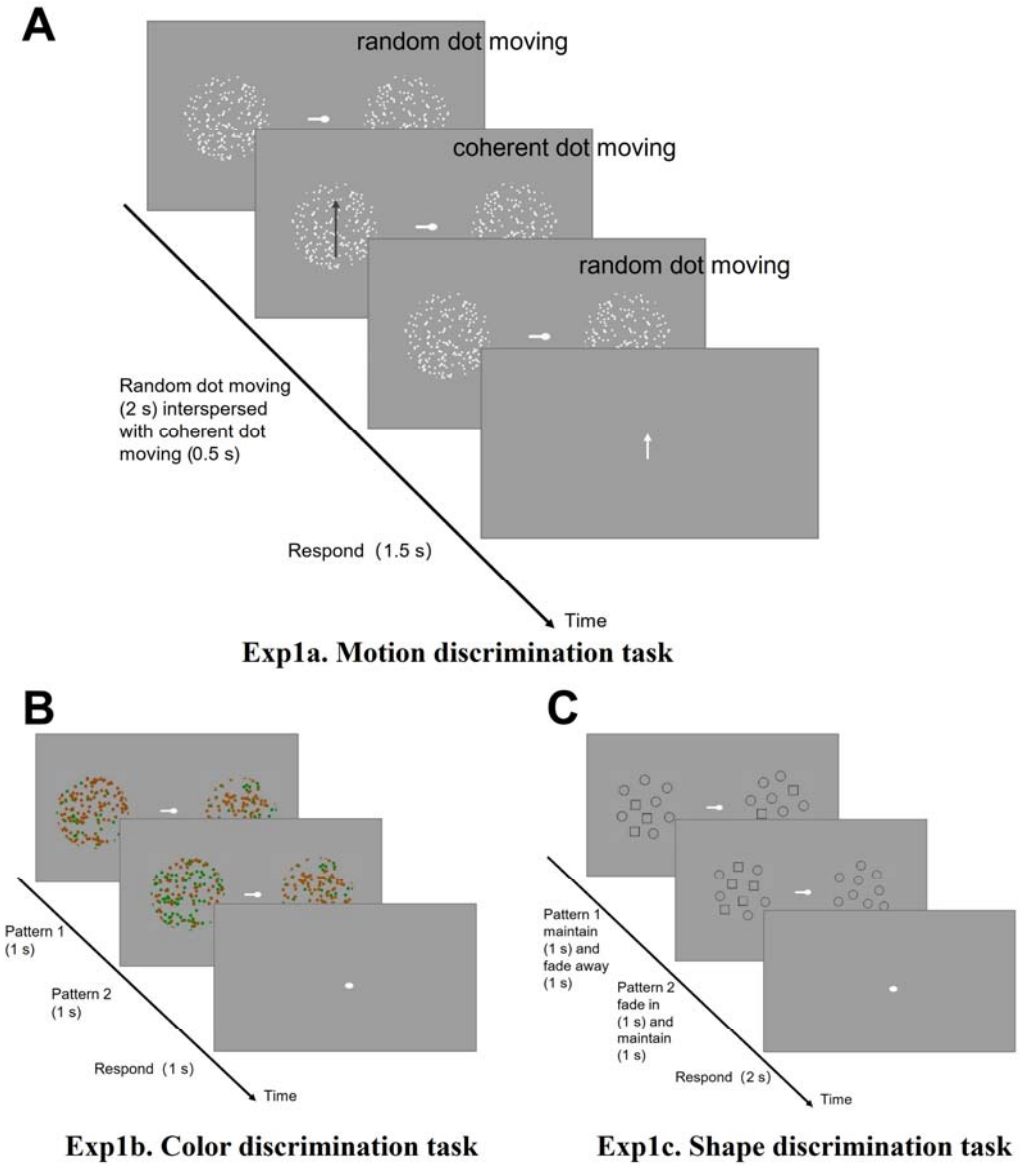

Figure 7. Schematic representation of the experimental paradigm for the Experiment 1. (A): The task was to report whether the direction of coherent motion on the attended side matched that of the white arrow. Note: The black arrow, representing one potential direction for the coherent dot movement, is used for illustrative purposes and was not actually presented during the experiment. (B): Pattern 1 and pattern 2, consisting of iso-luminant red and green dots, were presented sequentially. The task was to compare the color ratios of these patterns on the attended side and respond accordingly if any changes in the color ratio were detected. (C): Pattern 1 and pattern 2, consisting of equal number of small shapes (circles and squares) were presented sequentially. The task was to compare the shape ratios of these patterns on the attended side and respond accordingly if any changes in the shape ratio were detected.