Reviewed Preprint
v1 • January 12, 2026
Not revised

Reviewed Preprint
v2 • April 17, 2026
Revised by authors

Dissociable neural substrates of integration and segregation in exogenous attention

Yu-Jie Chen[1,†], Ai-Su Li[1,†], Yang Yu[2], Su Hu[2], Xun He[3] ✉, Yang Zhang[1] ✉

[1]Department of Psychology, Soochow University, Suzhou, China • [2]Department of Radiology, The First Affiliated Hospital of Soochow University, Suzhou, China • [3]School of Psychology, Bournemouth University, Poole, United Kingdom

✉ **For correspondence:**
xhe@bournemouth.ac.uk
yzhangpsy@suda.edu.cn

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

Competing interests: No competing interests declared

Funding: See page 21

Reviewing editor: Xilin Zhang, Key Laboratory of Brain, Cognition and Education Sciences, Ministry of Education, South China Normal University, China

eLife Assessment

This **important** study uses an optimized IOR-Stroop fMRI paradigm to dissociate integration and segregation processes and to show that attentional orienting modulates conflict processing at both the semantic and response levels. The evidence is **compelling**, supporting the integration-segregation theory of exogenous attention in inhibition of return while also deepening our understanding of how attentional orienting shapes downstream cognitive processing. The work will therefore be of broad interest to researchers in attention and cognitive control.

https://doi.org/10.7554/eLife.109842.2.sa4

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

Salient visual stimuli, such as abrupt onsets, involuntarily capture attention. This process, known as exogenous or reflexive attentional orienting, is crucial for efficient visual search in cluttered scenes (Klein et al., 2023; Li et al., 2023; Ma et al., 2011; Wang & Klein, 2010; Wolfe & Horowitz, 2017). Studies employing the cue-target paradigm (Posner & Cohen, 1984) have shown that the exogenous attentional orienting has 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; Lupianez et al., 2006; Posner et al., 1985; Seidel Malkinson et al., 2024). This characteristic shift from facilitation to inhibition has fueled decades of theoretical debate, giving rise to multiple competing accounts of its underlying mechanisms (Funes et al., 2008; Klein & Dick, 2002; Lupiáñez, 2010; Lupiáñez et al., 2001; Milliken et al., 2000; Prime & Jolicoeur, 2009; Taylor & Klein, 1998; Vivas et al., 2007).

The Integration-Segregation Theory

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; Lupiáñez et al., 2001; Milliken et al., 2000). Rooted in the object file framework (Kahneman et al., 1992), this theory attributes the biphasic pattern to the dynamic competition between cue-target integration and segregation. 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), whereas the segregation process tends to create a new episodic representation for targets at the uncued locations (Funes et al., 2008; Lupiáñez & Funes, 2005; Lupiáñez et al., 2001; Milliken et al., 2000). At short SOAs (e.g., around less than 250 ms), integration on the cued trials is more efficient than segregation on 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 cue-activated object file 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; Funes et al., 2008; Hu et al., 2011; Li et al., 2018; Luo et al., 2010; Lupiáñez et al., 2001; Lupiáñez et al., 2007; Zu et al., 2023). Originally developed to explain spatial IOR (Funes et al., 2008; Lupiáñez & Funes, 2005; Lupiáñez et al., 2001; Milliken et al., 2000), it has subsequently been extended to nonspatial forms of IOR (e.g., color-, shape-, or frequency-based IOR) in both visual and auditory modalities (Chen et al., 2007; Hu et al., 2011). When targets and cues 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; Hu et al., 2011).

More recently, such integrative interference has also been observed in cross-modal IOR, revealing supramodal mechanisms involving abstract semantic features (Zu et al., 2023). Researchers have further extended the explanatory scope of the integration-segregation framework to electrophysiological data (Li et al., 2018; Martín-Arévalo et al., 2014, 2016). 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), 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, 2016).

Furthermore, the theory also succeeds in accounting for task-dependent variations in IOR that are difficult to explain by other attentional theories (Chen et al., 2007; Lupiáñez & Milliken, 1999; Lupiáñez et al., 2001; Lupiáñez et al., 2007). By positing that task demands modulate the timing of object-file closure, the framework predicts an earlier IOR onset in detection tasks (favoring an early file closure for new event encoding) than that in discrimination tasks (i.e., late closure due to information accumulation) (Lupiáñez & Milliken, 1999; Lupiáñez et al., 2001). Crucially, Lupiáñez et al. (2007) found that, at long SOAs, frequent targets elicited the expected IOR, whereas infrequent targets instead received facilitation. This dissociation cannot be explained by cue-driven attentional capture and disengagement accounts alone, and instead suggests a task-dependent cue-target integration (Lupiáñez, 2010). Similarly, in auditory attention, Chen et al. (2007) found that task-irrelevant features can either enhance or eliminate the IOR effect depending on whether the cue and target share the same task-relevant dimension, a pattern considered to be 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 unified theoretical framework of exogenous attention accommodating diverse stimulus features, modalities, and task demands.

The Challenge of Neural Verification

Despite the strength, current support for the integration-segregation theory remains largely inferential, due to the fact that the recently hypothesized dual processes of integration and segregation had not been directly evidenced in brain activities. Operating under the assumption that IOR reflects a time-dependent inhibitory state that builds up with the increase of SOA (Lepsien & Pollmann, 2002; Mayer, Seidenberg, et al., 2004), the past studies typically contrasted long versus short SOAs to capture the neural dynamics underlying the inhibitory phase of visual attentional orienting. This contrast was typically examined either by merging the cued and uncued trials (Lepsien & Pollmann, 2002; Mayer, Dorflinger, et al., 2004; Müller & Kleinschmidt, 2007; Zhou & Chen, 2008), or separately for each of the cued and uncued conditions without directly comparing these cueing conditions (Mayer, Seidenberg, et al., 2004). 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, these SOA-based contrasts were insufficient for testing the integration-segregation framework, as they only captured the temporal dynamics in attentional orienting and in the process missed the functional distinction between integration and segregation that characterizes the theory.

To directly test the functional distinction of event integration (for the cued targets) vs. segregation (for the uncued targets), it is necessary to compare the cued and uncued conditions. Some studies attempted this direct comparison have yielded mixed findings. While Chen et al. (2006) identified a cue-validity effect during the inhibitory period, this effect was confined to the left FEF, and other studies did not observe significant neural difference between the cued and uncued trials during the inhibitory period (Lepsien & Pollmann, 2002; Mayer, Seidenberg, et al., 2004).Given that the integration-segregation account predicts distinct neural processing for cued and uncued targets, clear evidence for such a dissociation during inhibition remains limited. This limitation may reflect 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; Liu, 2004; Liu & Frank, 2004; Liu et al., 2001; Wager & Nichols, 2003).

Overview of the Present Study

To obtain direct neuroimaging evidence for the integration-segregation theory, the present study employed ER-fMRI with a stimulus sequence optimized with a genetic algorithm (GA) (Wager & Nichols, 2003). This flexible optimization approach was adopted to maximize statistical power of contrast detection while maintaining a high estimation efficiency of the hemodynamic response function (HRF), thereby addressing the power limitations that had negatively impacted on previous neuroimaging IOR studies. Guided by the integration-segregation framework, we predicted dissociable neural signatures for the cued versus uncued targets corresponding to their divergent processing requirements. Specifically, targets at the cued locations would engage an integration process to update the existing file, recruiting regions associated with information integration and attentional reorienting such as the FEF (Astafiev et al., 2003; Corbetta & Shulman, 2002; Liu et al., 2023). In contrast, targets appearing at the uncued locations would engage a segregation process to establish a new object file, presumably recruiting regions involved in new episodic encoding such as the parahippocampal gyrus (PHG) (Burgess et al., 2002; Danieli et al., 2023; Hayes et al., 2007; Li et al., 2016; Menon et al., 2000; Torres-Morales & Cansino, 2024).

Beyond the primary focus on the neural mechanisms of the integration-segregation framework (i.e., the IOR-generation mechanisms), the current study also employed the experimental design to examine how IOR modulates ongoing cognitive processing (i.e., the IOR expression mechanisms). Specifically, we embedded a modified Stroop task within the cue-target paradigm to systematically manipulate cognitive conflict at the target-processing stage (De Houwer, 2003; Veen & Carter, 2005; Veen et al., 2001). These conflict types were operationalized through three distinct stimulus types: 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 manipulation allowed us to examine how spatial attention interacts with distinct levels of cognitive conflict. A previous fMRI study by Chen et al. (2006) reported dissociable neural signatures for the semantic and response conflicts when spatial attention was engaged. However, that study manipulated response eligibility by excluding certain incongruent color words from the response set (Milham et al., 2001). This design choice has been criticized for conflating the semantic and response conflicts, as ineligible distractors may not be processed in the same way as response-relevant words (Veen & Carter, 2005). To address this limitation, the present study adopted a refined Stroop design that clearly separates the semantic and response conflicts while keeping all stimuli response-relevant (De Houwer, 2003; Veen & Carter, 2005; Veen et al., 2001), providing a more precise test of how spatial attention modulates these two distinct conflict types in brain activities.

Following a further 450-ms ISI, the target, a colored Chinese character, appeared at one of the two target locations with equal probabilities and remained on the screen for 450 ms. The trial ended with a variable intertrial interval (ITI) of 850, 1050, 1250, or 1450 ms (with equal probabilities). **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.

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) (Fig 1A), 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). 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; Veen & Carter, 2005). 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), 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.

Behavioral Results

Mean reaction times (RTs) and accuracies are shown in Fig 2. 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), $F(1, 28) = 12.057$, $p = .002$, $\eta_p^2 = .301$, showing slower responses to targets at the cued location ($M = 642$ ms, $SE = 20$ ms) than at the uncued location (M = 632 ms, SE = 19 ms). The main effect of congruency was also significant, $F(1.53, 42.88) = 29.602$, $p < .001$, $\eta_p^2 = .514$ (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 (M = 624 ms, SE = 19 ms), followed by SI (M = 632 ms, SE = 20 ms) then RI (M = 654 ms, SE = 21ms). These data demonstrated typical Stroop interference effects (Veen & Carter, 2005) in both the semantic (SI-NE difference) and response conflicts (RI-SI difference). The interaction between cue validity and congruency did not approach significance, $F(2, 56) = 0.930$, $p = .401$, $\eta_p^2 = .032$. To further investigate whether cue validity modulated the two conflict components, we conducted planned analyses examining its interactions with semantic conflict (SI vs. NE) and response conflict (RI vs. SI) (Chen et al., 2006). The results showed that cue validity did not significantly interact with either semantic conflict ($F(1, 28) = 0.968$, $p = .334$, $\eta_p^2 = .033$) or response conflict ($F(1, 28) = 1.502$, $p = .231$, $\eta_p^2 = .051$).

The rm-ANOVA for the accuracy data (Fig 2B) only showed a significant main effect of Congruency, $F(2, 56) = 7.685$, $p = .001$, $\eta_p^2 = .215$. Post hoc comparisons confirmed that this came from a lower accuracy in the RI condition ($M = 0.943$, $SE = 0.009$) than in the NE (M = 0.966, SE = 0.009; $t(28) = -3.596$, $p = .002$, Cohen’s $d = 0.446$) and SI (M = 0.963, SE = 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 ($F(1, 28) = 0.021$, $p = .887$, $\eta_p^2 < .001$) and the interaction ($F(2, 56) = 1.298$, $p = .281$, $\eta_p^2 = .044$) were not significant. Given that accuracy did not decrease with faster response times, 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. Notably, in the left hemisphere, these activations formed a continuous cluster spanning both regions (labeled as PHG/STG in Table 1). 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 $ps < .001$, 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 behavioral results did not reveal any significant modulation of IOR in the magnitude of either semantic conflict or response conflict, differential neural modulations were observed between these conditions (summarized in Fig 4 and Table 2). The effect of IOR on the

Fig 1. **Experimental Materials.**

A. Trial sequence and display sizes. Each trial started with a 150-ms non-informative cue presented at one of the two peripheral boxes. After a 150-ms interstimulus interval (ISI), a 150-ms fixation cue was presented at the central fixation box.

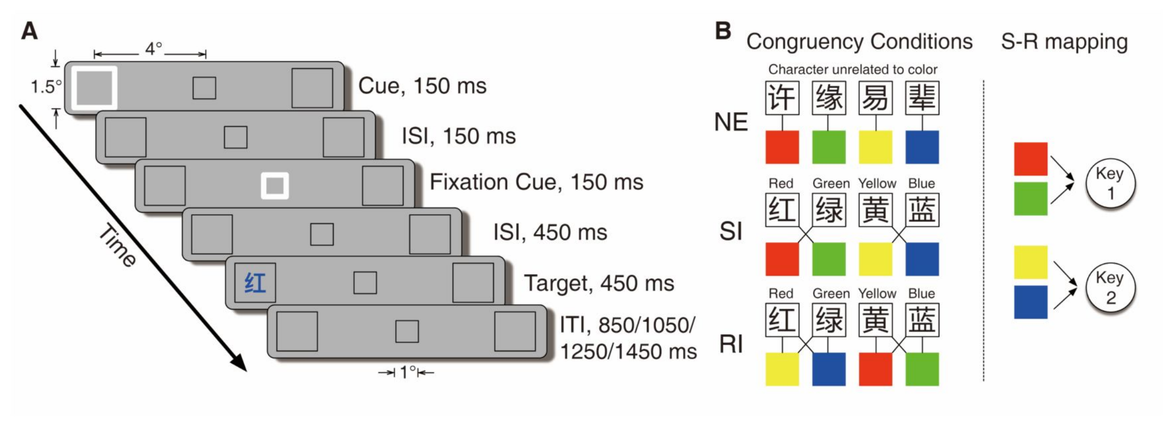
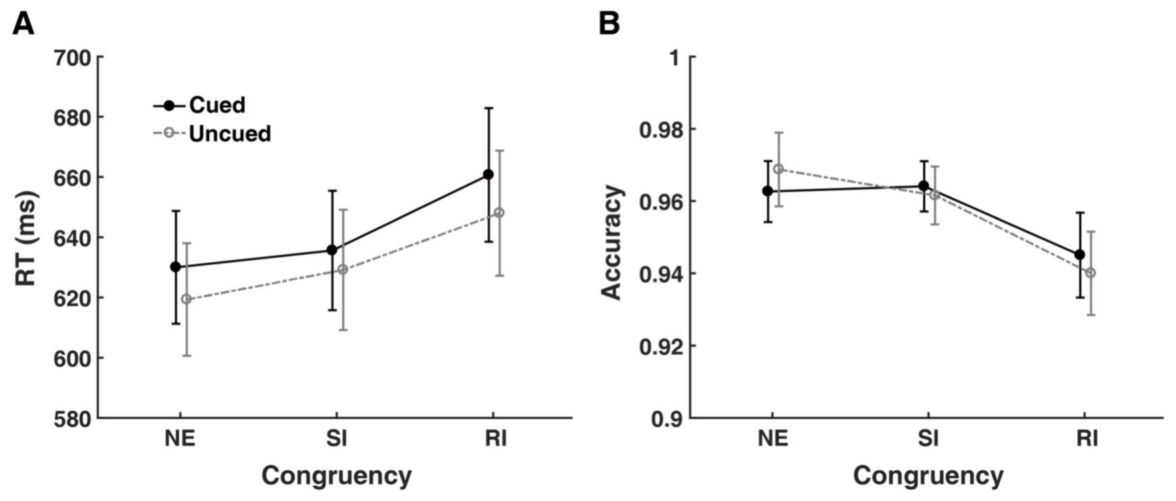


Fig 2. **Behavioral Results.**

Mean reaction times (**A**) and accuracies (**B**) as a function of cue validity and congruency. NE = neutral; SI = semantically incongruent; RI = response-incongruent. Error bars extend to one standard error of the mean (SEM).



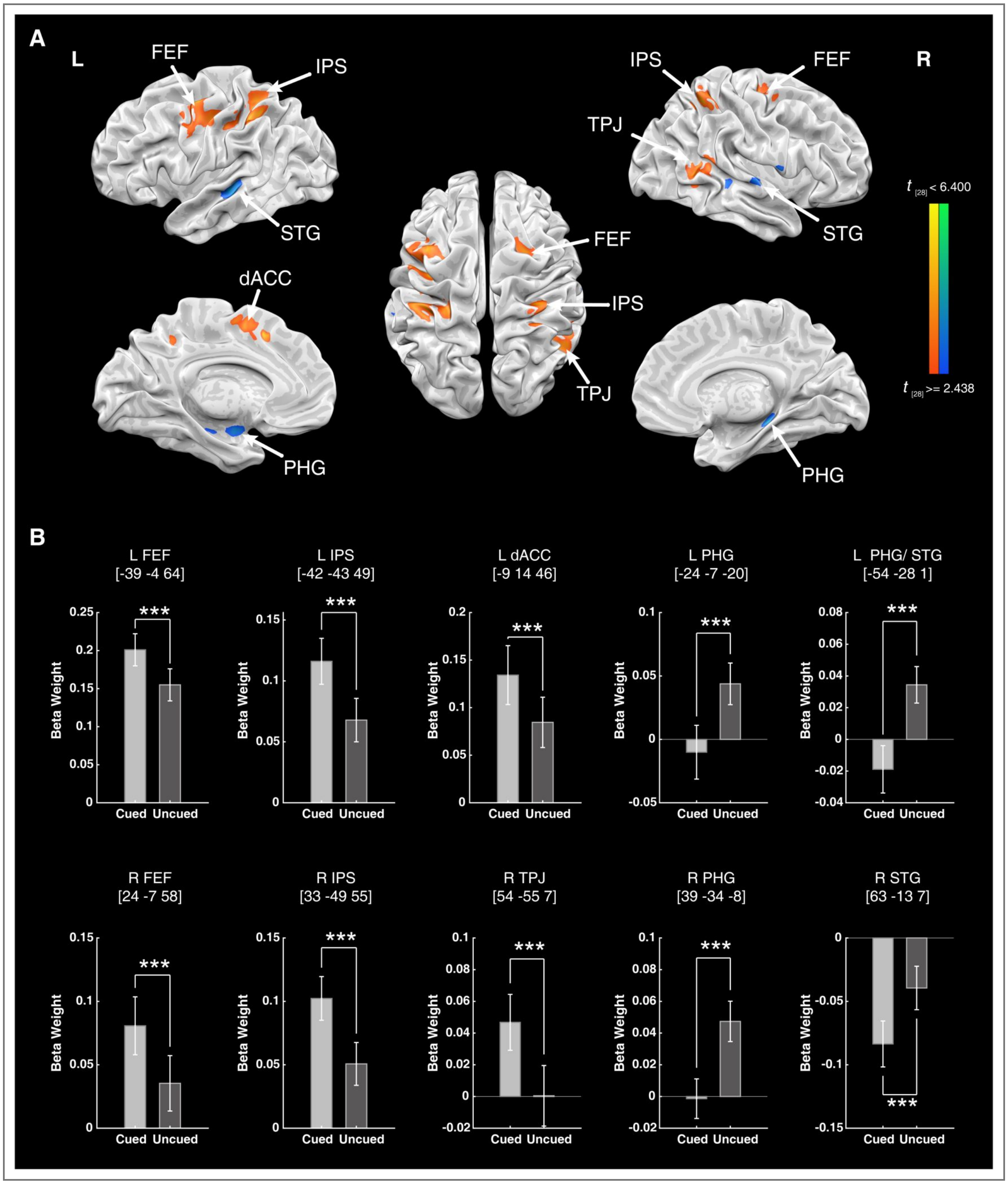


Fig 3. **IOR Effect in the Neutral Condition and Parameter Estimation.**

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 mm³ (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 coordinates peak			T_{max}	T_{mean}	BA
			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. "/" indicates that no BA could be assigned.

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

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, $F(1,28) = 15.946$, $p < .001$, $\eta_p^2 = .363$. 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 $d = 0.187$).

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 and Table 2). The right superior parietal cortex (SPC) showed a significant activation reduction (Fig 4B, left), while the right putamen exhibited an activation enhancement (Fig 4B, right). A two-way rm-ANOVA on the beta weights revealed a significant interaction in the right SPC, $F(1,28) = 20.833$, $p < .001$, $\eta_p^2 = .427$. 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 ($F(1,28) = 26.686$, $p < .001$, $\eta_p^2 = .488$), 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

Neural Substrates of Integration and Segregation

The integration-segregation theory has emerged as an influential framework for explaining the dynamic effects of exogenous attention (Chen et al., 2007; Funes et al., 2008; Hu et al., 2011; Li et al., 2018; Lupiáñez et al., 2001; Lupiáñez et al., 2007; Zu et al., 2023), 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; Lupiáñez & Funes, 2005; Lupiáñez et al., 2001; Milliken et al., 2000). In the current study, by contrasting the cued versus uncued targets under long SOA, we provided the first direct neuroimaging evidence supporting this theory by dissociating brain activation patterns associated with these two processes. Specifically, the cued and uncued targets engaged distinct neural systems that respectively map onto the functional demands of re-engaging an existing representation and those of encoding a novel spatial event.

The heightened activation of the bilateral FEF and IPS, and right TPJ for targets appearing at the cued locations reflected the increased attentional demand associated with the integration process. Within the integration-segregation theoretical framework, this demand is conceptualized as the need to maintain or "re-open" an object file. Specifically, this framework posits that in the long SOA conditions, the cue-initiated object file is likely to have closed or begun closing, hindering immediate integration of the subsequent targets (Funes et al., 2008; Lupiáñez & Funes, 2005; Lupiáñez et al., 2001; Milliken et al., 2000). Thus, 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; Corbetta & Shulman, 2002; Fox et al., 2006; Vossel et al., 2014). These regions act in concert to support the attentional shifts and reorienting necessary for reopening the object file for integration. Furthermore, the observed increase in the left dACC activity under the cued relative to the uncued condition likely reflected the engagement of cognitive control mechanisms (Botvinick et al., 2004; Chung et al., 2024; Mayer et al., 2012; Veen & Carter, 2005), particularly in resolving the conflict between the task-driven requirement of target integration and the reduced accessibility of the cue-initiated representation.

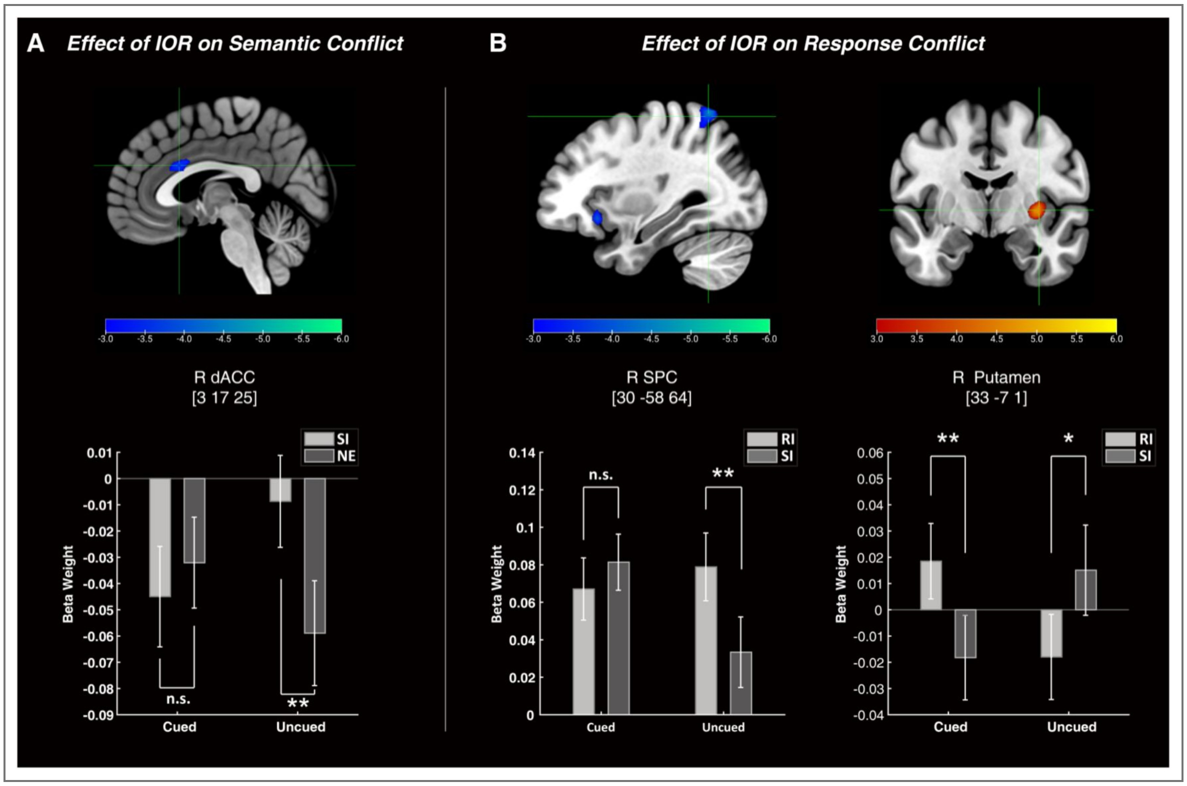
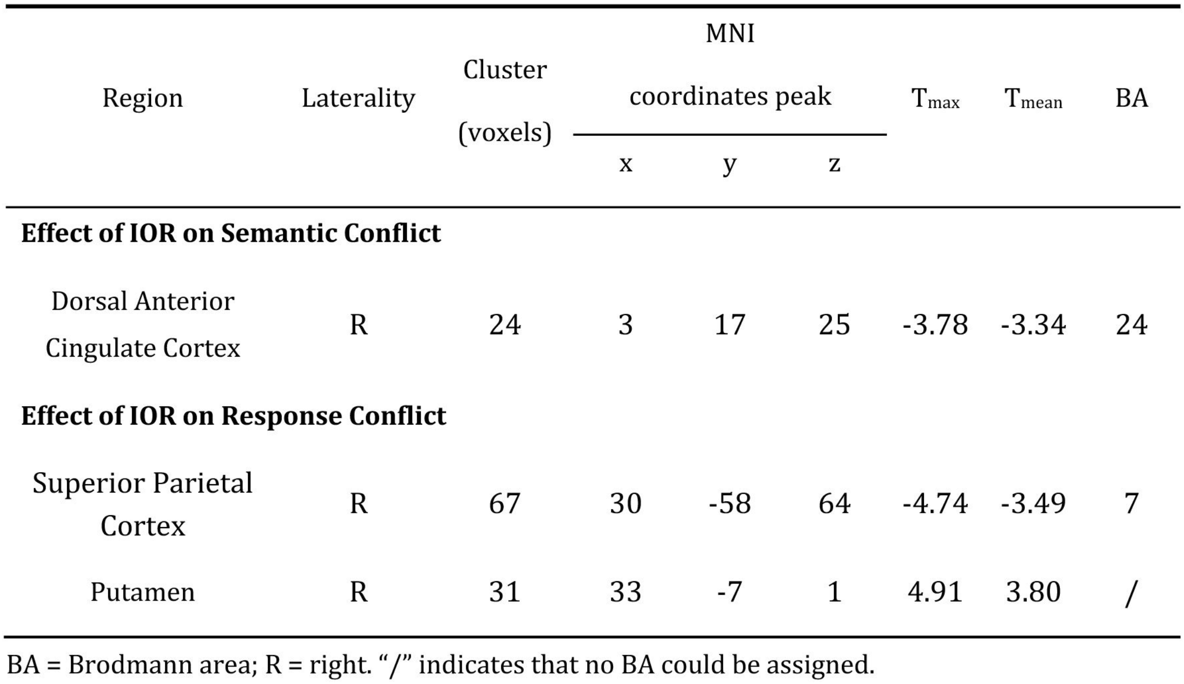


Fig 4. **Effect of IOR in semantic conflict and response conflict.**

(A) Regions showing the IOR modulation of semantic conflict, defined as (cued-SI – cued-NE) > (uncued-SI – uncued NE). **(B)** Regions showing the IOR modulation of response conflict, defined as (cued-RI – cued-SI) > (uncued-RI – uncued-SI). NE = neutral; SI = semantically incongruent; RI = response-incongruent. Parameter estimations were based on a threshold of $p <$ .005 (uncorrected), with a minimum cluster size of 540 mm³ (20 voxels), yielding a threshold of corrected $p < .05$ based on 2500 Monte Carlo simulations in BrainVoyager. Error bars extend to 1 SEM. $^{**}p < .01$, $^{*}p < .05$, n.s. = non-significant.



Region	Laterality	Cluster (voxels)	MNI coordinates peak			T_{max}	T_{mean}	BA
			x	y	z			
Effect of IOR on Semantic Conflict								
Dorsal Anterior Cingulate Cortex	R	24	3	17	25	-3.78	-3.34	24
Effect of IOR on Response Conflict								
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. “/” indicates that no BA could be assigned.

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).

In this context, the heightened activation of dACC may also reflect its role in fulfilling the inhibitory bias toward the cued location (Mayer, Seidenberg, et al., 2004) and discouraging inefficient integration attempts at a location marked as less relevant.

On the other hand, the recruitment of the bilateral PHG and the STG in the uncued condition supports the conceptualization of segregation as an active process of creating a new object file. According to the integration-segregation theory, when a target appears at an unced location, the brain will register it as a new and separate event. Our results suggest that the brain recruits mechanisms specialized for spatial novelty to support this segregation process. Specifically, PHG is involved in episodic encoding of novel visual or spatial stimuli (Burgess et al., 2002; Danieli et al., 2023; Hayes et al., 2007; Li et al., 2016; Menon et al., 2000; Ranganath & Rainer, 2003; Torres-Morales & Cansino, 2024), while the STG supports the detection of salient and unexpected changes in the sensory environment (Corbetta & Shulman, 2002; Downar et al., 2002). Thus, the enhanced PHG and STG activities observed in the uncued condition may signify the active engagement of a novelty-detection system required for encoding new spatial representations and establishing new object files.

Connections and Discrepancies with Previous IOR Neuroimaging Studies

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; Mayer, Dorflinger, et al., 2004; Mayer, Seidenberg, et al., 2004; Müller & Kleinschmidt, 2007) or a limited number of trials (e.g., 30 trials per condition; Lepsien & Pollmann, 2002), 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, effectively enhancing the sensitivity of detecting differences between experimental conditions (Baker et al., 2021; Chen et al., 2022). Finally, task design differences may further contribute to the observed inconsistencies. The earlier studies often employed simple localization or detection tasks (Lepsien & Pollmann, 2002; Mayer, Dorflinger, et al., 2004; Mayer, Seidenberg, et al., 2004; Müller & Kleinschmidt, 2007), while the current study adopted a discrimination task. According to the integration-segregation theory (Funes et al., 2008; Lupiáñez & Funes, 2005; Lupiáñez et al., 2001; Lupiáñez et al., 2007; Milliken et al., 2000), 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.

Neural Interactions Between IOR and Cognitive Conflict Processing

Another novelty of the current study is integrating the IOR and the modified Stroop tasks, which were separately studied for the semantic- and response-related conflicts (De Houwer, 2003; Veen & Carter, 2005). 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; Vivas & Fuentes, 2001) 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; Milham et al., 2001; Veen & Carter, 2005), appeared to serve as a critical neural interface for the interaction between semantic conflict and IOR. Specifically, in the uncued condition, the semantic incongruency 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), 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), which posits that, when attention is drawn away from a cued location, stimuli presented there are temporarily tagged with inhibition (Fuentes et al., 2000; Fuentes et al., 1999; Vivas & Fuentes, 2001). ERP evidence supporting this mechanism was reported by Zhang et al. (2012), 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), 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). This highlights that spatial attention can affect subsequent cognitive processes at the semantic level (Cristescu & Nobre, 2008; Zhang & Zhang, 2007).

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 the control of response interference (Schmidt et al., 2018; Schmidt et al., 2020). For example, Schmidt et al. (2020) 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). 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.

Methodological Considerations and Limitations

While the neural interactions between IOR and conflict processing offer novel insights, we need to be cautious when interpreting the results given that the neural interactions were observed without any corresponding behavioral effect. One likely explanation for this dissociation is the differences in measurement sensitivity between the behavioral and neural indices (Chen et al., 2006; Wilkinson & Halligan, 2004). As noted by Wilkinson and Halligan (2004), RTs and accuracies are not perfect measures of cognition, whereas neural signals can reveal finer-grained or “hidden” processes that precede overt behavior. Consistent with this view, Chen et al. (2006) reported a similar dissociation in which response conflict modulation by IOR was clearly seen in neural data but not in behavior. This suggests that neural modulations sometimes emerge even in the absence of detectable behavioral differences. In addition, the usage of a GA-optimized sequence in the current study may have partly accounted for the observed dissociation. While this optimization enhances both detection and HRF estimation efficiency, it may result in partially clustered event sequence that resemble a block-like structure, thereby reducing event counterbalancing and increasing sequential predictability (Wager & Nichols, 2003). As a result, the participants may have formed expectations about upcoming events and weakened the correspondence between the neural and behavioral findings. Future studies are required to address this limitation by employing more optimized designs that consider some psychological factors (e.g., event counterbalancing; Wager & Nichols, 2003) to better validate the observed neural mechanisms.

Conclusion

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; Lupiáñez & Funes, 2005; Lupiáñez et al., 2001; Milliken et al., 2000). 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. All participants were right-handed and reported no history of neurological or psychiatric 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; aged 18-30 years, *M* = 22.69, SD = 2.58). 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) for a within-subjects rm-ANOVA. To achieve an 80% statistical power at the threshold of α = .05 (Chen et al., 2006), 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). 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 ≥ 28).

Experimental Design

The experiment adopted a within-subjects design with two factors, namely cue validity (cued and uncued) and congruency (SI, RI, and NE). The targets appeared at the cued location on the cued trials and at the other peripheral location on the uncued trials. The congruency factor referred to the relationship between the ink color and the meaning of the Chinese characters (i.e., targets) according to the predefined stimulus-response mapping. In total, eight characters and four colors were used (see Fig 1B). Ink colors of the stimuli were mapped onto two response keys, with red and green assigned to one response key (Key 1) while yellow and blue assigned to the other response key (Key 2). In an SI trial, the ink color and the character meaning were incongruent but mapped to the same response key (e.g., character “红” [meaning “red”] being displayed in green ink, with both colors being mapped to the same response key). Thus, the SI trials involved a semantic conflict without inducing a response conflict. In an RI trial, the ink color and the character meaning differed and were also mapped to different response keys (e.g., “红” [red] being displayed in yellow ink, with red and yellow being related to different responses), leading to both the semantic and response conflicts. The NE trials used characters that were not related to any color in their meanings and shared the same orthographic structures (character complexity and form) as the color characters. In addition to the six experimental conditions, a null condition (no Chinese character presented) was included as an implicit baseline to facilitate estimation of the effects of interest in the ER-fMRI analysis (Burock et al., 1998; Friston et al., 1999; Liu, 2004).

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; Wager & Nichols, 2003). To ensure high design efficiency, we optimized the stimulus sequences employing the genetic algorithm (see Supplementary Information for details) (Wager & Nichols, 2003). This optimization improves the detection efficiency for the contrasts of interest by moderately sacrificing the efficiency of less relevant contrasts (Wager & Nichols, 2003). 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) 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°×1°, line width of 0.02°) and two black placeholder boxes (1.5°×1.5°, 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. 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°×1.4°) 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). The color category-button mapping was counterbalanced across participants. Furthermore, to avoid a possible occurrence of the Simon effect (Klein & Ivanoff, 2011), 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, with equal numbers of participants scanned at each site ($N = 16$ per site). 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; see Zhang et al. (2018) for spatiotemporal properties) viewed through a head-coil-mounted mirror at an optical distance 115 cm. 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). Responses were collected using an MRI-compatible 2-button fiber-optic response pad (Tang et al., 2025).

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) on an LCD monitor (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. Responses were collected using an MRI-compatible 2-button fiber-optic response pad.

Data Analysis

Behavioral Analysis

Trials with incorrect responses (4.29% of trials) and RTs outliers (1.24% trials, RT shorter than 150 ms, considered anticipatory responses, or longer than 1,300 ms, reflecting extremely slow responses) were excluded from statistical analyses (Ratcliff, 1993; Whelan, 2008). 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). 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 (resulting in the exclusion of two runs). 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. Data quality was further assessed using the variance in the global signal (mean signal across all voxels within each run). Runs with global variance ≥ 0.1% were excluded, resulting in the exclusion of eight runs (see Supplementary Information for details). Ultimately, three participants were excluded because neither run met the quality criteria. All remaining participants retained both runs, except for three individuals who each contributed only one valid run. The retained functional images were then co-registered to each participant’s high-resolution anatomical scan in native space and subsequently normalized to the Montreal Neurological Institute (MNI) 152 template, with a resampled voxel size of 3×3×3 mm^3.

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 ±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^3 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).

Supplementary Materials

Sequence Optimization

The statistical power of effects in rapid ER-fMRI depends greatly on the particular sequence of events chosen (Liu & Frank, 2004; Wager & Nichols, 2003). Here, we compared the performance of Genetic Algorithm (Wager & Nichols, 2003), M-sequence (Liu, 2004; Liu & Frank, 2004; Liu et al., 2001), and random sequences under two different settings of ISI (Burock et al., 1998; Dale, 1999). Using simulation, we evaluated how these different stimulus sequences contribute to statistical power in the current study, with a specific focus on detection efficiency (i.e., the sensitivity of detecting activations) and estimation efficiency (i.e., the ability to accurately estimate the hemodynamic response function, HRF).

Specifically, following the definitions by Wagner and Nichols (2003), the efficiency of a sequence is defined as:

$$\xi = 1/\,\mathrm{trace}\left\{\mathrm{diag}(w) C Z^{-} K V K^{'} \left(Z^{-}\right)^{'} C^{'}\right\} \tag{1}$$

Where *diag*(*w*) is a diagonal matrix comprised of the elements of *w* (the weight vector for the prior comparisons); *Z* represents the filtered design matrix (*Z* = *KX*, *K*: filtering matrix; *X*: design matrix); and *V* is the correlation matrix of the errors. When *C* is the identity matrix, ξreflects the estimation efficiency for the HRF shape. Alternatively, when *C* represents the canonical HRF, it corresponds to the detection efficiency for the prior contrasts.

Additionally, we incorporated a high-pass filter with a cutoff time constant of 128s. To mitigate any nonlinear effects from ISIs shorter than 2 s, we set the activation upper bound to twice the maximum value of the classic HRF function (Wager & Nichols, 2003). Moreover, we used the first-order autoregressive model(Friston et al., 2000) to account for autocorrelation. Using the formula and the settings described above, the detection efficiency and estimation efficiency can be calculated for any stimulus sequence.

The simulation procedure was as follows: 1) using the GA toolbox provided byWager and Nichols (2003), we generated multiple GA stimulus sequences tailored to this experiment (300 runs, 30 repetitions each); 2) generated M-sequences using the Mseq2 program (Liu, 2004; Liu & Frank, 2004); 3) created 500 random sequences with our custom Optseq2 algorithm and selected the best-performing ones; 4) these three optimal sequences underwent ISI jittering (2.2s, 2.4s, 2.6s, 2.8s) for 500 iterations.

Fig S1 presents the detection efficiency (**A**) and estimation efficiency (**B**) for the different stimulus sequences. The simulation results indicate that, in the fixed (i.e., non-jitter) ISI condition, the GA-generated sequence exhibited the highest detection efficiency (approximately 30), followed by the random sequence (max value around 25.5), then the M-sequence which had the lowest detection efficiency (about 20). After Jittering the ISI, the trend in the detection efficiency remained consistent. In terms of the HRF estimation efficiency, the M-sequence performed best in both the jittered and fixed ISI conditions, followed by the GA sequence, while the random sequence showed the lowest efficiency. Notably, ISI jittering significantly improved the HRF estimation efficiency, with all three sequences performing better under jittered ISI compared to fixed ISI.

In summary, the use of the GA sequences effectively enhanced the detection efficiency. Moreover, by implementing a jittered ISI design, the algorithm ensured the optimal HRF estimation efficiency.

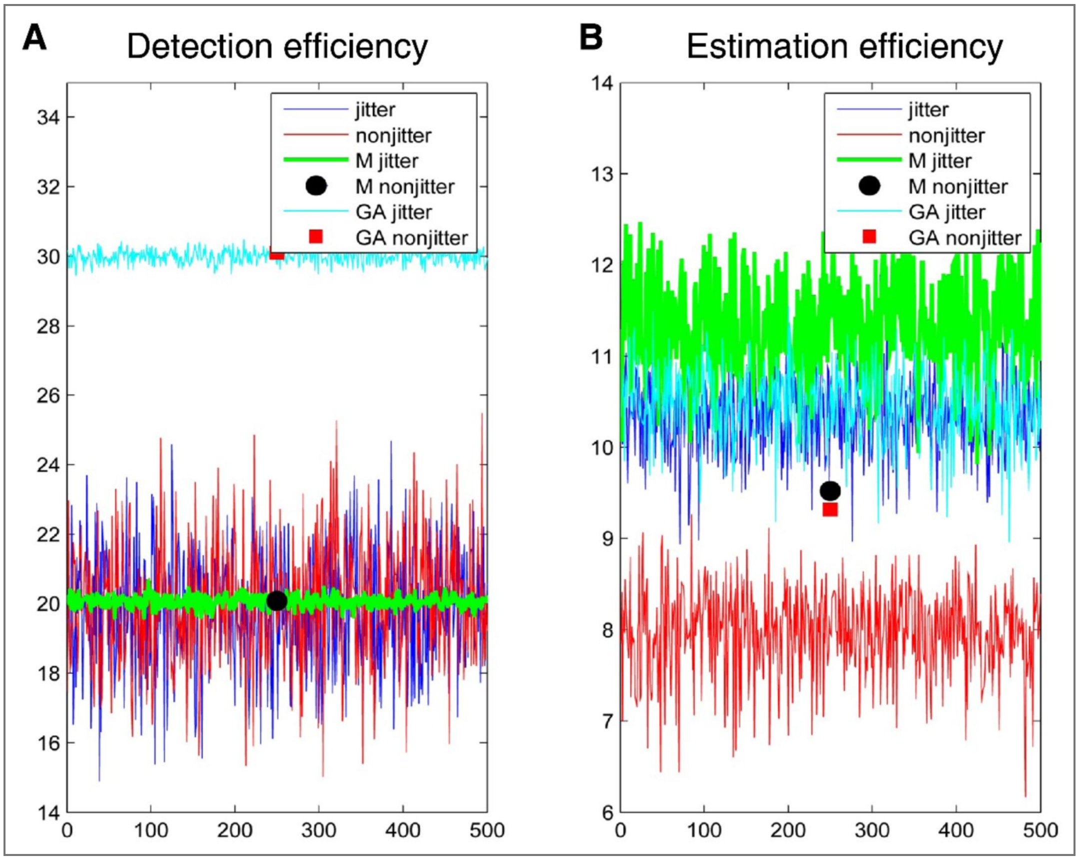


Fig. S1. Simulation results of different stimulus sequences. A. Detection efficiency. **B**. Estimation efficiency for the HRF shape.

Global Variance Quality Control

Global variance was used as a quality-control index to identify runs with abnormal signal variance. This procedure was applied after the standard preprocessing steps (including slice timing correction, 3D motion correction, spatial smoothing, linear trend removal, and high-pass temporal filtering). For each run, the global signal was computed as the mean signal across voxels with intensity values greater than 100 in each volume, thereby excluding background and non-brain voxels. The 100 threshold was applied to effectively distinguish brain tissue from low-intensity background noise. This resulted in a single global signal value per volume (TR). Global variance was then computed as the variance of the percentage global signal normalized to the mean (baseline). Specifically, for each run, the global signal time course GS_t was first normalized to its run mean $G\bar{S}$ and expressed in percentage units as $x_t = 100 \cdot \frac{GS_t}{G\bar{S}}$. Global variance was then computed as the sample variance of x_t, defined as $\frac{1}{N-1}\sum_{t=1}^{N}\left(x_t-\bar{x}\right)^2$, where N denotes the number of time point. Runs with global variance values equal to or over 0.1% were considered demonstrating abnormal signal variances (i.e., signal instability) and therefore excluded from further analyses. Fig. S2 shows the global signal time series of all the excluded runs that exceeded the 0.1% global variance threshold. Each panel corresponds to one excluded run, with the global variance value indicated in the panel title. The figure shows the excluded runs from Subjects 12 and 16 scanned at the Imaging Center for Brain Research, Beijing Normal University, and Subjects 14, 21, and 22 scanned at the Imaging Center of the First Affiliated Hospital of Soochow University.

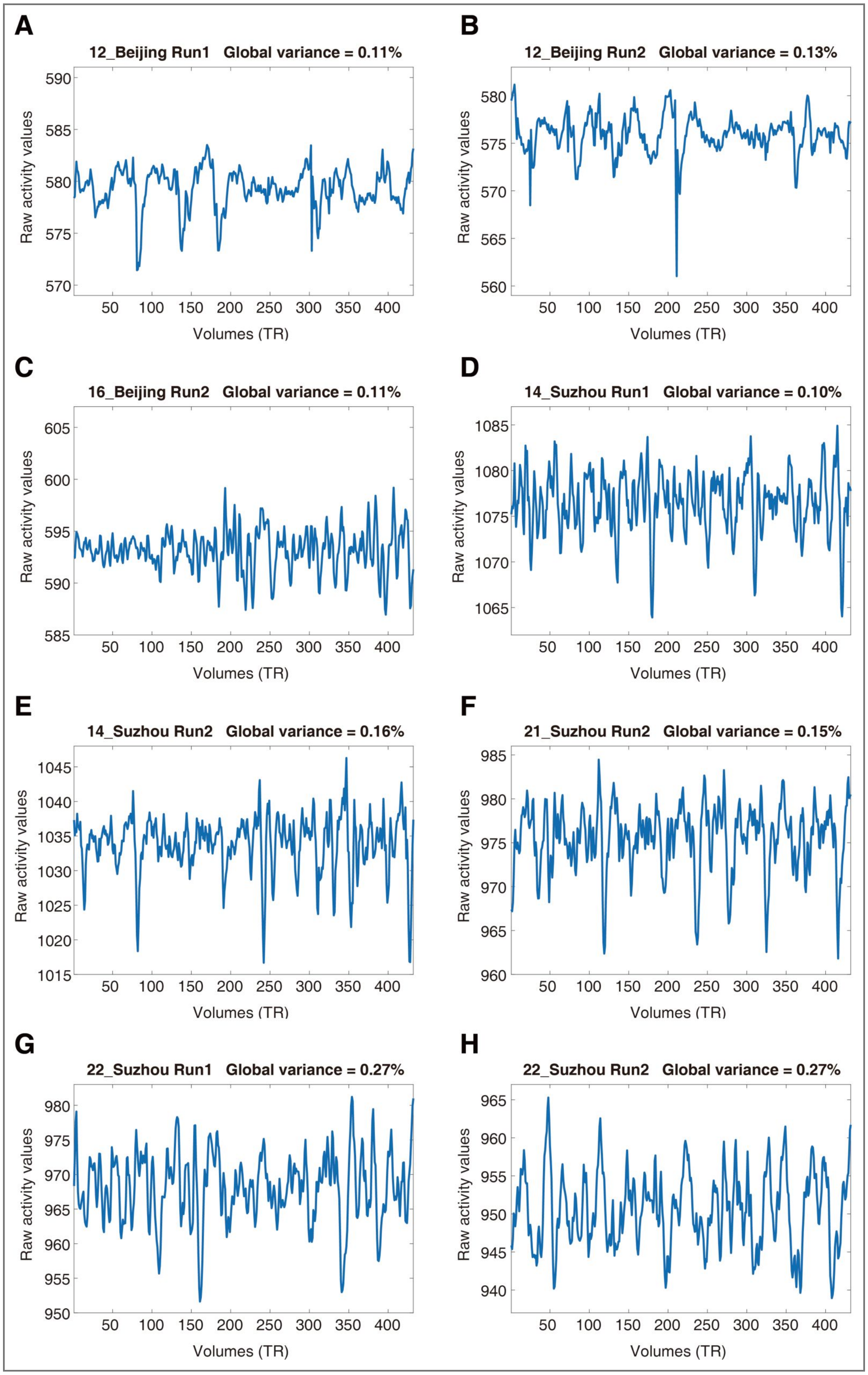


Fig. S2. Global signal time series for runs excluded from further analysis. Each panel shows the global signal time course of one run that had a global variance over the exclusion threshold (≥ 0.1%). Eight runs were excluded across Subjects 12 and 16 (Imaging Center for Brain Research, Beijing Normal University) and Subjects 14, 21, and 22 (Imaging Center of the First Affiliated Hospital of Soochow University).

Data availability

The processed data used for the final analyses are available at *https://github.com/yangzhangpsy/ER-fMRI-IOR*. All other study-related materials (raw data, analysis scripts, etc.) 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), Jiangsu Funding Program for Excellent Postdoctoral Talent (2024ZB496, 2025ZB642), and Natural Science Foundation of Jiangsu Province (BK20250796).

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
China Postdoctoral Science Foundation (中国博士后科学基金会职)	2024M752310	Yujie Chen
Jiangsu Funding Program for Excellent Postdotoral Talent	2024ZB496	Yujie Chen
Jiangsu Funding Program for Excellent Postdotoral 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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

Brainard D. H (1997) The psychophysics toolbox. *Spatial Vision* **10**:433-436 https://doi.org/10.1163/156856897x00357 | PubMed

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 | PubMed

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 | PubMed

Burock M. A., Buckner R. L., Woldorff M. G., Rosen B. R., Dale A. M (1998) Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *Neuroreport* **9**:3735-3739 https://doi.org/10.1097/00001756-199811160-00030 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

Chung R. S., Cavaleri 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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

Downar J., Crawley A. P., Mikulis D. J., Davis K. D (2002) A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. *Journal Of Neurophysiology* **87**:615-620 https://doi.org/10.1152/jn.00636.2001 | PubMed

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 | PubMed

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 | PubMed

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/10.1080/713755797

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”. *Psicologica* **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 | PubMed

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 | PubMed

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 | PubMed

Holm S (1979) A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics* **6**:65-70

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 | PubMed

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 | PubMed

Klein R. M (2000) Inhibition of return. *Trends Cogn Sci* **4**:138-147 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 | PubMed

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 | PubMed

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

Lepsien 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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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

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

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

Ranganath C., Rainer G (2003) Neural mechanisms for detecting and remembering novel events. *Nature Reviews Neuroscience* **4**:193-202 https://doi.org/10.1038/nrn1052 | PubMed

Ratcliff R (1993) Methods for dealing with reaction time outliers. *Psychological Bulletin* **114**:510-532 https://doi.org/10.1037/0033-2909.114.3.510 | PubMed

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 | PubMed

Satel J., Wilson N. R., Klein R. M (2019) What neuroscientific studies tell us about inhibition of return. *Vision* **3**:58 https://doi.org/10.3390/vision3040058 | PubMed

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. *Front Hum Neurosci* **12**:414 https://doi.org/10.3389/fnhum.2018.00414 | PubMed

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 | PubMed

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 | PubMed

Tang N., Li A., Chen B., Qiao Y., Lin Z., Zhang Y (2025) PsyRTDevice: Open-source toolkit for sub-millisecond evaluation of reaction time device accuracy. *The Innovation Informatics* **1** https://doi.org/10.59717/j.xinn-inform.2025.100017

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

Weber J. (2017) NeuroElf (Version 1.1). In [Computer program]. http://neuroelf.net

Whelan R (2008) Effective Analysis of Reaction Time Data. *The Psychological Record* **58**:475-482 https://doi.org/10.1007/BF03395630

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 | PubMed

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 | PubMed

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 | PubMed

Zhang G. L., Li A. S., Miao C. G., He X., Zhang M., Zhang Y (2018) A consumer-grade LCD monitor for precise visual stimulation. *Behav Res Methods* **50**:1496-1502 https://doi.org/10.3758/s13428-018-1018-7 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

Burock M. A., Buckner R. L., Woldorff M. G., Rosen B. R., Dale A. M (1998) Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. *NeuroReport* **9**:3735-3739 https://doi.org/10.1097/00001756-199811160-00030 | PubMed

Dale A. M (1999) Optimal experimental design for event-related fMRI. *Hum Brain Mapp* **8**:109-114 https://doi.org/10.1002/(sici)1097-0193(1999)8:2/3<109::aid-hbm7>3.0.co;2-w | PubMed

Friston K. J., Josephs O., Zarahn E., Holmes A. P., Rouquette S., Poline J. B (2000) To smooth or not to smooth? Bias and efficiency in fMRI time-series analysis. *NeuroImage* **12**:196-208 https://doi.org/10.1006/nimg.2000.0609 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 | PubMed

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 modulate 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.

Comments on revisions:

I appreciate the authors' thorough and thoughtful revisions, which have successfully addressed all of my prior concerns.

https://doi.org/10.7554/eLife.109842.2.sa3

Reviewer #2 (Public review):

This study provides neuroimaging evidence supporting the integration-segregation theory of inhibition of return (IOR), a widely studied attentional phenomenon. It also explores the neural interactions between IOR and cognitive conflict, demonstrating that conflict processing is potentially modulated by attentional orienting.

The integration-segregation theory was investigated using a sophisticated, well-executed experimental task that accounted for cognitive conflict processing, which is phenomenologically related to IOR but is non-spatial. The behavioral and neuroimaging data were carefully analyzed.

The authors have thoughtfully addressed all my previous concerns. By demonstrating how attentional orienting can modulate neural processing of cognitive conflict, this study helps to advance a more unified and mechanistic understanding of the cognitive and neural processes that govern our visual perception and response selection.

https://doi.org/10.7554/eLife.109842.2.sa2

Reviewer #3 (Public review):

Summary:

This study provides 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 rigorous, with appropriate preprocessing, modeling, and converging analyses across multiple contrasts. The results are theoretically coherent, demonstrating plausible dissociations between integration-related activity in the fronto-parietal attention network (e.g., FEF, IPS, TPJ, dACC) and segregation-related activity in medial temporal regions (e.g., PHG, STG). Importantly, the findings provide much-needed neural support for the integration-segregation framework and clarify how IOR modulates conflict processing.

Revisions and Evaluation:

The authors have responded thoroughly and convincingly to the concerns raised in the previous round of review. In particular, issues related to the interpretation of dACC activity, the functional characterization of PHG and STG, and reporting clarity have been carefully addressed. The manuscript has been improved in terms of transparency, consistency of reporting, and overall readability.

As a result, I no longer see any major weaknesses. The study is now clearly presented, methodologically sound, and theoretically informative. It makes a valuable contribution to the literature on attention and cognitive control.

Comments on revisions:

I appreciate the authors' efforts in addressing the previous comments. They have responded thoroughly to the concerns raised in the prior round of review. The work is well executed and makes a meaningful contribution to the field.

https://doi.org/10.7554/eLife.109842.2.sa1

Author response:

The following is the authors’ response to the original reviews

> ***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.*

We appreciate the positive reception to our manuscript. In the revised manuscript, we have further clarified the logic underlying the task design, adopted a more cautious tone in interpreting the behavioral and neuroimaging results, and enhanced the transparency of reporting and presentation.

> ***Public 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.*

We thank the reviewer for these helpful comments. In the revised manuscript, we have provided additional clarification regarding the SI and RI conditions (page 29), expanded the justification for the behavioral trial exclusion criteria (page 32), and clarified how the null condition was modeled and incorporated into the analyses (page 29). In addition, we have revised the description of the behavioral results to adopt more cautious wording, particularly given the absence of a significant interaction effect. For detailed responses to these specific points, please refer to the "Recommendations for the Authors" section below.

> ***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.*

We thank the reviewer for raising these important points. In the revised manuscript, we have further clarified the rationale of the experimental design and the motivation for introducing cognitive conflict.

First, we clarified that previous neuroimaging studies relied primarily on SOA-based contrasts, which capture the temporal dynamics of attentional orienting but do not directly distinguish the functional processes of integration and segregation. We therefore established the direct comparison between cued and uncued targets in the long SOA as the critical test required by the theory, as these conditions are hypothesized to engage integration and segregation processes, respectively (pages 6-7, “The Challenge of Neural Verification”). Crucially, to successfully implement this comparison, we highlighted the specific methodological advantage of our study: the use of a Genetic Algorithm (GA) to optimize the stimulus sequence. We explained how this design maximizes statistical power specifically for contrast detection (i.e., cued vs. uncued) while maintaining high estimation efficiency, thereby directly overcoming the power constraints that had likely obscured these subtle neural signatures in prior ER-fMRI work (pages 7-8).

Second, we clarified that the manipulation of cognitive conflict was introduced with the additional aim of examining IOR expression mechanisms, specifically investigating how spatial attention modulates ongoing cognitive processing after target onset, rather than the generation of IOR itself. We have now provided a clearer rationale for embedding a modified Stroop task within the cue-target paradigm, and explained how this design allows us to dissociate semantic and response conflicts while avoiding methodological confounds present in previous studies (page 8).

> *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.*

We thank the reviewer for pointing out the interpretational challenges in Figure 4. To address this, we have revised Figure 4 and provided a clearer and more precise interpretation of these interaction effects in the manuscript.

First, we have added explicit panel titles to Figure 4 (page 17). Panel A is now clearly labeled as the “Effect of IOR on Semantic Conflict”, while Panel B is labeled as the “Effect of IOR on Response Conflict”. We hope this visual labeling helps readers clearly identify the IOR modulation effects specific to each conflict type.

Second, we have revised the figure caption to explicitly define the interaction contrasts used to quantify these modulations, providing specific formulas (e.g., [UncuedRI – Uncued-SI] > [Cued-RI – Cued-SI] for response conflict) to ensure transparency.

Finally, regarding the reviewer’s comment on “deactivation”, we realized that our original figure terminology (e.g., “IOR effect under...”) might have caused confusion by mixing the interaction effect with the IOR effect itself. We have clarified that Figure 4 specifically illustrates the “Effect of IOR on the Semantic Conflict and the Response Conflict” (i.e., interaction effect between IOR and cognitive conflict). To interpret this interaction, we further examined the simple effects of conflict under each cueing condition. Specifically, we analyzed the neural signatures of semantic conflict (SI minus NE) and response conflict (RI minus SI) separately for the cued and uncued targets. Importantly, regarding the nature of the IOR effect itself (as displayed in Figure 3, page 14), it is not simply a uniform deactivation. Instead, by directly comparing the cued and uncued conditions for the neutral words, we observed neural changes in two directions: some specific regions exhibited an increased activation (Cued > Uncued), while others showed a reduced activation (Uncued > Cued). These differential patterns involved distinct brain networks and corresponded to the distinct integration and segregation mechanisms, respectively, rather than a global loss of activation (pages 20-21).

> ***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.*

We thank the reviewer for the positive and constructive assessment of our study. In response to the concerns raised, we have carefully revised the manuscript and addressed all points in detail below. In brief, we have clarified key interpretation issues in the Discussion section, including the complementary roles of dACC activation and the distinction between statistical clustering and functional interpretation of PHG and STG activations (pages 20-21). We have also improved transparency and reporting throughout the manuscript by providing more detailed sample characteristics, clarifying exclusion criteria and global variance computation, adding illustrative supplementary figures, and standardizing statistical reporting and abbreviations (pages 28, 33). Finally, we have added a concise paragraph on limitations of the study to provide a more balanced interpretation of the findings (pages 26-27). Detailed, point-by-point responses to all specific comments are provided below (see the “Recommendations for the authors” Section).

> ***Recommendations for the authors:***

> ***Reviewer #1 (Recommendations for the authors):***
>
> *Specific comments:*
>
> *(1) The figure caption contains an unclear sentence (lines 195-196): "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." This description is ambiguous and should be revised for clarity.*

Thanks for pointing this out. In the revised manuscript, we have rephrased the figure caption to improve clarity as follows (pages 9-10):

“Each trial started with a 150-ms non-informative cue presented at one of the two peripheral boxes. After a 150-ms interstimulus interval (ISI), a 150-ms fixation cue was presented at the central fixation box. Following a further 450-ms ISI, the target, a colored Chinese character, appeared at one of the two target locations with equal probabilities and remained on the screen for 450 ms. The trial ended with a variable intertrial interval (ITI) of 850, 1050, 1250, or 1450 ms (with equal probabilities).”

> *(2) Please provide a more detailed and clearer description of the SI and RI experimental conditions in the Methods section.*

Thanks for this helpful suggestion. We have revised the Methods section to provide a more detailed description of the SI and RI conditions. Specifically, we have further described the stimulus-response mapping and clarified how the SI and RI conditions are defined based on whether the ink color and the character meaning fell into the same or different response categories under this mapping. In addition, we have added a clarification in the Methods section to make it clearer that the SI trials involved semantic conflict without response conflict, whereas RI trials involve both semantic and response conflicts (page 29).

> *(3) As the data were collected across two research centers, please clarify the number of participants enrolled at each site.*

Thanks for this suggestion. We have now explicitly stated in the Apparatus and Data Acquisition section that 16 participants were enrolled at each site. The revised text reads (page 31)：

“The imaging data were acquired at two research sites following comparable protocols, with equal numbers of participants scanned at each site (n = 16 per site).”

> *(4) In the behavioral data analysis, please provide the rationale or justification for the criteria used to exclude trials.*

Thanks for this comment. In the revised manuscript (page 32), we have clarified that reaction times (RTs) shorter than 150 ms were excluded as anticipatory responses, and RTs longer than 1,300 ms were excluded to limit the influence of unusually slow responses. These exclusion criteria are commonly adopted in RT research and were applied consistently across all conditions (Ratcliff, 1993; Whelan, 2008).

> *(5) Given that the behavioral interaction effect was not statistically significant, the conclusion on lines 236-237, "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)" appears overstated and should be softened accordingly.*

We thank the reviewer for this important comment. We have clarified that our original statement was intended to highlight the successful isolation of conflict types based on the significant main effects of congruency (validating the task design), rather than implying a significant interaction effect. However, we agree that the original phrasing appeared unclear in this context. We have therefore revised the sentence to adopt a more cautious tone in the revised manuscript (page 12):

“These data demonstrated typical Stroop interference effects (Veen & Carter, 2005) in both the semantic (SI-NE difference) and response conflicts (RI-SI difference).”

> *(6) The statement on lines 281-282, "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" lacks supporting statistical evidence. Please report the relevant test statistics.*

We appreciate the reviewer’s careful reading and note that the relevant statistical evidence was missing from the original manuscript. This has now been added in the revised version. Specifically, we examined the interactions between cue validity and semantic conflict (SI vs. NE) as well as between cue validity and response conflict (RI vs. SI). Neither interaction was significant (see revised Results for full statistics on page 12), supporting our original statement that cue validity did not modulate either conflict component in behavioral performance.

> *(7) The manuscript mentions that a null condition (with no Chinese character presented) was included to increase statistical power for detecting differences across conditions. However, it is unclear how this null condition was actually used in the data analyses. Please clarify the role of the null condition in both the behavioral and neuroimaging analyses.*

Thanks for this comment. We regret that this was not sufficiently clear in the original manuscript. The null condition was included for neuroimaging purposes and was not used in the behavioral analyses, as no response was required in these trials. In the fMRI analyses, null trials served as the implicit baseline and were not modeled as regressors of interest. Task-related activities for all experimental conditions were therefore estimated relative to this null baseline, facilitating estimations of task-related responses in randomized event-related designs (Burock et al., 1998; Friston et al., 1999; Liu, 2004). We have clarified this point in the revised manuscript (page 29).

References

Burock, M. A., Buckner, R. L., Woldorff, M. G., Rosen, B. R., & Dale, A. M. (1998). Randomized event-related experimental designs allow for extremely rapid presentation rates using functional MRI. NeuroReport, 9(16), 3735-3739. *https://doi.org/10.1097/00001756-199811160-00030*

Friston, K. J., Zarahn, E., Josephs, O., Henson, R. N. A., & Dale, A. M. (1999). Stochastic designs in event-related fMRI. NeuroImage, 10(5), 607-619. *https://doi.org/10.1006/nimg.1999.0498*

Liu, T. T. (2004). Efficiency, power, and entropy in event-related fMRI with multiple trial types: Part II: design of experiments. NeuroImage, 21(1), 401-413. *https://doi.org/10.1016/j.neuroimage.2003.09.031*

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. Psychological Bulletin, 114(3), 510-532. *https://doi.org/10.1037/0033-2909.114.3.510*

Whelan, R. (2008). Effective analysis of reaction time data. The Psychological Record, 58(3), 475-482. *https://doi.org/10.1007/BF03395630*

> ***Reviewer #2 (Recommendations for the authors):***

> *(1) The paper is a bit too lengthy, with a lot of information that is hard for non-experts to grasp.*

We thank the reviewer for this comment. We realized that the Introduction was the most challenging section for general readers. In the revision, we refined the text in the Introduction for a better structure and more reader-friendly wording to improve readability. In addition, following the reviewer's suggestion (Recommendation 4 below), we have added short subsection titles to the Introduction, Results, and Discussion sections to better organize the content and highlight the main ideas. We hope these revisions make the manuscript more accessible and easier for a broader audience to follow.

> *(2) Please double-check the stats, as some of the results presented in the main text do not align well with the figures. Take Figure 2 as an example.*

We appreciate the reviewer's concern and have double-checked all statistics. All the results are consistent between the figures and the main text. Take Figure 2 as an example (page 12), the perceived discrepancy probably was caused by the fact that the descriptive values reported in the main text are marginal means for the main effects (i.e., the overall average of one factor, collapsed over the other factor), whereas Figure 2 shows the mean for each Congruency × Cue Validity condition (i.e., simple effect).

> *(3) The reasoning that the neuroimaging findings support the dissociation between integration and segregation needs to be improved.*

We thank the reviewer for this important comment. In the revised Discussion (pages 1921), we have strengthened the reasoning linking our neuroimaging findings to the dissociation between the integration and segregation processes. Specifically, we make it clear how the distinct activation patterns observed for the cued and uncued targets map onto the different functional demands proposed by the integration-segregation theory. The cued targets were theorized to recruit the frontoparietal attentional control networks, consistent with the re-engagement of an existing object file (integration). On the other hand, the uncued targets should engage the medial temporal and temporal association regions responsible for novelty detection and episodic encoding, consistent with the creation of a new object file (segregation). We hope the reviewer finds that the revision offers a clearer explanation of how the observed neural patterns are consistent with a dissociation between the integration and segregation processes.

> *(4) Please use short section titles to organize the introduction, results, and discussion sections. For instance, the discussion section is a long chunk of text (almost 9 pages) and is pretty dense, making it hard to quickly grasp the ideas the authors want to convey.*

Thanks for this helpful suggestion. Following the reviewer's recommendation, we have now added short subsection titles to the Introduction and Discussion sections to improve structure and readability. For the Results section, we have maintained and further refined the existing subheadings to ensure consistent organization.

> ***Reviewer #3 (Recommendations for the authors):***
>
> *I found this manuscript to be a timely and substantive contribution to the study of attention and cognitive neuroscience. To my knowledge, it provides the first direct neuroimaging evidence relevant to the integration-segregation theory of exogenous attention, a framework that has been influential in behavioral work for more than two decades but has lacked clear neural support. The study is conceptually well motivated, methodologically solid, and generally clearly reported. The findings differentiate neural substrates associated with integration and segregation processes and further show how*

> *inhibition of return (IOR) interacts with semantic and response conflicts at the neural level.*
>
> *The manuscript is well organized, the writing is mostly clear, and the progression from theory to hypotheses and methods is easy to follow. The combination of IOR with a modified Stroop paradigm is a clever choice that extends the theoretical scope of exogenous attention research. The use of an optimized event-related fMRI design based on a genetic algorithm is also a strength and reflects careful attention to design efficiency.*
>
> *The main results are internally consistent and theoretically meaningful. Integration related activity in the fronto-parietal attention network (including FEF, IPS, TPJ, and dACC) and segregation-related activity in medial temporal areas (PHG and STG) it well with the proposed framework, and the pattern of activations is coherent across analyses.*
>
> *Overall, I think this is a carefully executed study that offers much-needed neural evidence bearing on the integration-segregation theory of exogenous attention. I would recommend the following revisions.*
>
> *Suggestions:*
>
> *(1) In the Discussion (pp. ~17-18), dACC activation is described both in terms of general cognitive control demands and as reflecting a possible inhibitory bias toward the cued direction. It would help the reader if you could briefly indicate whether you see these as complementary (e.g., dual roles within the same region) or as more competing interpretations.*

We thank the reviewer for this helpful comment. We have clarified in the revised manuscript that dACC exerts general cognitive control demands and biasing against the cued direction are complementary rather than competing interpretations. Specifically, we described how the dACC is involved in both the cognitive control required for target integration and the inhibitory bias toward the cued location, thereby highlighting its dual roles within the same region. The revised section reads as follows (page 20):

“Furthermore, the observed increase in the left dACC activity under the cued relative to the uncued condition likely reflected the engagement of cognitive control mechanisms (Botvinick et al., 2004; Chung et al., 2024; Mayer et al., 2012; Veen & Carter, 2005), particularly in resolving the conflict between the task-driven requirement of target integration and the reduced accessibility of the cue-initiated representation. In this context, the heightened activation of dACC may also reflect its role in fulfilling the inhibitory bias toward the cued location (Mayer et al., 2004) and discouraging inefficient integration attempts at a location marked as less relevant.”

> *(2) In the Discussion, you could consider adding a short paragraph explicitly acknowledging a few limitations and how they might constrain generalization of the findings. A concise reflection of this kind would give a more balanced picture without undermining the main conclusions.*

We appreciate this helpful suggestion. In the revised manuscript, we have added a concise paragraph explicitly addressing a key limitation of the present study (pages 26-27). Specifically, we acknowledge that the absence of behavioral interactions alongside clear neural effects requires cautious interpretation. We discussed how this dissociation may reflect differences in measurement sensitivity between behavioral and neural indices, consistent with prior findings (Chen et al., 2006; Wilkinson & Halligan, 2004). We also note that the use of a GA-optimized sequence, while improving statistical efficiency, may have introduced unintended regularities in event order that could influence behavioral strategies.

> *(3) Since the dataset is hosted on GitHub, adding a short note in the Data Availability section about whether the repository will also include analysis scripts or future replication data would further enhance transparency and long-term usefulness.*

Thanks for this helpful suggestion. We have revised the Data Availability section (page 35) to clarify that the GitHub repository contains the processed data used in the final analyses. Analysis scripts and additional materials for replication are available from the authors upon reasonable request.

> *(4) In the Results section, the formatting of statistics is not fully consistent. For example, some reports use spaces around symbols (e.g., "η^2 = 0.301") whereas others do not (e.g., "p< .001"). It would be good to standardize this (e.g., "p < .001", "η^2 = .30") across the manuscript.*

Done as suggested.

> *(5) A few abbreviations appear before they are defined-for instance, SPC (superior parietal cortex) shows up in the Results (response conflict section) before the full name is given. Ensuring that each abbreviation is defined at first mention would help readers who may be less familiar with all of the regional acronyms.*

Thanks for this comment. We have conducted a thorough check of the manuscript and ensured that all abbreviations are defined upon their first occurrence.

> *(6) The text sometimes refers to "PHG/STG" as a combined cluster, while at other points, PHG and STG are described separately. It would be useful to clarify under what circumstances they are treated as a single functional cluster versus distinct regions of interest, and to keep the nomenclature as consistent as possible between the main text and the tables.*

Thanks for raising this point. In the revised manuscript, we have clarified this issue by distinguishing between statistical clustering and functional interpretation. In the whole brain analysis, activations in the left hemisphere formed a single continuous cluster spanning the PHG and STG; therefore, this cluster is labeled as “PHG/STG” in Table 1. We have explicitly noted the continuous nature of this cluster in the Results section (page 15) to ensure clarity:

“Notably, in the left hemisphere, these activations formed a continuous cluster spanning both regions (labeled as PHG/STG in Table 1).”

> *(7) It would be helpful to provide a bit more detail about the sample characteristics (e.g., age range, handedness, and inclusion/exclusion criteria) and to state explicitly how many participants, if any, were excluded from the analyses and for what reasons. This would help readers better evaluate data quality and generalizability.*

Thanks for this helpful suggestion. We have revised the Participants section (page 28) to provide the full details regarding our sample:

“32 healthy participants with normal or corrected-to-normal vision and normal color vision were recruited. All participants were right-handed and reported no history of neurological or psychiatric 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; aged 18-30 years, M = 22.69, SD = 2.58).”

Furthermore, we have provided a clearer description of the exclusion criteria in the Data Analysis section (pages 33-34) as follows:

"Runs with motions exceeding one voxel length in any direction were excluded (resulting in the exclusion of two runs) …Runs with global variance equal to or over 0.1% were excluded, resulting in the exclusion of eight runs (see Supplementary Information for details). Ultimately, three participants were excluded because neither run met the quality criteria. All remaining participants retained both runs, except for three individuals who each contributed only one valid run."

> *(8) Given that participants were excluded based on global variance exceeding 0.1%, it would be very informative to include, in the Supplementary Materials, an illustrative figure showing the signal time series (or global signal variance over time) for excluded participants.*

We appreciate this valuable suggestion. In the revised Supplementary Materials, we have included a new figure (Figure S2) that plots the global signal time series for the excluded runs to illustrate the signal patterns that led to their exclusion based on global variance.

> *(9) Relatedly, it may help to more explicitly describe how global variance was computed (e.g., over which time window, after which preprocessing steps, and whether it was calculated on whole-brain signal or within specific masks). A concise clarification would make the exclusion criterion easier to interpret.*

Thanks for this helpful suggestion. We have now clarified in the manuscript how global variance was computed (page 33) and have also provided a more detailed description of the computation procedure in the Supplementary Materials (page 4). Specifically, after the standard preprocessing (slice timing correction, 3D motion correction, spatial smoothing, linear trend removal, and high-pass temporal filtering), the global signal was computed for each run as the mean signal across voxels with intensity values greater than 100 in each volume. Global variance was then quantified as the temporal variance of this run-wise global-signal time course across all volumes, providing a quality-control index of signal stability.

> *(10) Rather than only reporting a single overall exclusion rate (e.g., 5.52% of total trials), it would be informative to break this down by source, reporting separately the proportion of trials excluded as RT outliers and the proportion excluded due to response errors. This would further improve transparency regarding the behavioral preprocessing pipeline.*

Thanks for this helpful suggestion. We have now broken down the overall exclusion rate by source in the revised manuscript. Specifically, we reported that 4.29% of trials were excluded due to incorrect responses, and 1.24% of trials were excluded as RT outliers (page 32).

References

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. Trends in Cognitive Sciences, 8(12), 539-546. *https://doi.org/10.1016/j.tics.2004.10.003*

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(11), 1937-1946. *https://doi.org/10.1162/jocn.2006.18.11.1937*

Chung, R. S., Cavaleri, J., Sundaram, S., Gilbert, Z. D., Del Campo-Vera, R. M., Leonor, A., Tang, A. M., Chen, K.-H., Sebastian, R., Shao, A., Kammen, A., Tabarsi, E., Gogia, A. S., Mason, X., Heck, C., Liu, C. Y., Kellis, S. S., & Lee, B. (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*

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(7), 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(12), 2843-2855. *https://doi.org/10.1002/hbm.21405*

Veen, V. V., & Carter, C. S. (2005). Separating semantic conflict and response conflict in the Stroop task: A functional MRI study. Neuro Image, 27(3), 497-504. *https://doi.org/10.1016/j.neuroimage.2005.04.042*

Wilkinson, D., & Halligan, P. (2004). The relevance of behavioural measures for functional imaging studies of cognition. Nature Reviews Neuroscience, 5(1), 67-73. *https://doi.org/10.1038/nrn1302*

https://doi.org/10.7554/eLife.109842.2.sa0