

Reviewed Preprint  
v1 • January 12, 2026  
Not revised

✉ For correspondence:

xhe@bournemouth.ac.uk

yhangpsy@suda.edu.cn

† These authors contributed equally to the work as co-first authors

**Competing interests:** No competing interests declared

**Funding:** See page 17

**Reviewing editor:** Xilin Zhang,  
South China Normal University,  
China

© 2026, Chen et al. This article is distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use and redistribution provided that the original author and source are credited.

# Dissociable neural substrates of integration and segregation in exogenous attention

Yu-Jie Chen<sup>1,†</sup>, Ai-Su Li<sup>1,†</sup>, Yang Yu<sup>2</sup>, Su Hu<sup>2</sup>, Xun He<sup>3</sup>✉, Yang Zhang<sup>1</sup>✉

<sup>1</sup>Department of Psychology, Soochow University, Suzhou, China, 215000 • <sup>2</sup>Department of Radiology, The First Affiliated Hospital of Soochow University, Suzhou, China, 215000 • <sup>3</sup>School of Psychology, Bournemouth University, Poole, United Kingdom

## eLife Assessment

This **important** study provides the first direct neuroimaging evidence for the integration-segregation theory of exogenous attention underlying inhibition of return, using an optimized IOR-Stroop fMRI paradigm to dissociate integration and segregation processes and to demonstrate that attentional orienting modulates semantic- and response-level conflict processing. Although the empirical evidence is **compelling**, clearer justification of the experimental logic, more cautious framing of behavioral and regional interpretations, and greater transparency in reporting and presentation are needed to strengthen the conclusions. The work will be of broad interest to researchers investigating visual attention, perception, cognitive control, and conflict processing.

<https://doi.org/10.7554/eLife.109842.1.sa3>

## Abstract

The integration-segregation theory proposes that early facilitation and later inhibition (i.e., inhibition of return, IOR) in exogenous attention arise from the competition between cue-target event integration and segregation. Although widely supported behaviorally, the theory lacked direct neural evidence. Here, we used event-related fMRI with an optimized cue-target paradigm to test this account. Cued targets elicited stronger activation in the frontoparietal attention networks, including the bilateral frontal eye field (FEF) and intraparietal sulcus (IPS), right temporoparietal junction (TPJ), and left dorsal anterior cingulate cortex (dACC), consistent with the notion of attentional demand of reactivating the cue-initiated representations for integration. In contrast, uncued targets engaged the medial temporal cortex, particularly the bilateral parahippocampal gyrus (PHG) and superior temporal gyrus (STG), reflecting the segregation processes associated with new object-file creation and novelty encoding. These dissociable activations provide the first direct neuroimaging evidence for the integration-segregation theory. Moreover, we observed neural interactions between IOR and cognitive conflict, suggesting a potential modulation of conflict processing by attentional orienting. Taken together, these findings provide new insights into exogenous attention by clarifying the neural underpinnings of integration and segregation and uncovering the interaction between spatial orienting and conflict processing.

## Impact Statement

This study offers the first direct neuroimaging evidence for the integration-segregation theory of exogenous attention by identifying dissociable neural signatures of the competing integration and segregation processes, strengthening the understanding of attentional orienting mechanisms.

## Introduction

Attention can be voluntarily directed following one's behavioral goals (goal-directed or endogenous attention) or reflexively oriented by salient information (such as abrupt onsets) in the environment (stimulus-driven or exogenous attention). The exogenous attentional orienting is crucial for efficient visual search in cluttered scenes (Klein et al., 2023 [2](#); Li et al., 2023 [2](#); Ma et al., 2011 [2](#); Wang & Klein, 2010 [2](#); Wolfe & Horowitz, 2017 [2](#)). Studies employing the cue-target paradigm (Posner & Cohen, 1984 [2](#)) have shown that the exogenous attentional orienting shows a biphasic temporal pattern. An uninformative peripheral cue initially facilitates subsequent target processing at the cued location at short stimulus-onset asynchronies (SOAs), and later turns into inhibiting responses at the cued location at long SOAs (typically over 200 ms), a phenomenon known as inhibition of return (IOR) (Klein, 2000 [2](#); Lupiáñez et al., 2006 [2](#); Posner et al., 1985 [2](#); Seidel Malkinson et al., 2024 [2](#)). This characteristic shift from facilitation to inhibition at the cued location has fueled decades of theoretical debate, giving rise to multiple competing accounts of its underlying mechanisms (Funes et al., 2008 [2](#); Klein & Dick, 2002 [2](#); Lupiáñez, 2010 [2](#); Lupiáñez et al., 2001 [2](#); Milliken et al., 2000 [2](#); Prime & Jolicoeur, 2009 [2](#); Taylor & Klein, 1998 [2](#); Vivas et al., 2007 [2](#)).

Among these, one of the most influential and extensively developed theories explaining the biphasic effect is the integration-segregation theory proposed by Lupiáñez and colleagues (Funes et al., 2008 [2](#); Lupiáñez et al., 2001 [2](#); Milliken et al., 2000 [2](#)). Rooted in the object file framework (Kahneman et al., 1992 [2](#)), this theory attributes the biphasic pattern to the dynamic competition between cue-target integration and segregation, two opposing processes. While the integration process favors integrating the targets at the cued locations into an existing episodic representation (an object file) that has been activated by the preceding peripheral cue (Kahneman et al., 1992 [2](#)), the segregation process tends to create a new episodic representation for targets at the uncued locations (Funes et al., 2008 [2](#); Lupiáñez & Funes, 2005 [2](#); Lupiáñez et al., 2001 [2](#); Milliken et al., 2000 [2](#)). At short SOAs (e.g., around less than 250 ms), the integration process in the cued trials is more efficient than the segregation process in the uncued trials, resulting in faster responses for the cued than the uncued trials (i.e., the facilitation effect). At longer SOAs, however, the original object file activated by the cue gradually closed, making it less efficient to integrate new stimuli. Consequently, constructing a new object file at the uncued location gradually becomes easier than updating the closing one, resulting in the IOR effect.

Over the past two decades, the integration-segregation theory has been widely accepted as a flexible and extensible framework for explaining the accumulating IOR research findings (Chen et al., 2007 [2](#); Funes et al., 2008 [2](#); Hu et al., 2011 [2](#); Li et al., 2018 [2](#); Luo et al., 2010 [2](#); Lupiáñez et al., 2001 [2](#); Lupiáñez et al., 2007 [2](#); Zu et al., 2023 [2](#)), especially showing strengths in explaining the diverse research findings across sensory modalities, feature domains, and task contexts.

Originally developed to explain spatial IOR (Funes et al., 2008 [2](#); Lupiáñez & Funes, 2005 [2](#); Lupiáñez et al., 2001 [2](#); Milliken et al., 2000 [2](#)), it has subsequently been extended to nonspatial (e.g., color-, shape-, or frequency-based) IOR in complex environments (Chen et al., 2007 [2](#); Hu et al., 2011 [2](#)). When the target and the cue share both non-spatial and spatial features, these consecutive stimuli are integrated into a single event representation, hindering the detection of the target (Chen et al., 2007 [2](#); Hu et al., 2011 [2](#)). Such integrative interference has also been observed across sensory domains—including the auditory modality—and more recently, in supramodal contexts involving abstract semantic features (Zu et al., 2023 [2](#)). Researchers have further extended the explanatory scope of the integration-segregation framework to electrophysiological data (Li et al., 2018 [2](#); Martín-Arévalo et al., 2014 [2](#), 2016 [2](#)). For example, enhanced P3 amplitudes at the uncued locations have been interpreted as reflecting greater cognitive demands associated with new object-file creation underlying IOR (Li et al., 2018 [2](#)), whereas reduced P1 amplitudes at the cued locations have been taken to index a perceptual detection cost arising from disrupted cue-target integration (Martín-Arévalo et al., 2014 [2](#), 2016 [2](#)).

In addition, the integration-segregation theory also succeeds in accounting for findings that are difficult to explain by other attentional theories, particularly the task-dependent variations in IOR (Chen et al., 2007 [2](#); Lüpáñez & Milliken, 1999 [2](#); Lüpáñez et al., 2001 [2](#); Lüpáñez et al., 2007 [2](#)). According to this framework, the timing of IOR is attributed to task demands, with exogenous orientation being modulated top-down by task relevance. It predicts earlier IOR onsets in detection tasks (favoring an early file closure for new event encoding) and delayed onsets in discrimination tasks (supporting late closure for information accumulation) (Lüpáñez & Milliken, 1999 [2](#); Lüpáñez et al., 2001 [2](#)). Lüpáñez et al. (2007) [2](#) further extended this view by showing that, in discrimination tasks, facilitation for infrequent targets also occurred at long SOAs when IOR for frequent targets took place. This finding is incompatible with the attentional capture and disengagement accounts, and instead suggests cue-target integration dependent on the task set (see also its three-factor extension; Lüpáñez, 2010 [2](#)). In auditory attention research, IOR has also been found to vary with task demands, as task-irrelevant features can either enhance or eliminate the IOR effect depending on whether the cue and target share the same task-relevant dimension (Chen et al., 2007 [2](#)). As reported by Chen et al. (2007) [2](#), this pattern is better explained by the integration-segregation theory than by the traditional accounts. Collectively, by emphasizing the cue-target interplay, the integration-segregation account provides a comprehensive interpretation of exogenous attention within a unified theoretical framework by incorporating stimulus features, modalities, and task demands.

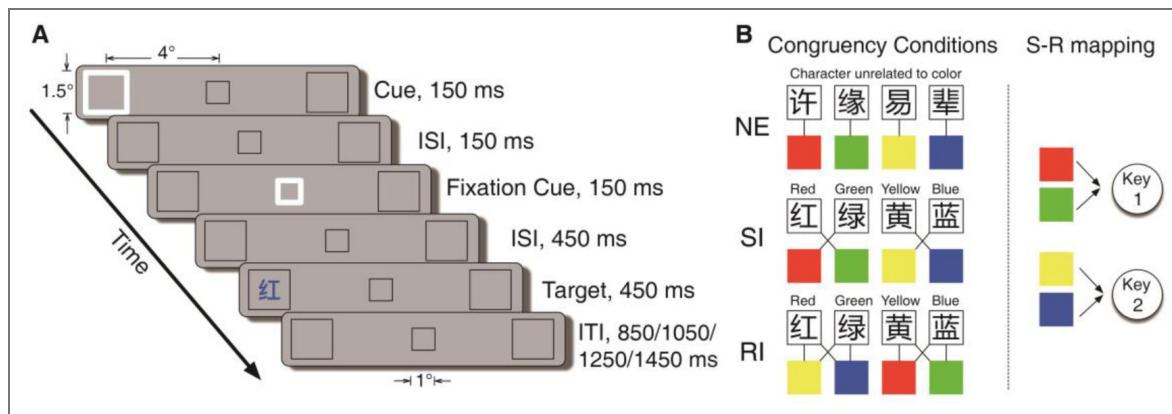
Despite these strengths, current support for the theory remains largely inferential. Specifically, the hypothesized dual processes of integration and segregation have not been directly evidenced in brain activities. This gap likely reflects that the empirical research had not kept up pace with the theoretical advances, with most of the early neuroimaging work being conducted when the theoretical account of IOR was still evolving and the integration-segregation framework not yet fully formulated. These studies generally assumed that IOR reflects a time-dependent inhibitory state that builds up with the increasing SOA, and that stronger brain activation was expected at longer SOAs regardless of cue validity (Lepsién & Pollmann, 2002 [2](#); Mayer, Seidenberg, et al., 2004 [2](#)). Following this assumption, the past studies typically contrasted long and short SOAs to capture the neural dynamics underlying the inhibitory phase of visual attentional orienting, either by collapsing the cued and uncued trials together (Lepsién & Pollmann, 2002 [2](#); Mayer, Dorflinger, et al., 2004 [2](#); Müller & Kleinschmidt, 2007 [2](#); Zhou & Chen, 2008 [2](#)) or by performing the short-vs. long-SOA comparison separately for each cueing condition (Mayer, Seidenberg, et al., 2004 [2](#)). These studies observed the involvement of the frontoparietal attention network, particularly the frontal eye fields (FEF), anterior cingulate cortex (ACC), and inferior parietal lobule (IPL). However, such SOA-based contrasts are insufficient for testing the integration-segregation framework, because they only capture the temporal changes in attentional orienting rather than the functional distinction (integration vs. segregation) that characterizes the theory. To directly test the dual processes of event integration (for the cued targets) and segregation (for the uncued targets), it is necessary to compare the cued and uncued conditions. Some studies did attempt this direct comparison but did not reveal reliable neural differences between the cued and uncued trials during the inhibitory period (Chen et al., 2006 [2](#); Lepsién & Pollmann, 2002 [2](#); Mayer, Seidenberg, et al., 2004 [2](#)). For instance, Chen et al. (2006) [2](#) reported a cue validity effect, but only in the left FEF. The limited neuroimaging evidence for distinct neural responses to the cued and uncued targets in IOR research was likely due to statistical power constraints inherent in event-related functional magnetic resonance imaging (ER-fMRI) experiments (despite their high psychological validity, i.e., estimation efficiency), further aggravated by the suboptimal temporal structure of stimulus sequences, and the limited sample sizes and trial numbers (Buracas & Boynton, 2002 [2](#); Liu, 2004 [2](#); Liu & Frank, 2004 [2](#); Liu et al., 2001 [2](#); Wager & Nichols, 2003 [2](#)). In the current study, we aimed at obtaining direct neuroimaging evidence for the integration-segregation theory employing ER-fMRI with an optimized stimulus sequence following the genetic algorithm (GA) framework (Wager & Nichols, 2003 [2](#)). GA is a class of flexible search algorithms that iteratively optimize sequences based on multiple fitness measures to improve the statistical power of contrast detection while taking into account the estimation efficiency of hemodynamic response function (HRF). According to the integration-segregation theory, targets appearing at the

previously cued locations will be engaged by the integration process for an update of the existing cue-activated object file. In contrast, targets appearing at the uncued locations should trigger the segregation process to form a new object file. Therefore, we predicted dissociable neural activations for the cued versus uncued targets. Specifically, we hypothesized that the uncued targets would show a greater engagement of the regions involved in new episodic encoding, such as the parahippocampal gyrus (PHG) (Burgess et al., 2002 [2](#); Danieli et al., 2023 [3](#); Hayes et al., 2007 [4](#); Li et al., 2016 [5](#); Menon et al., 2000 [6](#); Torres-Morales & Cansino, 2024 [7](#)), and that the cued targets would show a stronger activation of the regions involved in information integration and attentional reorienting, such as the FEF (Astafiev et al., 2003 [8](#); Corbetta & Shulman, 2002 [9](#); Liu et al., 2023 [10](#)).

In addition to the main comparison of the cued vs. uncued locations (i.e., IOR generation mechanisms), we also took the opportunity to investigate the IOR expression by systematically manipulating the type of target stimuli by varying the color-response key mappings using a Stroop paradigm (De Houwer, 2003 [11](#); Veen & Carter, 2005 [12](#); Veen et al., 2001 [13](#)). Three types of stimuli were presented: a neutral condition (non-color words shown in color, producing no conflict), a semantic conflict condition (word meaning and ink color were incongruent, but mapped to the same response), and a combined semantic-response conflict condition (word meaning and ink color were mismatched and mapped to different responses). This design enabled us to examine how spatial attention interacts with distinct levels of cognitive conflict. A previous fMRI study by Chen et al. (2006) [14](#) reported dissociable neural signatures for the semantic and response conflicts when spatial attention was engaged. Specifically, the increased dorsolateral prefrontal cortex (DLPFC) activity at the cued locations was interpreted within the inhibitory tagging framework as reflecting a temporary disconnection of pre-response representations from response processes. However, that study manipulated response eligibility by excluding certain incongruent color words from the response set (Milham et al., 2001 [15](#)). This method was criticized for conflating the semantic and response conflicts, as ineligible distractors may not be processed in the same way as task-relevant words (Veen & Carter, 2005 [12](#)). To overcome this limitation, the present study adopted an improved Stroop design that clearly separates the semantic and response conflicts (De Houwer, 2003 [11](#); Veen & Carter, 2005 [12](#); Veen et al., 2001 [13](#)). This approach provided a more precise test of whether spatial attention differentially modulates these two conflict types and their associated neural mechanisms.

## Results

Participants performed a spatial cueing task (long SOA to elicit IOR) combined with a Stroop paradigm adapted for colored Chinese characters (Chen et al., 2006 [14](#)) (Fig 1A [16](#)), with the characters appearing at either the cued or the uncued location. The experimental manipulation dissociated the semantic and response conflicts, following a well-established three-condition design (Fig 1B [17](#)). These conditions were neutral (NE; non-color characters), semantically incongruent (SI; the character and the color are incongruent but mapped to the same response, causing only the semantic conflict), and response-incongruent (RI; the character and the color are incongruent and mapped to opposite responses, causing both the semantic and response conflicts) (De Houwer, 2003 [11](#); Veen & Carter, 2005 [12](#)). Participants responded using two keys, each assigned to two colors. To ensure sufficient statistical power for detecting condition-specific neural differences, the ER-fMRI design was optimized using the Genetic Algorithm (Wager & Nichols, 2003 [18](#)), with the goal of maximizing experimental efficiency for three contrasts of interest. To directly evaluate the prediction of the integration-segregation theory, we first examined the brain activity differences between the conditions of cued-NE (targets at the cued location in the neutral condition) and uncued-NE (targets at the uncued location in the neutral condition). Subsequent analyses examined how IOR modulated the conflict-related neural activities. The contrast of cued-SI minus cued-NE vs. uncued-SI minus uncued-NE was used to assess the effect of IOR on semantic conflict processing, whereas the contrast of cued-RI minus cued-SI vs. uncued-RI minus uncued-SI was employed to capture the modulation of response conflict by IOR.



**Fig 1. Experimental Materials.**

**A.** Trial sequence and display sizes. Each trial starts with a 150-ms non-informative cue at one of the peripheral boxes, followed by a 150-ms fixation cue with a 300-ms SOA. The target was a 450-ms colored Chinese character presented 600 ms after the fixation cue onset at the two target locations with equal probabilities. ISI = interstimulus interval; ITI = intertrial interval. **B.** The character-color combinations in the three congruency conditions. In the neutral condition (first row), the characters were not color-related. In the other conditions, the characters were color names (translation added for illustration purposes). S-R mapping = stimulus-response mapping; NE = neutral; SI = semantically incongruent; RI = response-incongruent.

## Behavioral Results

Mean reaction times (RTs) and accuracies are shown in Fig 2A. A two-way (Cue Validity x Congruency) repeated-measures analysis of variance (rm-ANOVA) for the RTs revealed a significant IOR effect (main effect of Cue Validity),  $F(1, 28) = 12.057, p = .002, \eta_p^2 = 0.301$ , showing slower responses to targets at the cued location ( $M = 642$  ms,  $SE = 20$  ms) than at the uncued location ( $632 \pm 19$  ms). The main effect of Congruency was also significant,  $F(1.53, 42.88) = 29.602, p < .001, \eta_p^2 = 0.514$  (the Greenhouse-Geisser correction was applied due to the violation of the sphericity assumption). Post hoc comparisons with the Holm-Bonferroni correction (Holm, 1979) revealed significant differences among all conditions (NE vs. RI:  $t(28) = -2.179, p = .038$ , Cohen's  $d = 0.071$ ; NE vs. SI:  $t(28) = -5.957, p < .001$ , Cohen's  $d = 0.275$ ; SI vs. RI:  $t(28) = -6.715, p < .001$ , Cohen's  $d = 0.203$ ), with NE showing the shortest RT ( $624 \pm 19$  ms), followed by SI ( $632 \pm 20$  ms) then RI ( $654 \pm 21$  ms). These data clearly isolated the two distinct conflict effects in the Stroop effect, namely the semantic conflict (SI-NE difference) and the response conflict (RI-SI difference). The interaction did not approach significance,  $F(2, 56) = 0.930, p = .401, \eta_p^2 = 0.032$ .

The rm-ANOVA for the accuracy data (Fig 2B) only showed a significant main effect of Congruency,  $F(2, 56) = 7.685, p = .001, \eta_p^2 = 0.215$ . Post hoc comparisons confirmed that this came from a lower accuracy in the RI condition ( $M = 0.943 \pm SE = 0.009$ ) than in the NE ( $0.966 \pm 0.009$ ;  $t(28) = -3.596, p = .002$ , Cohen's  $d = 0.446$ ) and SI ( $0.963 \pm 0.009$ ;  $t(28) = -3.150, p = .005$ , Cohen's  $d = 0.391$ ) conditions. No significant difference was found between the NE and SI conditions ( $t(28) = 0.446, p = .657$ , Cohen's  $d = 0.055$ ). The main effect of Cue Validity ( $F(1, 28) = 0.021, p = .887, \eta_p^2 < 0.001$ ) and the interaction ( $F(2, 56) = 1.298, p = .281, \eta_p^2 = 0.044$ ) were not significant. No speed-accuracy trade-off was noticed in the current data.

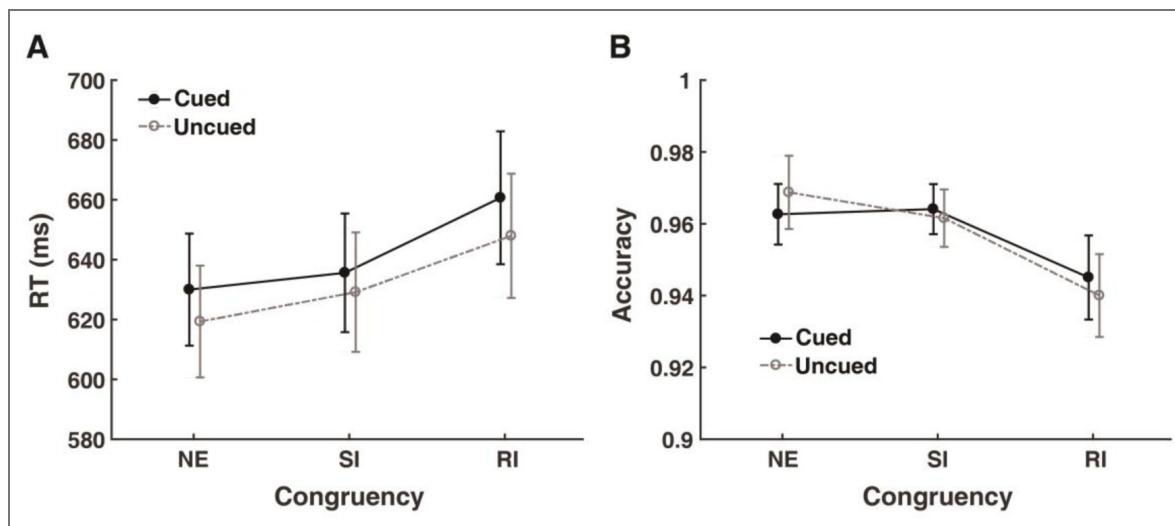
## Neuroimaging Results

### IOR Effect in the Neutral Condition

The contrast between the cued-NE and uncued-NE conditions was examined to identify the underlying neural mechanisms of the IOR effect during the processing of neutral targets. Whole-brain fMRI findings revealed two distinct activation patterns in response to these conditions (Fig 3A). Relative to the uncued-NE condition, the cued-NE condition showed enhanced activations in the dorsal attention network (DAN) including the bilateral FEF and IPS, along with the right-lateralized TPJ from the ventral attention network (VAN), and the left dACC. In contrast, the uncued-NE condition demonstrated stronger activations in the bilateral PHG and STG than the cued-NE condition. To further compare the activity levels in each brain region between the cued-NE and uncued-NE conditions, paired *t*-tests were conducted on the average parameter estimates (beta weights) in the left and right IPS, FEF, PHG, and STG, as well as the left dACC and right TPJ (all  $p < .001$ , see Fig 3B). Detailed information on the activated regions' coordinates, cluster sizes, and statistical significance is provided in Table 1.

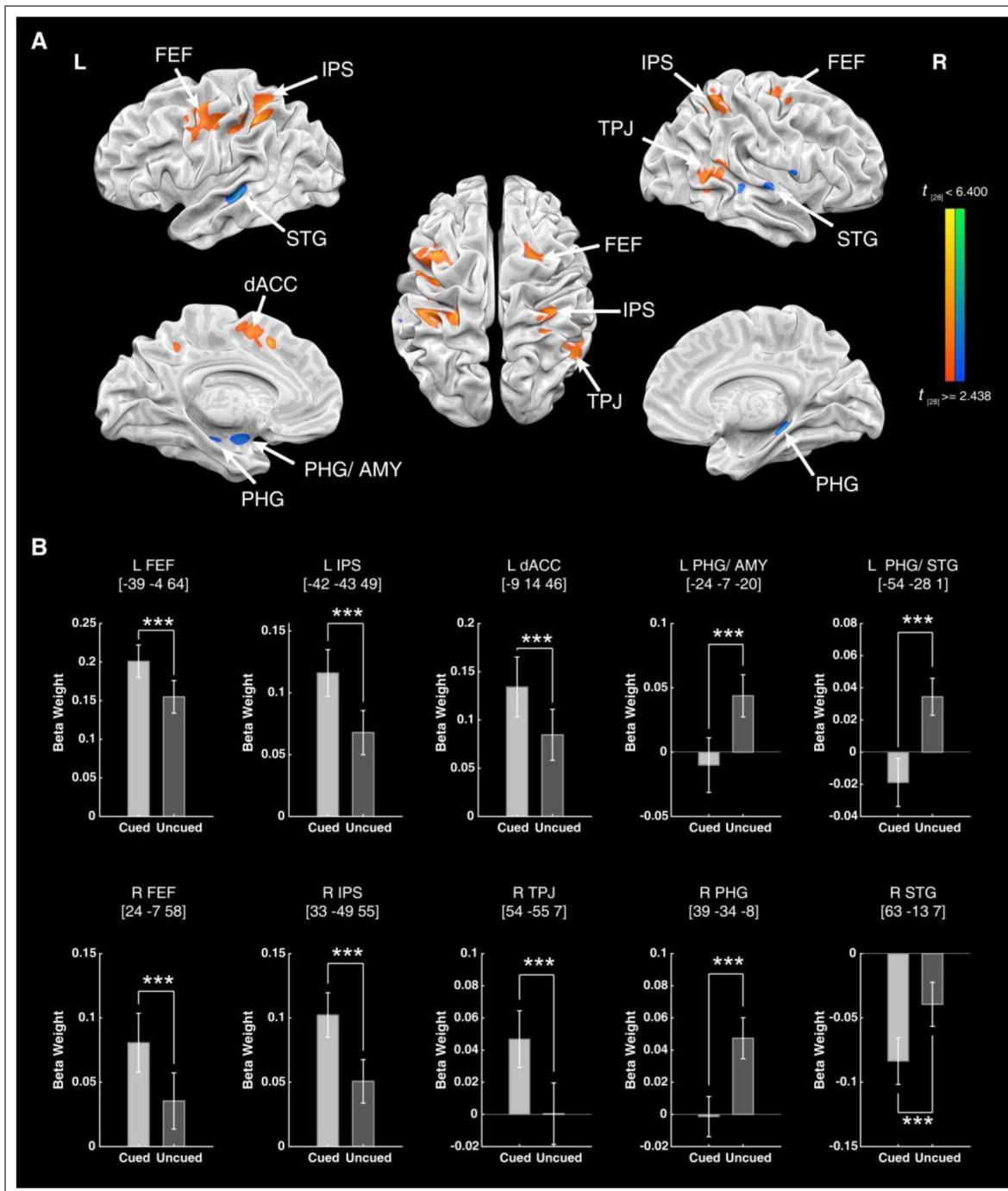
### Effect of IOR on Semantic Conflict

Although the IOR effect showed no effect on either the semantic conflict difference (SI-NE) or the response conflict difference (RI-SI) in the behavioral performance, differential neural activities were observed between these conditions (summarized in Fig 4A and Table 2). The effect of IOR on the semantic conflict was examined as the contrast between the SI-NE differences (SI minus NE) in the cued and the uncued conditions. As illustrated in Fig 4A, the right dACC showed significantly reduced activation. A two-way rm-ANOVA was conducted on the average parameter estimates (beta weights) obtained from these contrasts for each activated region. The results confirmed a significant interplay between semantic conflict and IOR in the right dACC,  $F(1, 28) = 15.946, p < .001, \eta_p^2 = 0.363$ . Greater neural activities were found in the SI condition compared to the NE condition when the targets were presented at the uncued location ( $t(28) = 3.262, p = .003$ , Cohen's  $d = 0.606$ ), but not for the targets at the cued location ( $t(28) = -1.010, p = .321$ , Cohen's  $d = 0.187$ ).



**Fig 2. Behavioral Results.**

Mean reaction times (A) and accuracies (B) as a function of cue validity and congruency. Error bars extend to one standard error of the mean (SEM).



**Fig 3. IOR Effect in the Neutral Condition and Parameter Estimation.**

**A.** Brain regions showing significant activations in the contrast between the cued-NE and uncued-NE conditions, with a threshold of  $p < 0.005$  (uncorrected) with a minimum cluster size of  $540 \text{ mm}^3$  (20 voxels), yielding a corrected  $p < 0.05$  based on 2,500 Monte Carlo simulations in BrainVoyager. Warm colors represent stronger activations in the cued condition, and cold colors represent stronger activations in the uncued condition. **B.** Parameter estimates for each activation region. Error bars extend to 1 SEM. L = left; R = right. \*\*\*  $p < .001$ .

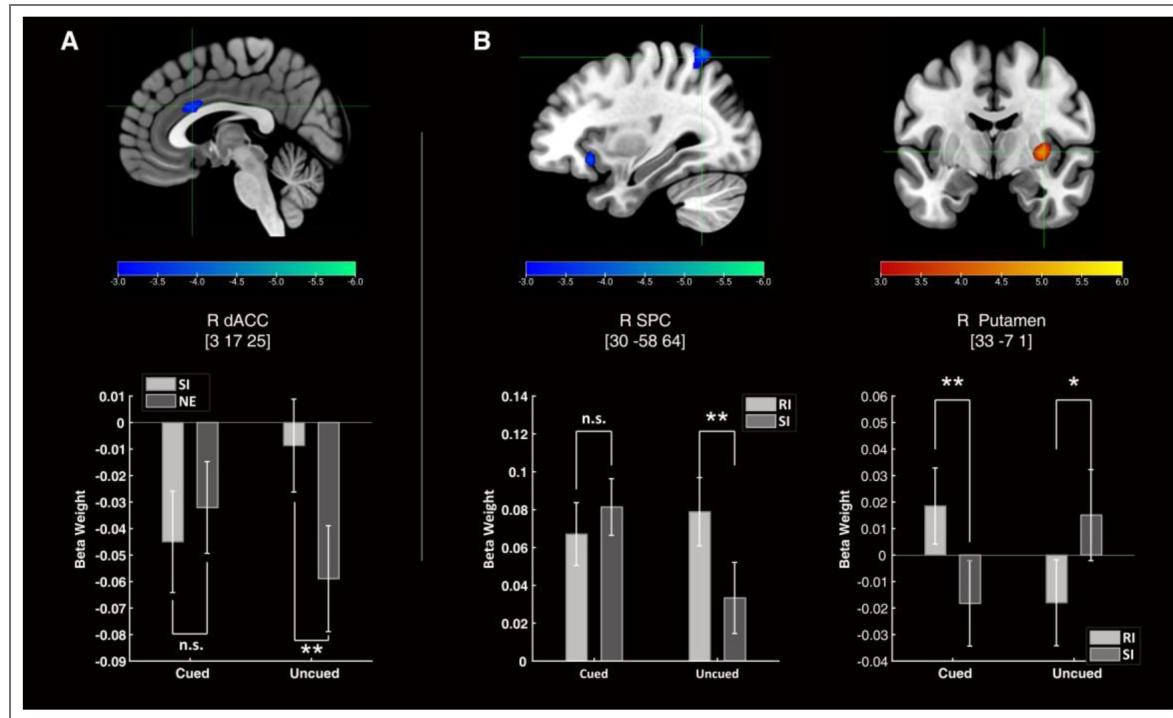
Regions	Laterality	Cluster (voxels)	MNI			T <sub>max</sub>	T <sub>mean</sub>	BA			
			coordinates peak								
			x	y	Z						
Frontal Eye Field	L	146	-39	-4	64	4.51	3.43	6			
	R	28	24	-7	58	4.32	3.52	6			
Intraparietal Sulcus	L	120	-42	-43	49	4.73	3.57	40			
	R	82	33	-49	55	5.12	3.78	40			
Dorsal Anterior Cingulate Cortex	L	36	-9	14	46	5.05	3.46	32			
Temporoparietal Junction	R	75	54	-55	7	4.23	3.45	39			
Parahippocampal Gyrus	L	22	-24	-7	-20	-4.11	-3.46	/			
	R	44	39	-34	-8	-5.44	-3.83	/			
Parahippocampal Gyrus/ Superior Temporal Gyrus	L	84	-54	-28	1	-4.42	-3.49	21			
Superior Temporal Gyrus	R	32	63	-13	7	-4.22	-3.48	22			

BA = Brodmann area; L = left; R = right.

**Table 1.** Brain regions showing significant activation differences between the cued-NE and uncued-NE conditions.

**Fig 4. IOR Effect under the Semantic Conflict (A) and the Response Conflict (B).**

These two effects were quantified as cued-SI minus cued\_NE > uncued-SI minus uncued-NE and cued-RI minus cued-SI > uncued-RI minus uncued-SI, respectively. Parameter estimations were based on a threshold of  $p < 0.005$  (uncorrected), with a minimum cluster size of  $540 \text{ mm}^3$  (20 voxels), yielding a corrected  $p < 0.05$  based on 2500 Monte Carlo simulations in BrainVoyager. Error bars extend to 1 SEM.  $**p < .01$ ,  $*p < .05$ , n.s. = non-significant.



**Table 2.** Brain regions showing a significant modulation effect of IOR on semantic conflict (cued-SI minus cued\_NE > uncued-SI minus uncued-NE) or response conflict (cued-RI minus cued-SI > uncued-RI minus uncued-SI).

Region	Laterality	Cluster (voxels)	MNI			$T_{\max}$	$T_{\text{mean}}$	BA
			coordinates	peak	X			
<b>Effect of IOR on Semantic Conflict</b>								
Dorsal Anterior Cingulate Cortex	R	24	3	17	25	-3.78	-3.34	24
<b>Effect of IOR on Response Conflict</b>								
Superior Parietal Cortex	R	67	30	-58	64	-4.74	-3.49	7
Putamen	R	31	33	-7	1	4.91	3.80	/

BA = Brodmann area; R = right.

## Effect of IOR on Response Conflict

To explore the influence of IOR on response conflict, we compared the cued (RI-SI) and the uncued (RI-SI) conditions (Fig 4B [2](#) and Table 2 [2](#)). The right superior parietal cortex (SPC) showed a significant activation reduction (Fig 4B, left [2](#)), while the right putamen exhibited an activation enhancement (Fig 4B, right [2](#)). A two-way rm-ANOVA on the beta weights revealed a significant interaction in the right SPC,  $F(1,28) = 20.833, p < .001, \eta_p^2 = 0.427$ . Specifically, it showed greater activations in the RI condition compared to the SI condition when the targets were presented at the uncued location ( $t(28) = 3.447, p = .002$ , Cohen's  $d = 0.640$ ), but not for the cued location ( $t(28) = -0.962, p = .344$ , Cohen's  $d = 0.179$ ). The right putamen also demonstrated a significant interaction ( $F(1,28) = 26.686, p < .001, \eta_p^2 = 0.488$ ), but with a different pattern. The activation was stronger in the RI than the SI conditions for the cued location ( $t(28) = 2.983, p = .006$ , Cohen's  $d = 0.554$ ); and the opposite pattern was observed for the uncued location ( $t(28) = -2.404, p = .023$ , Cohen's  $d = 0.446$ ).

## Discussion

The integration-segregation theory has emerged as an influential framework for explaining the dynamic effects of exogenous attention (Chen et al., 2007 [2](#); Funes et al., 2008 [2](#); Hu et al., 2011 [2](#); Li et al., 2018 [2](#); Lupiáñez et al., 2001 [2](#); Lupiáñez et al., 2007 [2](#); Zu et al., 2023 [2](#)), attributing the turning from the early attentional facilitation to the later IOR to the dynamic processes of cue-target integration and segregation (Funes et al., 2008 [2](#); Lupiáñez & Funes, 2005 [2](#); Lupiáñez et al., 2001 [2](#); Milliken et al., 2000 [2](#)). In the current study, by contrasting the cued versus the uncued targets, we provided the first direct neuroimaging evidence supporting this theory by dissociating brain activation patterns associated with these two processes. Stronger responses were found in the bilateral FEF, IPS, right TPJ, and left dACC in the cued than the uncued targets. Relative to the cued targets, the uncued targets showed stronger activations in the bilateral PHG and STG.

The heightened activation observed for targets appearing at the cued locations (particularly in the bilateral FEF, IPS, and right TPJ) reflects the attentional demand associated with the integration process. According to the integration-segregation theory (Funes et al., 2008 [2](#); Lupiáñez & Funes, 2005 [2](#); Lupiáñez et al., 2001 [2](#); Milliken et al., 2000 [2](#)), the cue-initiated object file is likely to have closed or begin closing under the long SOA conditions, hindering immediate integration of the subsequent targets. To integrate a target appearing again at the cued location, the object file needs to be reopened with reallocation of attentional resources. Our neuroimaging data captured this process by showing coordinated activation in the bilateral FEF and IPS (key nodes of the dorsal attention network) and in the right TPJ (a core region of the ventral attention network) (Ahrens et al., 2019 [2](#); Corbetta & Shulman, 2002 [2](#); Fox et al., 2006 [2](#); Vossel et al., 2014 [2](#)). These regions act in concert to support the attentional shifts and reorienting necessary for reopening the object file for integration. Additionally, the observed increase in left dACC activity under the cued relative to the uncued condition likely reflects the engagement of cognitive control mechanisms (Botvinick et al., 2004 [2](#); Chung et al., 2024 [2](#); Mayer et al., 2012 [2](#); Veen & Carter, 2005 [2](#)) to resolve the difficulty with the task-driven requirement of target integration facing the reduced accessibility of the cue-initiated representation. It is also possible that the heightened activation of dACC represents the inhibitory bias toward the direction of the cued location (Mayer, Seidenberg, et al., 2004 [2](#)).

In contrast, the stronger activation in the uncued condition was observed in the bilateral PHG extending into the STG. Located in the medial temporal lobe surrounding the hippocampus, the PHG is critically involved in episodic encoding, particularly for novel visual or spatial stimuli (Burgess et al., 2002 [2](#); Danieli et al., 2023 [2](#); Hayes et al., 2007 [2](#); Li et al., 2016 [2](#); Menon et al., 2000 [2](#); Torres-Morales & Cansino, 2024 [2](#)). In our task, the targets appearing at the uncued locations likely represented novel events requiring new spatial registration. Following the theorized segregation process, such events require the creation of new object files (Kahneman et

al., 1992) by engaging brain regions involved in novelty detection and contextual updating. The enhanced PHG/STG activities observed in the uncued condition may reflect the encoding of new spatial representations of the segregation process.

Our data provided clear support for the integration-segregation theory. It is also noteworthy that, although prior studies investigated the neural mechanisms of IOR (Bourgeois et al., 2013a, 2013b; Hanlon et al., 2017; Lepsien & Pollmann, 2002; Mayer, Dorflinger, et al., 2004; Mayer et al., 2007; Mayer, Seidenberg, et al., 2004; Müller & Kleinschmidt, 2007; Satel et al., 2019; Yang & Mayer, 2014; Zhou & Chen, 2008), none identified distinct activation patterns corresponding to the integration and segregation processes as in our data. Specifically, most of the previous IOR studies did not show significant brain activations when contrasting the cued and uncued conditions (Lepsien & Pollmann, 2002; Mayer, Seidenberg, et al., 2004), except Chen et al. (2006) reported a cue-validity effect confined to the left FEF. Instead, some indirect approaches, such as comparing long- and short-SOA trials while collapsing over the cueing conditions, reported activations in regions like the FEF, TPJ, ACC, and posterior parietal cortex (Lepsien & Pollmann, 2002; Mayer, Dorflinger, et al., 2004; Mayer, Seidenberg, et al., 2004; Müller & Kleinschmidt, 2007; Zhou & Chen, 2008), showing some similarity with the integration-related network observed in the current study. However, the findings were inconsistent across studies, with some reporting only a limited subset of regions and others showing lateralized instead of bilateral effects (e.g., stronger right-hemisphere FEF activation; Lepsien & Pollmann, 2002; Mayer, Dorflinger, et al., 2004). Similar frontoparietal engagement has also been observed in auditory and cross-modal IOR studies (Hanlon et al., 2017; Mayer et al., 2009; Mayer et al., 2007; Yang & Mayer, 2014), typically present across various SOAs (e.g., sustained activation in both the dorsal and ventral frontoparietal regions regardless of SOA length; Hanlon et al., 2017) or showing SOA-dependent effects (e.g., reversed direction of activation differences between short and long SOAs; Mayer et al., 2007). Complementing these observations, transcranial magnetic stimulation (TMS) studies have provided causal evidence for the contribution of frontoparietal regions in IOR (Bourgeois et al., 2013a, 2013b; Chica et al., 2011; Ro et al., 2003). For instance, stimulation over the right FEF during the cue-target interval has been shown to eliminate the typical IOR effect for the cued targets in the ipsilateral hemifield (Ro et al., 2003). Similarly, TMS applied to the right IPS/TPJ also disrupted the IOR effect (Bourgeois et al., 2013a; Chica et al., 2011), whereas stimulation over their left-hemisphere counterparts did not cause much change in IOR (Bourgeois et al., 2013b). These findings suggest a possible right-lateralized neural organization of the integration process. However, this lateralization notion conflicts with the largely bilateral activation pattern observed in our study. The lack of systematic testing for the left-hemisphere contribution in previous TMS studies leaves this asymmetry open to further investigation. Notably, despite offering partial (and often lateralized) support for the integration process, none of these prior studies have addressed the neural mechanisms underlying the segregation process, which is uniquely revealed by the present neuroimaging findings.

The above discrepancies between our findings and the previous studies may stem from several methodological and design factors. Firstly, the prior studies likely introduced confounds when investigating IOR indirectly. When comparing long and short SOAs, the observed effects may have been jointly influenced by factors unrelated to IOR, such as working memory (i.e., increased demand of maintaining the cue representation over longer intervals; Mayer et al., 2007) and temporal attention (i.e., distinct temporal expectations formed by variations in SOA; Nobre & van Ede, 2018). Moreover, the IOR effect depends not only on cue-induced attentional orienting, but also on the dynamic interaction between the target onset and the ongoing cue-related neural activity (Lupiáñez, 2010; Nobre & van Ede, 2018; Taylor & Donnelly, 2002). These confounds could potentially obscure the genuine IOR effect. Secondly, differences in statistical power may also account for the discrepancies. In the present study, we employed an optimized GA stimulus sequence (Wager & Nichols, 2003), which provides greater statistical power than simple random sequences while maintaining a high estimation efficiency (for details, see the Methods and Supplementary Information sections). This optimization likely enhanced the reliability of the estimated neural responses (Wager & Nichols, 2003). In addition, the previous neuroimaging

studies on IOR often relied on relatively small sample sizes (around 10-12 participants; Chen et al., 2006 [\(2\)](#); Mayer, Dorflinger, et al., 2004 [\(2\)](#); Mayer, Seidenberg, et al., 2004 [\(2\)](#); Müller & Kleinschmidt, 2007 [\(2\)](#)) or a limited number of trials (e.g., 30 trials per condition; Lepsién & Pollmann, 2002 [\(2\)](#)), leading to much reduced statistical power and a higher probability of false negatives. In contrast, the current study increased both the number of trials and the sample size, having effectively enhanced the sensitivity of detecting differences between experimental conditions (Baker et al., 2021 [\(2\)](#); Chen et al., 2022 [\(2\)](#)). Finally, task design differences may further contribute to the observed inconsistencies. The earlier studies often employed simple localization or detection tasks (Lepsién & Pollmann, 2002 [\(2\)](#); Mayer, Dorflinger, et al., 2004 [\(2\)](#); Mayer, Seidenberg, et al., 2004 [\(2\)](#); Müller & Kleinschmidt, 2007 [\(2\)](#)), while the current study adopted a discrimination task. According to the integration-segregation theory (Funes et al., 2008 [\(2\)](#); Lupiáñez & Funes, 2005 [\(2\)](#); Lupiáñez et al., 2001 [\(2\)](#); Lupiáñez et al., 2007 [\(2\)](#); Milliken et al., 2000 [\(2\)](#)), more complex stimuli may require greater cognitive resources to establish object files, leading to enhanced processing of object files and heightened detectability of the underlying integration and segregation processes.

Another novelty of the current study is integrating the IOR and the modified Stroop tasks, which separately studied semantic-and response-level conflicts (De Houwer, 2003 [\(2\)](#); Veen & Carter, 2005 [\(2\)](#)). Through this design, we made an additional discovery about how IOR modulates the ongoing Stroop interference effect at the inhibited (i.e., cued) locations. Behaviorally, our results showed no significant interaction between IOR and any conflict in the Stroop task, not replicating the previous findings (Chen et al., 2006 [\(2\)](#); Vivas & Fuentes, 2001 [\(2\)](#)) of reduced Stroop interference at the cued relative to the uncued locations. Yet, at the neural level, the brain regions involved in conflict processing were engaged in the interaction between IOR and the Stroop effect. Specifically, the right dACC, which is involved in semantic conflict processing (Li et al., 2017 [\(2\)](#); Milham et al., 2001 [\(2\)](#); Veen & Carter, 2005 [\(2\)](#)), appeared to serve as a critical neural interface for the interaction between semantic conflict and IOR. Specifically, in the uncued condition, the semantic incongruity elicited stronger activations compared to the neutral condition, a pattern that disappeared or even reversed in the cued condition. Regarding the interaction between response conflict and IOR, brain regions such as the right SPC, which are involved in detecting response conflict and orienting spatial attention (Li et al., 2017 [\(2\)](#)), played a key role. Similarly, this region exhibited stronger conflict effects (i.e., greater activation in the RI than SI condition) in the uncued condition compared to the cued condition. These results can be interpreted by the inhibitory tagging mechanism proposed by Fuentes et al. (1999 [\(2\)](#)), which posits that, when attention is drawn away from a cued location, stimuli presented there are temporarily tagged with inhibition (Fuentes et al., 2000 [\(2\)](#); Fuentes et al., 1999 [\(2\)](#); Vivas & Fuentes, 2001 [\(2\)](#)). ERP evidence supporting this mechanism was reported by Zhang et al. (2012 [\(2\)](#)), who showed that the Stroop conflict-related N450 effects were delayed and attenuated at the cued compared to the uncued locations, suggesting a temporary disruption of the stimulus-response link. Such inhibitory tagging may attenuate or even disrupt conflict processing at the inhibited location, offering a plausible account for the neural interactions between IOR and Stroop conflicts observed in our study. The current results could also potentially suggest that the effects of inhibitory tagging are not limited to stimulus-response connections (as proposed by Fuentes et al., 1999 [\(2\)](#)), but also extend to semantic representations, as evidenced by the modulation of the right dACC observed in our study. This notion is consistent with a previous finding that the N400 ERP component (a biomarker of semantic processing) had a decreased amplitude for the cued position (Zhang & Zhang, 2007 [\(2\)](#)). This highlights that spatial attention can affect subsequent cognitive processes at the semantic level (Cristescu & Nobre, 2008 [\(2\)](#); Zhang & Zhang, 2007 [\(2\)](#)).

Furthermore, we observed pronounced neural responses in the right putamen when contrasting the RI and SI conditions at the cued versus the uncued locations. The putamen is a subcortical nucleus in the basal ganglia and has been found to be involved in control of response interference (Schmidt et al., 2018 [\(2\)](#); Schmidt et al., 2020 [\(2\)](#)). For example, Schmidt et al. (2020 [\(2\)](#)) demonstrated that the dorsal striatum, including the putamen, is engaged during Simon-type interference by supporting task-appropriate response selection and suppression of competing alternatives, and that its damage leads to less efficient interference control (Schmidt et al., 2018 [\(2\)](#)). These findings support the view that the putamen is recruited when interference arises at the response-selection

level. Building on this, we speculated that the enhanced putamen activation in the cued conditions in the current study reflects an increased demand for response control when attentional resources were reduced by IOR. Taken together, our findings highlight a potential neural basis for the interaction between IOR and conflict processing encompassing both semantic and response domains.

However, these results should be interpreted with caution given the absence of behavioral support. One potential explanation of this dissociation is the usage in the current study of the GA-optimized sequence, which prioritizes both detection and HRF estimation efficiency. This optimization made the event order converging towards a block-like structure, thereby having reduced event counterbalancing and increased sequential predictability (Wager & Nichols, 2003 [2](#)). Such unintended regularities may have influenced participants' behavioral strategies (e.g., forming expectations about upcoming events), thereby weakening the correspondence between the neural and behavioral findings. Future studies should address this limitation by employing more optimized designs that consider some psychological factors (e.g., event counterbalancing; Wager & Nichols, 2003 [2](#)) to better validate the observed neural mechanisms. Alternatively, the observed neural-behavioral response dissociation may reflect differences in measurement sensitivity between the neural and behavioral indices (Chen et al., 2006 [2](#); Wilkinson & Halligan, 2004 [2](#)). A similar pattern of dissociation was reported by Chen et al. (2006 [2](#)), who found that the response conflict was not modulated by IOR behaviorally, yet the regions associated with conflict resolution exhibited stronger activations in the cued condition. As noted by Wilkinson and Halligan (2004 [2](#)), RTs and accuracies are not perfect measures of cognition, whereas neural signals can reveal finer-grained or "hidden" processes that precede overt behavior. Thus, neural modulations may emerge even in the absence of detectable behavioral differences.

In conclusion, the current study provides the first direct neuroimaging evidence lending support to the hypothesis of the integration-segregation theory (Funes et al., 2008 [2](#); Lupiáñez & Funes, 2005 [2](#); Lupiáñez et al., 2001 [2](#); Milliken et al., 2000 [2](#)). We revealed distinct neural mechanisms for processing of the cued and uncued targets during IOR, with attentional integration engaging the frontoparietal attention network (FEF, IPS, TPJ, dACC) and segregation recruiting the medial temporal regions (PHG-STG) associated with new object-file formation and novelty encoding. These dissociated activations offered direct support for the dynamic interplay between the integration and segregation processes. We also identified interactions between IOR and cognitive conflict in brain activities, suggesting that attentional orienting can modulate conflict processing at both the semantic and response levels. Taken together, our findings revealed the neural underpinnings of the integration-segregation theory and advanced our understanding of the neural mechanisms linking exogenous attentional orienting and cognitive control.

## Methods

### Participants

32 healthy participants with normal or corrected-to-normal vision and normal color vision were recruited. None reported a history of neurological disorders. Data from three participants were excluded due to excessive head movements and high global variances (see fMRI Data Analysis), leaving 29 participants for analysis (18 female, 11 male; mean age  $\pm$  SD = 22.69  $\pm$  2.58 years). All participants were naïve to the purpose of the study, provided written informed consent approved by the Ethics Committees of Northeast Normal University and Soochow University, and received monetary compensation. The sample size was informed by a power analysis using MorePower 6.0 (Campbell & Thompson, 2012 [2](#)) for a within-subjects rm-ANOVA. To achieve an 80% statistical power at the threshold of  $\alpha = .05$  (Chen et al., 2006 [2](#)), 14 participants were required. In addition, we also acknowledged that effect sizes from published studies are often inflated due to the publication bias (Albers & Lakens, 2018 [2](#)). To mitigate this potential risk, we determined to acquire data from a sample sized at least double of the suggested size by the power analysis (i.e.,  $N \geq 28$ ).

## Experimental Design

The experiment adopted a within-subjects design with two factors, namely Cue Validity (cued vs. uncued) and Congruency (SI, RI, and NE). The targets appeared at the cued location in the cued trials and at the other peripheral location in the uncued trials. The congruency factor referred to the relationship between the color and meaning of the Chinese characters (i.e., targets). In total, eight characters and four colors were used (see Fig 1B [2](#)). In an SI trial, the color and character meaning differed but were mapped to the same response key (e.g., character “红” [“red”] displayed in green, which was related to the same response key). In an RI trial, the color and character meaning differed and were mapped to different response keys (e.g., “红” displayed in yellow, which was related to another key). The NE trials used characters that were not related to color in the meanings and shared the same orthographic structures (character complexity and form) of the color-meaning characters. In addition to the six experimental conditions, we added a null condition, in which no Chinese character was shown, to increase the statistical power of detecting differences across conditions (Friston et al., 1999 [3](#)).

It is worth noting that the statistical power of effects in rapid ER-fMRI depends greatly on specific sequences of stimulus events (Liu & Frank, 2004 [4](#); Wager & Nichols, 2003 [5](#)). To ensure high design efficiency, we optimized the stimulus sequences employing the genetic algorithm (see Supplementary Information for details) (Wager & Nichols, 2003 [5](#)). This optimization improves the detection efficiency for the contrasts of interest by moderately sacrificing the efficiency of less relevant contrasts (Wager & Nichols, 2003 [5](#)). In the current study, we focused on three contrasts, including cued-NE vs. uncued-NE, cued-SI minus cued-NE vs. uncued-SI minus uncued-NE, and cued-RI minus cued-SI vs. uncued-RI minus uncued-SI. These contrasts respectively examined the IOR effect, the modulation of semantic conflict processing by IOR, and the modulation of response conflict processing by IOR. The optimized sequences were used for all but two participants, whose trial sequences were constructed using a truncated M-sequence (Buracas & Boynton, 2002 [6](#)) implemented in an earlier version of the experiment.

## Stimuli and Procedure

Each participant completed two functional scans (i.e., experimental runs) and one anatomical scan in a single session. Each experimental run employed a rapid event-related design and had each of the seven conditions (six experimental conditions plus the null condition) repeated 48 times (336 trials per run). Across the two runs, this yielded a total of 672 trials (96 trials per condition).

All trials displayed a three-box display over a gray background, including a central black fixation box ( $1^\circ \times 1^\circ$ , line width of 0.02°) and two black placeholder boxes ( $1.5^\circ \times 1.5^\circ$ , line width of 0.02°) positioned 4° (center-to-center) to the left and right of the fixation box. Each run began and ended with this display for 16 and 20 s, respectively. The trial sequence is illustrated in Fig 1A [2](#). In a null trial, only the three boxes were shown for the trial duration. In any of the six experimental conditions, each trial started with one of the peripheral boxes changing to a white color with a line width of 0.05° for 150 ms to attract attention to this peripheral location (cue). 150 ms after the offset of the peripheral cue, the central fixation box turned into white with a line width of 0.05° for 150 ms to force attention back to the central location (central cue). After another 450 ms, a colored Chinese character (in the STSong font,  $1.4^\circ \times 1.4^\circ$ ) was presented (target) for 450 ms inside one of the two peripheral boxes with equal probabilities. Participants were required to ignore the meaning of the character and identify the word color as quickly and accurately as possible by pressing one of the two keys designated for the color categories (red/green and blue/yellow) with their middle and index fingers, respectively (Fig 1B [2](#)). The color category-button mapping was counterbalanced across participants. Furthermore, to avoid a possible occurrence of the Simon effect (Klein & Ivanoff, 2011 [7](#)), the response keys were vertically arranged. Each trial ended with an inter-trial interval (ITI) with a duration of 850, 1,050, 1,250, or 1,450 ms (randomized with equal probabilities). The average trial duration was 2,500 ms.

Before the scanning, all participants had two practice parts outside the scanner to familiarize themselves with the task and the stimuli. In the first part, the participants practiced on a discrimination task with only color patches (no Chinese characters) using the predefined color category-button mapping. Once having reached an accuracy of 96%, the participants did the second part and completed 24 practice trials of the experimental task as in the scanning runs.

## Apparatus and Data Acquisition

The imaging data were acquired at two research sites following comparable protocols. At the Imaging Center for Brain Research of Beijing Normal University, the stimuli were presented with E-Prime (Psychological Software Tools, Pittsburgh, PA) on an LCD monitor (1024×768 resolution, 60 Hz refresh rate) viewed through a head-coil-mounted mirror (115 cm optical distance). The data were collected using a Siemens 3-Tesla Tim Trio scanner with a head coil. The functional data were acquired through a T2\*-weighted echo planar imaging (EPI) sequence (TR = 2,000 ms; TE = 30 ms; flip angle = 90°; FOV = 220×220 mm; matrix size = 64×64). Thirty-three transversal slices covering the whole brain (slice thickness = 4 mm; in-plane resolution = 3.44×3.44 mm; slice gap = 0.4 mm) were acquired in an interleaved ascending order. Each participant completed two functional runs of 400 volumes (including 8 initial dummy volumes). High-resolution anatomic images were collected using a T1\*-weighted magnetization-prepared rapid gradient echo (MP-RAGE) sequence consisting of 128 sagittal slices (TR = 2,300 ms; TE = 3.9 ms; flip angle = 8°; FOV = 256×256 mm, matrix size = 256×256, voxel resolution = 1.33×1×1 mm, slice gap = 0 mm).

At the Imaging Center of the First Affiliated Hospital of Soochow University, the stimuli were presented with MATLAB (The MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997 [\[2\]](#)) on a monitor (also 1920×1080 and 60 Hz) viewed through a mirror at an optical distance of 251 cm. The imaging data were recorded using a 3-Tesla Philips Ingenia scanner equipped with a head coil. The functional images featured a matrix size of 80×80, an in-plane resolution of 2.75×2.74 mm, and no slice gap. The structural images were acquired with a voxel resolution of 1×1×1 mm across 180 slices (FOV = 240×240 mm; matrix size = 240×240). The other parameters remained the same as those used at Beijing Normal University.

## Data Analysis

### Behavioral analysis

Trials with incorrect responses and RTs shorter than 100 ms or longer than 1,300 ms (5.52% of total trials) were excluded from statistical analyses. Mean RTs on correct trials and response accuracies were entered into the two-way rm-ANOVA.

### fMRI Data Analysis

The fMRI preprocessing and analysis were conducted with the BrainVoyager QX (version 2.2, Brain Innovation) software package (Goebel et al., 2006 [\[2\]](#)). The initial eight functional volumes of each scan were discarded to allow signal equilibration. For the remaining functional images, slice timing correction was applied using sinc interpolation, followed by 3D motion correction with trilinear/sinc interpolation for intra-session alignment to the middle volume. Each run for each participant was examined for the six head motion parameters (three rotations and three translations). Runs with motions exceeding one voxel length in any direction were excluded. An isotropic Gaussian kernel of an 8-mm full width at half maximum (FWHM) was then applied to spatially smooth the images. Finally, linear trend removal was performed, along with high-pass temporal filtering at a cutoff of approximately 0.0081 Hz (corresponding to seven cycles per run), to remove low-frequency nonlinear drifts. After these steps, we checked the variance in the global signals to assess the data quality, specifically for detection of any abrupt changes in the time course of each run. Volumes containing abrupt changes were interpolated, and runs exhibiting a global variance exceeding 0.1% were excluded. For each participant, the accepted fMRI data were co-registered with an anatomical scan in the native space and further transformed to the standard Montreal Neurological Institute (MNI) 152 space, with a resampled voxel size of 3×3×3 mm.

After the preprocessing, statistical analyses were performed using a random effects general linear model (RFX-GLM) analysis within BrainVoyager, executing a multi-subject GLM with distinct predictors for each participant. Using a deconvolution and multiple regression approach, we modeled six experimental conditions and one “error” term (including all the error trials) for each participant, with each condition including six sampling points taken from the 0-12 s period after the cue presentation (i.e., one sampling point every 2 seconds). The functional images occurring 5-10 s after the cue onsets, corresponding to the peak of the hemodynamic response function (HRF, Cohen, 1997), were used to provide parameter estimates for the amplitudes of the HRF. Volumes deviating in intensity by  $\pm 3$  SDs or more from the individual means were removed by a weighted vector that was included in the model as a covariate of no interest. In addition, the six mean-centered head motion parameters were modeled as covariates of no interest to further remove any residual variance due to head motion. To mitigate noise related to global physiological processes, the model incorporated the global signal, which represented the normalized average activity across all voxels at each time point in the standard space, as an additional predictor. We examined the three contrasts of interest introduced earlier. Corrections for multiple comparisons at  $p < 0.05$  were made through the Cluster Threshold plugin (BrainVoyager) using 2,500 Monte Carlo simulations. Minimum cluster sizes (540 mm<sup>3</sup> corresponding to 20 voxels) corresponding to significance at a threshold of  $p < 0.005$  (uncorrected) were computed for each contrast (Forman et al., 1995). The approximate Brodmann areas (BAs) and the corresponding anatomical labels of the peak voxel of the significant clusters in the MNI space were identified using the Neuroelf toolbox v1.1 (Weber, 2017 [\[2\]](#)).

## Data availability

The processed data used for the final analyses are available at <https://github.com/yangzhangpsy/ER-fMRI-IOR> [\[2\]](#). Raw data for this study can be requested from the Lead Contact, and the authors confirm that all reasonable requests will be fulfilled.

## Acknowledgements

This research was supported by grants from the Brain Science and Brain-like Intelligence Technology-National Science and Technology Major Project (2025ZD0215702), the National Natural Science Foundation of China (32171049), the Social Science Foundation of Jiangsu Province (22JYB015), the China Postdoctoral Science Foundation (2024M752310), and Jiangsu Funding Program for Excellent Postdoctoral Talent (2024ZB496, 2025ZB642).

## Additional files

[Supplemental Materials](#) [\[2\]](#)

## Additional information

### Funding

Funder	Grant reference number	Author
Brain Science and Brain-like Intelligence Technology - National Science and Technology Major Project	2025ZD0215702	Yang Zhang
MOST   National Natural Science Foundation of China (NSFC)	32171049	Yang Zhang
Social Science Foundation of Jiangsu Province (江苏省社会科学基金项目)	22JYB015	Yang Zhang

博士后科学基金)	2024M752310	Yujie Chen
Jiangsu Funding Program for Excellent Postdoctoral Talent	2024ZB496	Yujie Chen
Jiangsu Funding Program for Excellent Postdoctoral Talent	2025ZB642	Ai-Su Li

## Author ORCID iDs

Yang Zhang: <https://orcid.org/0000-0002-1827-9383>

## References

Ahrens M., Veniero D., Freund I. M., Harvey M., Thut G (2019) Both dorsal and ventral attention network nodes are implicated in exogenously driven visuospatial anticipation. *Cortex* **117**:168-181 <https://doi.org/10.1016/j.cortex.2019.02.031>

Albers C., Lakens D (2018) When power analyses based on pilot data are biased: Inaccurate effect size estimators and follow-up bias. *Journal Of Experimental Social Psychology* **74**:187-195 <https://doi.org/10.1016/j.jesp.2017.09.004>

Astafiev S. V., Shulman G. L., Stanley C. M., Snyder A. Z., Van Essen D. C., Corbetta M. (2003) Functional organization of human intraparietal and frontal cortex for attending, looking, and pointing. *Journal of Neuroscience* **23**:4689-4699 <https://doi.org/10.1523/jneurosci.23-11-04689.2003>

Baker D. H., Vilidaite G., Lygo F. A., Smith A. K., Flack T. R., Gouws A. D., Andrews T. J (2021) Power contours: Optimising sample size and precision in experimental psychology and human neuroscience. *Psychological Methods* **26**:295-314 <https://doi.org/10.1037/met0000337>

Botvinick M. M., Cohen J. D., Carter C. S (2004) Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences* **8**:539-546 <https://doi.org/10.1016/j.tics.2004.10.003>

Bourgeois A., Chica A. B., Valero-Cabré A., Bartolomeo P (2013a) Cortical control of inhibition of return: Causal evidence for task-dependent modulations by dorsal and ventral parietal regions. *Cortex* **49**:2229-2238 <https://doi.org/10.1016/j.cortex.2012.10.017>

Bourgeois A., Chica A. B., Valero-Cabré A., Bartolomeo P (2013b) Cortical control of Inhibition of Return: Exploring the causal contributions of the left parietal cortex. *Cortex* **49**:2927-2934 <https://doi.org/10.1016/j.cortex.2013.08.004>

Brainard D. H (1997) The Psychophysics Toolbox. *Spatial Vision* **10**:433-436 <https://doi.org/10.1163/156856897x00357>

Buracas G. T., Boynton G. M (2002) Efficient design of event-related fMRI experiments using M-sequences. *NeuroImage* **16**:801-813 <https://doi.org/10.1006/nimg.2002.1116>

Burgess N., Maguire E. A., O'Keefe J (2002) The human hippocampus and spatial and episodic memory. *Neuron* **35**:625-641 [https://doi.org/10.1016/S0896-6273\(02\)00830-9](https://doi.org/10.1016/S0896-6273(02)00830-9)

Campbell J. I. D., Thompson V. A (2012) MorePower 6.0 for ANOVA with relational confidence intervals and Bayesian analysis. *Behavior Research Methods* **44**:1255-1265 <https://doi.org/10.3758/s13428-012-0186-0>

Chen G., Pine D. S., Brotman M. A., Smith A. R., Cox R. W., Taylor P. A., Haller S. P (2022) Hyperbolic trade-off: The importance of balancing trial and subject sample sizes in neuroimaging. *NeuroImage* **247**:118786 <https://doi.org/10.1016/j.neuroimage.2021.118786>

Chen Q., Wei P., Zhou X (2006) Distinct neural correlates for resolving stroop conflict at inhibited and noninhibited locations in inhibition of return. *Journal Of Cognitive Neuroscience* **18**:1937-1946 <https://doi.org/10.1162/jocn.2006.18.11.1937>

Chen Q., Zhang M., Zhou X (2007) Interaction between location- and frequency-based inhibition of return in human auditory system. *Experimental Brain Research* **176**:630-640  
<https://doi.org/10.1007/s00221-006-0642-0>

Chica A. B., Bartolomeo P., Valero-Cabré A (2011) Dorsal and ventral parietal contributions to spatial orienting in the human brain. *The Journal of Neuroscience* **31**:8143-8149  
<https://doi.org/10.1523/JNEUROSCI.5463-10.2010>

Chung R. S., Cavalieri J., Sundaram S., Gilbert Z. D., Del Campo-Vera R. M., Leonor A., Tang A. M., Chen K.-H., Sebastian R., Shao A., et al. (2024) Understanding the human conflict processing network: A review of the literature on direct neural recordings during performance of a modified stroop task. *Neuroscience Research* **206**:1-19 <https://doi.org/10.1016/j.neures.2024.03.006>

Corbetta M., Shulman G. L (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience* **3**:201-215 <https://doi.org/10.1038/nrn755>

Cristescu T. C., Nobre A. C (2008) Differential modulation of word recognition by semantic and spatial orienting of attention. *Journal Of Cognitive Neuroscience* **20**:787-801  
<https://doi.org/10.1162/jocn.2008.20503>

Danieli K., Guyon A., Bethus I (2023) Episodic Memory formation: A review of complex Hippocampus input pathways. *Progress in Neuro-Psychopharmacology and Biological Psychiatry* **126**:110757  
<https://doi.org/10.1016/j.pnpbp.2023.110757>

De Houwer J. (2003) On the role of stimulus-response and stimulus-stimulus compatibility in the Stroop effect. *Memory & Cognition* **31**:353-359 <https://doi.org/10.3758/bf03194393>

Fox M. D., Corbetta M., Snyder A. Z., Vincent J. L., Raichle M. E (2006) Spontaneous neuronal activity distinguishes human dorsal and ventral attention systems. *Proceedings of the National Academy of Sciences of the United States of America* **103**:10046-10051 <https://doi.org/10.1073/pnas.0604187103>

Friston K. J., Zarahn E., Josephs O., Henson R. N. A., Dale A. M (1999) Stochastic designs in event-related fMRI. *NeuroImage* **10**:607-619 <https://doi.org/10.1006/nimg.1999.0498>

Fuentes L. J., Boucart M., Vivas A. B., Alvarez R., Zimmerman M. A (2000) Inhibitory tagging in inhibition of return is affected in schizophrenia: evidence from the stroop task. *Neuropsychology* **14**:134-140  
<https://doi.org/10.1037/0894-4105.14.1.134>

Fuentes L. J., Vivas A. B., Humphreys G. W (1999) Inhibitory tagging of stimulus properties in inhibition of return: Effects on semantic priming and flanker interference. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology* **52A**:149-164  
<https://doi.org/10.1080/027249899391269>

Funes M. J., Lupiáñez J., Milliken B (2008) The modulation of exogenous Spatial Cueing on Spatial Stroop interference: Evidence of a set for “cue-target event segregation”. *Psicológica* **29**:65-95

Goebel R., Esposito F., Formisano E (2006) Analysis of functional image analysis contest (FIAC) data with brainvoyager QX: From single-subject to cortically aligned group general linear model analysis and self-organizing group independent component analysis. *Human Brain Mapping* **27**:392-401  
<https://doi.org/10.1002/hbm.20249>

Hanlon F. M., Dodd A. B., Ling J. M., Bustillo J. R., Abbott C. C., Mayer A. R (2017) From behavioral facilitation to Inhibition: The neuronal correlates of the orienting and reorienting of auditory attention. *Frontiers in Human Neuroscience* **11**:293 <https://doi.org/10.3389/fnhum.2017.00293>

Hayes S. M., Nadel L., Ryan L (2007) The effect of scene context on episodic object recognition: parahippocampal cortex mediates memory encoding and retrieval success. *Hippocampus* **17**:873-889 <https://doi.org/10.1002/hipo.20319>

Holm S (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* **6**:65-70 <http://www.jstor.org/stable/4615733>

Hu F. K., Samuel A. G., Chan A. S (2011) Eliminating inhibition of return by changing salient nonspatial attributes in a complex environment. *The Journal of Experimental Psychology: General* **140**:35-50  
<https://doi.org/10.1037/a0021091>

Kahneman D., Treisman A., Gibbs B. J (1992) The reviewing of object files: Object-specific integration of information. *Cognitive Psychology* **24**:175-219 [https://doi.org/10.1016/0010-0285\(92\)90007-o](https://doi.org/10.1016/0010-0285(92)90007-o)

Klein R. M (2000) Inhibition of return. *Trends Cogn Sci* **4**:138-147 [https://doi.org/10.1016/s1364-6613\(00\)01452-2](https://doi.org/10.1016/s1364-6613(00)01452-2)

Klein R. M., Dick B (2002) Temporal dynamics of reflexive attention shifts: a dual-stream rapid serial visual presentation exploration. *Psychol Sci* **13**:176-179 <https://doi.org/10.1111/1467-9280.00432>

Klein R. M., Ivanoff J (2011) The components of visual attention and the ubiquitous Simon effect. *Acta Psychologica* **136**:225-234 <https://doi.org/10.1016/j.actpsy.2010.08.003>

Klein R. M., Redden R. S., Hilchey M. D (2023) Visual search and the inhibitions of return [Review]. *Frontiers in Cognition* **2** <https://doi.org/10.3389/fcogn.2023.1146511>

Lepsién J., Pollmann S (2002) Covert reorienting and inhibition of return: an event-related fMRI study. *Journal Of Cognitive Neuroscience* **14**:127-144 <https://doi.org/10.1162/089892902317236795>

Li A., Li Y., He X., Zhang Y (2023) Inhibition of return as a foraging facilitator in visual search: Evidence from long-term training. *Attention, Perception, & Psychophysics* **85**:88-98 <https://doi.org/10.3758/s13414-022-02605-0>

Li A., Miao C., Han Y., He X., Zhang Y (2018) Electrophysiological correlates of the effect of task difficulty on inhibition of return [Original Research]. *Frontiers in Psychology* **9** <https://doi.org/10.3389/fpsyg.2018.02403>

Li M., Lu S., Zhong N (2016) The parahippocampal cortex mediates contextual associative memory: evidence from an fMRI study. *BioMed Research International* **2016**:9860604 <https://doi.org/10.1155/2016/9860604>

Li Q., Yang G., Li Z., Qi Y., Cole M. W., Liu X (2017) Conflict detection and resolution rely on a combination of common and distinct cognitive control networks. *Neuroscience & Biobehavioral Reviews* **83**:123-131 <https://doi.org/10.1016/j.neubiorev.2017.09.032>

Liu J., Bayle D. J., Spagna A., Sitt J. D., Bourgeois A., Lehongre K., Fernandez-Vidal S., Adam C., Lambrecq V., Navarro V., et al. (2023) Fronto-parietal networks shape human conscious report through attention gain and reorienting. *Communications Biology* **6**:730 <https://doi.org/10.1038/s42003-023-05108-2>

Liu T. T (2004) Efficiency, power, and entropy in event-related fMRI with multiple trial types: Part II: design of experiments. *NeuroImage* **21**:401-413 <https://doi.org/10.1016/j.neuroimage.2003.09.031>

Liu T. T., Frank L. R (2004) Efficiency, power, and entropy in event-related FMRI with multiple trial types. Part I: theory. *NeuroImage* **21**:387-400 <https://doi.org/10.1016/j.neuroimage.2003.09.030>

Liu T. T., Frank L. R., Wong E. C., Buxton R. B (2001) Detection power, estimation efficiency, and predictability in event-related fMRI. *NeuroImage* **13**:759-773 <https://doi.org/10.1006/nimg.2000.0728>

Luo C., Lupiáñez J., Fu X., Weng X (2010) Spatial Stroop and spatial orienting: the role of onset versus offset cues. *Psychological Research* **74**:277-290 <https://doi.org/10.1007/s00426-009-0253-z>

Lupiáñez J. (2010) Inhibition of return. In: Nobre A. C., Coull J. T. (Eds). *Attention and Time* Oxford University Press. pp. 17-34 <https://doi.org/10.1093/acprof:oso/9780199563456.003.0002>

Lupiáñez J., Funes M. J (2005) Peripheral spatial cues modulate spatial congruency effects: Analysing the "locus" of the cueing modulation. *European Journal of Cognitive Psychology* **17**:727-752 <https://doi.org/10.1080/09541440540000103>

Lupiáñez J., Klein R. M., Bartolomeo P (2006) Inhibition of return: Twenty years after. *Cognitive Neuropsychology* **23**:1003-1014 <https://doi.org/10.1080/02643290600588095>

Lupiáñez J., Milliken B (1999) Inhibition of return and the attentional set for integrating versus differentiating information. *Journal Of General Psychology* **126**:392-418 <https://doi.org/10.1080/00221309909595373>

Lupiáñez J., Milliken B., Solano C., Weaver B., Tipper S. P (2001) On the strategic modulation of the time course of facilitation and inhibition of return. *Quarterly Journal Of Experimental Psychology* **54**:753-773 <https://doi.org/10.1080/713755990>

Lupiáñez J., Ruz M., Funes M. J., Milliken B (2007) The manifestation of attentional capture: Facilitation or IOR depending on task demands. *Psychological Research* **71**:77-91  
<https://doi.org/10.1007/s00426-005-0037-z>

Ma W. J., Navalpakkam V., Beck J. M., Berg R., Pouget A (2011) Behavior and neural basis of near-optimal visual search. *Nature Neuroscience* **14**:783-790 <https://doi.org/10.1038/nn.2814>

Martín-Arévalo E., Chica A. B., Lupiáñez J (2014) Electrophysiological modulations of exogenous attention by intervening events. *Brain and Cognition* **85**:239-250  
<https://doi.org/10.1016/j.bandc.2013.12.012>

Martín-Arévalo E., Chica A. B., Lupiáñez J (2016) No single electrophysiological marker for facilitation and inhibition of return: A review. *Behavioural Brain Research* **300**:1-10  
<https://doi.org/10.1016/j.bbr.2015.11.030>

Mayer A. R., Dorflinger J. M., Rao S. M., Seidenberg M (2004) Neural networks underlying endogenous and exogenous visual-spatial orienting. *NeuroImage* **23**:534-541  
<https://doi.org/10.1016/j.neuroimage.2004.06.027>

Mayer A. R., Franco A. R., Harrington D. L (2009) Neuronal modulation of auditory attention by informative and uninformative spatial cues. *Hum Brain Mapp* **30**:1652-1666  
<https://doi.org/10.1002/hbm.20631>

Mayer A. R., Harrington D. L., Stephen J., Adair J. C., Lee R. R (2007) An event-related fMRI Study of exogenous facilitation and inhibition of return in the auditory modality. *Journal Of Cognitive Neuroscience* **19**:455-467 <https://doi.org/10.1162/jocn.2007.19.3.455>

Mayer A. R., Seidenberg M., Dorflinger J. M., Rao S. M (2004) An event-related fMRI study of exogenous orienting: Supporting evidence for the cortical basis of inhibition of return?. *Journal Of Cognitive Neuroscience* **16**:1262-1271 <https://doi.org/10.1162/0898929041920531>

Mayer A. R., Teshiba T. M., Franco A. R., Ling J., Shane M. S., Stephen J. M., Jung R. E (2012) Modeling conflict and error in the medial frontal cortex. *Human Brain Mapping* **33**:2843-2855  
<https://doi.org/10.1002/hbm.21405>

Menon V., White C. D., Eliez S., Glover G. H., Reiss A. L (2000) Analysis of a distributed neural system involved in spatial information, novelty, and memory processing. *Hum Brain Mapp* **11**:117-129  
[https://doi.org/10.1002/1097-0193\(200010\)11:2<117::AID-HBM50>3.0.CO;2-M](https://doi.org/10.1002/1097-0193(200010)11:2<117::AID-HBM50>3.0.CO;2-M)

Milham M. P., Banich M. T., Webb A., Barad V., Cohen N. J., Wszalek T., Kramer A. F (2001) The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. *Cognitive Brain Research* **12**:467-473 [https://doi.org/10.1016/S0926-6410\(01\)00076-3](https://doi.org/10.1016/S0926-6410(01)00076-3)

Milliken B., Tipper S. P., Houghton G., Lupiáñez J (2000) Attending, ignoring, and repetition: On the relation between negative priming and inhibition of return. *Perception & Psychophysics* **62**:1280-1296 <https://doi.org/10.3758/BF03212130>

Müller N. G., Kleinschmidt A (2007) Temporal dynamics of the attentional spotlight: Neuronal correlates of attentional capture and inhibition of return in early visual cortex. *Journal Of Cognitive Neuroscience* **19**:587-593 <https://doi.org/10.1162/jocn.2007.19.4.587>

Nobre A. C., van Ede F. (2018) Anticipated moments: Temporal structure in attention. *Nature Reviews Neuroscience* **19**:34-48 <https://doi.org/10.1038/nrn.2017.141>

Posner M. I., Cohen Y. A (1984) Components of visual orienting. *Attention and Performance* **10**:531-556

Posner M. I., Rafal R. D., Choate L. S., Vaughan J (1985) Inhibition of return: Neural basis and function. *Cognitive Neuropsychology* **2**:211-228 <https://doi.org/10.1080/02643298508252866>

Prime D. J., Jolicoeur P (2009) On the relationship between occipital cortex activity and inhibition of return. *Psychophysiology* **46**:1278-1287 <https://doi.org/10.1111/j.1469-8986.2009.00858.x>

Ro T., Farnè A., Chang E (2003) Inhibition of return and the human frontal eye fields. *Experimental Brain Research* **150**:290-296 <https://doi.org/10.1007/s00221-003-1470-0>

Satel J., Wilson N. R., Klein R. M (2019) What neuroscientific studies tell us about inhibition of return. *Vision (Basel)* **3** <https://doi.org/10.3390/vision3040058>

Schmidt C. C., Timpert D. C., Arend I., Vossel S., Dovern A., Saliger J., Karbe H., Fink G. R., Henik A., Weiss P. H. (2018) Preserved but Less Efficient Control of Response Interference After Unilateral Lesions of the Striatum. *Frontiers in Human Neuroscience* **12**:414 <https://doi.org/10.3389/fnhum.2018.00414>

Schmidt C. C., Timpert D. C., Arend I., Vossel S., Fink G. R., Henik A., Weiss P. H. (2020) Control of response interference: Caudate nucleus contributes to selective inhibition. *Scientific Reports* **10**:20977 <https://doi.org/10.1038/s41598-020-77744-1>

Seidel Malkinson T., Bayle D. J., Kaufmann B. C., Liu J., Bourgeois A., Lehongre K., Fernandez-Vidal S., Navarro V., Lambrecq V., Adam C., et al. (2024) Intracortical recordings reveal vision-to-action cortical gradients driving human exogenous attention. *Nature Communications* **15**:2586 <https://doi.org/10.1038/s41467-024-46013-4>

Taylor T. L., Donnelly M. P (2002) Inhibition of return for target discriminations: the effect of repeating discriminated and irrelevant stimulus dimensions. *Perception & Psychophysics* **64**:292-317 <https://doi.org/10.3758/bf03195793>

Taylor T. L., Klein R. M (1998) On the causes and effects of inhibition of return. *Psychonomic Bulletin & Review* **5**:625-643 <https://doi.org/10.3758/BF03208839>

Torres-Morales C., Cansino S (2024) Brain representations of space and time in episodic memory: A systematic review and meta-analysis. *Cognitive, Affective, & Behavioral Neuroscience* **24**:1-18 <https://doi.org/10.3758/s13415-023-01140-1>

Veen V. V., Carter C. S (2005) Separating semantic conflict and response conflict in the Stroop task: A functional MRI study. *NeuroImage* **27**:497-504 <https://doi.org/10.1016/j.neuroimage.2005.04.042>

Veen V. V., Cohen J. D., Botvinick M. M., Stenger V. A., Carter C. S (2001) Anterior cingulate cortex, conflict monitoring, and levels of processing. *NeuroImage* **14**:1302-1308 <https://doi.org/10.1006/nimg.2001.0923>

Vivas A. B., Fuentes L. J (2001) Stroop interference is affected in inhibition of return. *Psychonomic Bulletin & Review* **8**:315-323 <https://doi.org/10.3758/bf03196167>

Vivas A. B., Fuentes L. J., Estevez A. F., Humphreys G. W (2007) Inhibitory tagging in inhibition of return: Evidence from flanker interference with multiple distractor features. *Psychonomic Bulletin & Review* **14**:320-326 <https://doi.org/10.3758/bf03194071>

Vossel S., Geng J. J., Fink G. R (2014) Dorsal and ventral attention systems: distinct neural circuits but collaborative roles. *Neuroscientist* **20**:150-159 <https://doi.org/10.1177/1073858413494269>

Wager T. D., Nichols T. E (2003) Optimization of experimental design in fMRI: a general framework using a genetic algorithm. *NeuroImage* **18**:293-309 [https://doi.org/10.1016/S1053-8119\(02\)00046-0](https://doi.org/10.1016/S1053-8119(02)00046-0)

Wang Z., Klein R. M (2010) Searching for inhibition of return in visual search: A review. *Vision Research* **50**:220-228 <https://doi.org/10.1016/j.visres.2009.11.013>

Weber J. (2017) NeuroElf. version: 1.1 <http://neuroelf.net>

Wilkinson D., Halligan P (2004) The relevance of behavioural measures for functional-imaging studies of cognition. *Nature Reviews Neuroscience* **5**:67-73 <https://doi.org/10.1038/nrn1302>

Wolfe J. M., Horowitz T. S (2017) Five factors that guide attention in visual search. *Nature Human Behaviour* **1**:0058 <https://doi.org/10.1038/s41562-017-0058>

Yang Z., Mayer A. R (2014) An event-related fMRI study of exogenous orienting across vision and audition. *Human Brain Mapping* **35**:964-974 <https://doi.org/10.1002/hbm.22227>

Zhang M., Zhang Y (2007) Semantic processing is affected in inhibition of return: evidence from an event-related potentials study. *Neuroreport* **18**:267-271 <https://doi.org/10.1097/WNR.0b013e32801231a9>

Zhang Y., Zhou X., Zhang M (2012) Temporary inhibitory tagging at previously attended locations: evidence from event-related potentials. *Psychophysiology* **49**:1191-1199 <https://doi.org/10.1111/j.1469-8986.2012.01412.x>

Zhou X., Chen Q (2008) Neural correlates of spatial and non-spatial inhibition of return (IOR) in attentional orienting. *Neuropsychologia* **46**:2766-2775  
<https://doi.org/10.1016/j.neuropsychologia.2008.05.017>

Zu G., Zhang T., Yang J., Wang A., Zhang M (2023) Does the construction retrieval account apply to cross-modal inhibition of return in semantic context?. *PsyCh Journal* **12**:211-221  
<https://doi.org/10.1002/pchj.618>

## Peer reviews

### Reviewer #1 (Public review):

#### Summary:

This study makes a significant and timely contribution to the field of attention research. By providing the first direct neuroimaging evidence for the integration-segregation theory of exogenous attention, it fills a critical gap in our understanding of the neural mechanisms underlying inhibition of return (IOR). The authors employ a carefully optimized cue-target paradigm combined with fMRI to elegantly dissociate the neural substrates of cue-target integration from those of segregation, thereby offering compelling support for the integration-segregation account. Beyond validating a key theoretical hypothesis, the study also uncovers an interaction between spatial orienting and cognitive conflict processing, suggesting that exogenous attention modulates conflict processing at both semantic and response levels. This finding shed new light on the neural mechanisms that connect exogenous attentional orienting with cognitive control.

#### Strengths:

The experimental design is rigorous, the analyses are thorough, and the interpretation is well grounded in the literature. The manuscript is clearly written, logically structured, and addresses a theoretically important question. Overall, this is an excellent, high-impact study that advances both theoretical and neural models of attention.

#### Weaknesses:

While this study addresses an important theoretical question and presents compelling neuroimaging findings, a few additional details would help improve clarity and interpretation. Specifically, more information could be provided regarding the experimental conditions (SI and RI), the justification for the criteria used for excluding behavioral trials, and how the null condition was incorporated into the analyses. In addition, given the non-significant interaction effect in the behavioral results, the claim that the behavioral data "clearly isolated" distinct semantic and response conflict effects should be phrased more cautiously.

<https://doi.org/10.7554/eLife.109842.1.sa2>

### Reviewer #2 (Public review):

#### Summary:

This study provides evidence for the integration-segregation theory of an attentional effect, widely cited as inhibition of return (IOR), from a neuroimaging perspective, and explores neural interactions between IOR and cognitive conflict, showing that conflict processing is potentially modulated by attentional orienting.

#### Strengths:

The integration-segregation theory was examined in a sophisticated experimental task that also accounted for cognitive conflict processing, which is phenomenologically related to IOR but "non-spatial" by nature. This study was carefully designed and executed. The behavioral and neuroimaging data were carefully analyzed and largely well presented.

#### Weaknesses:

The rationale for the experimental design was not clearly explained in the manuscript; more specifically, why the current ER-fMRI study would disentangle integration and segregation processes was not explained. The introduction of "cognitive conflict" into the present study was not well reasoned for a non-expert reader to follow.

The presentation of the results can be further improved, especially the neuroimaging results. For instance, Figure 4 is challenging to interpret. If "deactivation" (or a reduction in activation) is regarded as a neural signature of IOR, this should be clearly stated in the manuscript.

<https://doi.org/10.7554/eLife.109842.1.sa1>

### Reviewer #3 (Public review):

#### Summary:

This study aims to provide the first direct neuroimaging evidence relevant to the integration-segregation theory of exogenous attention - a framework that has shaped behavioral research for more than two decades but has lacked clear neural validation. By combining an inhibition-of-return (IOR) paradigm with a modified Stroop task in an optimized event-related fMRI design, the authors examine how attentional integration and segregation processes are implemented at the neural level and how these processes interact with semantic and response conflicts. The central goal is to map the distinct neural substrates associated with integration and segregation and to clarify how IOR influences conflict processing in the brain.

#### Strengths:

The study is well-motivated, addressing a theoretically important gap in the attention literature by directly testing a long-standing behavioral framework with neuroimaging methods. The experimental approach is creative: integrating IOR with a Stroop manipulation expands the theoretical relevance of the paradigm, and the use of a genetic-algorithm-optimized fMRI design ensures high efficiency. Methodologically, the study is sound, with rigorous preprocessing, appropriate modeling, and analyses that converge across multiple contrasts. The results are theoretically coherent, demonstrating plausible dissociations between integration-related activity in the fronto-parietal attention network (FEF, IPS, TPJ, dACC) and segregation-related activity in medial temporal regions (PHG, STG). The findings advance the field by supplying much-needed neural evidence for the integration-segregation framework and by clarifying how IOR modulates conflict processing.

#### Weaknesses:

Some interpretive aspects would benefit from clarification, particularly regarding the dual roles ascribed to dACC activation and the circumstances under which PHG and STG are treated as a single versus separate functional clusters. Reporting conventions are occasionally inconsistent (e.g., statistical formatting, abbreviation definitions), which may hinder readability. More detailed reporting of sample characteristics, exclusion criteria, and data-quality metrics-especially regarding the global-variance threshold-would improve transparency and reproducibility. Finally, some limitations of the study, including potential

constraints on generalization, are not explicitly acknowledged and should be articulated to provide a more balanced interpretation.

<https://doi.org/10.7554/eLife.109842.1.sa0>