



Published in final edited form as:

Neuroimage. 2025 May 01; 311: 121214. doi:10.1016/j.neuroimage.2025.121214.

Blending into naturalistic scenes: Cortical regions serving visual search are more strongly activated in congruent contexts

Ilenia Salsano^{a,b}, Nathan M. Petro^{a,b}, Giorgia Picci^{a,b,c}, Aubrie J. Petts^{a,b}, Ryan J. Glesinger^{a,b}, Lucy K. Horne^{a,b}, Anna T. Coutant^{a,b}, Grace C. Ende^{a,b}, Jason A. John^{a,b}, Danielle L. Rice^{a,b}, Grant M. Garrison^{a,b}, Kennedy A. Kress^{a,b}, Valerio Santangelo^{d,e}, Moreno I. Coco^{d,f}, Tony W. Wilson^{a,b,c,*}

^aInstitute for Human Neuroscience, Boys Town National Research Hospital, Boys Town, NE, United States

^bCenter for Pediatric Brain Health, Boys Town National Research Hospital, Boys Town, NE, United States

^cDepartment of Pharmacology & Neuroscience, Creighton University, Omaha, NE, United States

^dSanta Lucia Foundation IRCCS, Rome, Italy

^eDepartment of Philosophy, Social Sciences & Education, University of Perugia, Perugia, Italy

^fSapienza University of Rome, Rome, Italy

Abstract

Visual attention allows us to navigate complex environments by selecting behaviorally relevant stimuli while suppressing distractors, through a dynamic balance between top-down and bottom-up mechanisms. Extensive attention research has examined the object-context relationship. Some studies have shown that incongruent object-context associations are processed faster, likely due to semantic mismatch-related attentional capture, while others have suggested that schema-driven facilitation may enhance object recognition when the object and context are congruent. Beyond

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

*Corresponding author at: Brookhouser Endowed Chair & Director, Institute for Human Neuroscience, 14090 Mother Teresa Ln., Boys Town, NE 68010, United States., tony.wilson@boystown.org (T.W. Wilson).

CRediT authorship contribution statement

Ilenia Salsano: Writing – review & editing, Writing – original draft, Visualization, Validation, Investigation, Formal analysis, Data curation. **Nathan M. Petro:** Writing – review & editing, Investigation, Formal analysis, Data curation. **Giorgia Picci:** Writing – review & editing, Supervision, Investigation, Formal analysis. **Aubrie J. Petts:** Writing – review & editing, Investigation, Data curation. **Ryan J. Glesinger:** Writing – review & editing, Project administration, Investigation, Data curation. **Lucy K. Horne:** Writing – review & editing, Investigation, Data curation. **Anna T. Coutant:** Writing – review & editing, Investigation, Data curation. **Grace C. Ende:** Writing – review & editing, Project administration, Investigation, Data curation. **Jason A. John:** Writing – review & editing, Project administration, Formal analysis, Data curation. **Danielle L. Rice:** Writing – review & editing, Investigation, Data curation. **Grant M. Garrison:** Writing – review & editing, Investigation, Data curation. **Kennedy A. Kress:** Writing – review & editing, Investigation, Data curation. **Valerio Santangelo:** Writing – review & editing, Resources, Conceptualization. **Moreno I. Coco:** Writing – review & editing, Resources, Conceptualization. **Tony W. Wilson:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization.

Data and code availability

The data used in this article will be made publicly available through the COINS framework at the completion of the study (<https://coins.trendscenter.org/>).

Declaration of competing interest

The authors of this manuscript report no conflicts of interest, financial or otherwise.

the conflicting findings, translation of this work to real world contexts has been difficult due to the use of non-ecological scenes and stimuli when investigating the object-context congruency relationship. To address this, we employed a goal-directed visual search task and naturalistic indoor scenes during functional MRI (fMRI). Seventy-one healthy adults searched for a target object, either congruent or incongruent within the scene context, following a word cue. We collected accuracy and response time behavioral data, and all fMRI data were processed following standard pipelines, with statistical maps thresholded at $p < .05$ following multiple comparisons correction. Our results indicated faster response times for incongruent relative to congruent trials, likely reflecting the so-called pop-out effect of schema violations in the incongruent condition. Our neural results indicated that congruent elicited greater activation than incongruent trials in the dorsal frontoparietal attention network and the precuneus, likely reflecting sustained top-down attentional control to locate the targets that blend more seamlessly into the context. These findings highlight the flexible interplay between top-down and bottom-up mechanisms in real-world visual search, emphasizing the dominance of schema-guided top-down processes in congruent contexts and rapid attention capture in incongruent contexts.

Keywords

Visual attention; Frontoparietal; Ecological stimuli; Context incongruency; fMRI

1. Introduction

In daily life, visual attention helps us navigate a complex, stimulus-rich external world by suppressing irrelevant information and, at the same time, providing guidance towards relevant environmental stimuli based on a dynamic interplay between top-down internal goals and bottom-up perceptual features. In real world environments, objects are always embedded within complex scenes that are frequently arranged in predictable spatial and semantically coherent configurations (Henderson and Hollingworth, 1999; Bar, 2004; Pinker, 2007; Kaiser et al., 2019; McLean et al., 2023). Within this framework, investigating how the relationship between objects and their surrounding context affects attentional capture has been a controversial topic, with findings being generally mixed. Early studies argued that the recognition of objects within a scene is mediated by their semantic consistency, with object-context associations being learned through repeated exposure, forming high-level cognitive representations or schemas that encode contextual associations between objects and their surroundings (Biederman, 1972; Biederman et al., 1982; Chun, 2000; Bar and Aminoff, 2003). These schemas were thought to guide top-down processing, enabling the brain to anticipate and prioritize contextually congruent information (Biederman, 1972; Biederman et al., 1982; Chun, 2000; Bar and Aminoff, 2003). For instance, when searching for a pen, one is more likely to look on a desk than in a refrigerator, as schemas streamline the recognition of congruent objects in expected locations and configurations (Wolfe, 1994; Wolfe and Horowitz, 2004; Oliva and Torralba, 2007; Wolfe et al., 2011). As a result, congruent objects are often recognized faster and more accurately than incongruent ones, which often require more cognitive effort to reconcile their unexpected positioning within a scene (Biederman, 1972; Henderson and Hollingworth, 1999; Chun, 2000; Bar and Aminoff, 2003; Oliva and Torralba,

2007; Kaiser et al., 2019). This facilitative effect of congruence has been extensively documented in studies of object recognition, which emphasize the role of pre-activated schemas in facilitating perceptual and attentional processes (Biederman, 1972; Henderson and Hollingworth, 1999; Henderson et al., 1999; Davenport and Potter, 2004; Fenske et al., 2006). Congruent objects are also thought to benefit from priming effects, where the surrounding context activates semantic and spatial networks, reducing cognitive load and enhancing processing speed and accuracy (Henderson and Hollingworth, 1999; Chun, 2000; Oliva and Torralba, 2007). Conversely, incongruent objects, those that violate contextual expectations, would not benefit from schema-driven facilitation, leading to less accurate recognition (Biederman, 1972; Biederman et al., 1982; Boyce and Pollatsek, 1992; Bar and Ullman, 1996; Davenport and Potter, 2004; Fenske et al., 2006) and longer search times (De Graef et al., 1990).

However, contrary to this perspective, other studies have reported an advantage for incongruent objects (Loftus and Mackworth, 1978; Bonitz and Gordon, 2008; Underwood et al., 2008; LaPointe and Milliken, 2016; Borges et al., 2020). The interpretation of such findings is generally that out-of-context objects can introduce a breach of expectation, capturing attention and enhancing encoding and recognition (Pezdek et al., 1989; Hollingworth and Henderson, 1998). For example, a semantically incongruent object, such as a tennis ball in the fridge, rapidly captures attention because it stands out in this context (i.e., a pop-out effect), and this distinctiveness enhances the likelihood that the object will be later remembered (Hollingworth and Henderson, 1998; Pezdek et al., 1989). Of note, most studies supporting either congruence or incongruence advantages have relied on object recognition paradigms within non-ecological settings (e.g., using objects that were large, positioned close to the center of the screen, and/or presented very briefly). While these studies have provided valuable insights, their limited ecological validity warrants caution in translating findings to real-world complex environments.

Given these limitations, several recent studies have strived for more realistic settings by presenting objects embedded within naturalistic visual scenes and employing paradigms such as change detection, object recognition, or memory encoding (Silva et al., 2006; Underwood et al., 2006; Stirk and Underwood, 2007; Spotorno et al., 2013; Coco et al., 2014, 2020; Santangelo et al., 2015; see also Santangelo, 2015, for an extensive review). However, even in such naturalistic scenarios, findings have remained mixed, with some studies reporting a congruency advantage (Spotorno et al., 2013; Coco et al. 2014) and others showing the opposite (Underwood et al., 2006; Stirk and Underwood, 2007; Santangelo et al., 2015; Coco et al., 2020). These conflicting findings could reflect many possible sources, including differences in task design, task instructions, and stimuli even within the non-naturalistic categories. As an example, object recognition studies often rely on brief stimulus presentation (e.g., less than 500 ms), emphasizing rapid perceptual categorization and schema activation at the expense of deep semantic encoding. Conversely, studies employing naturalistic scenes typically extend stimulus presentation times beyond 1000 ms to facilitate encoding processes and enable more in-depth semantic integration (Santangelo et al., 2015).

The critical brain regions and processes for context perception have been studied extensively using electroencephalography (EEG) and functional MRI (fMRI; Hamm et al., 2002; Bar and Aminoff, 2003; Ganis and Kutas, 2003; Goh et al., 2004; Jenkins et al., 2010; Demiral et al., 2012; Rémy et al., 2013, 2014, 2020; Võ and Wolfe, 2013; Mudrik et al., 2014; McAndrews et al., 2016; Brandman and Peelen, 2017; Coco et al., 2017; Caplette et al., 2020; Li et al., 2023). Unfortunately, the vast majority of these neuroimaging studies have used more non-ecological scenes, such as superimposed objects and/or brief presentations, limiting their applicability to realistic scenarios. For example, one study using rapidly presented scenes (i.e., 100 ms) with objects superimposed found increased activation in incongruent relative to congruent scenes in the right anterior parahippocampal cortex, which was a region of interest, as well as the right middle and inferior frontal gyri in their whole brain analysis (Rémy et al., 2014). Nonetheless, these studies have offered critical insight on the underlying brain regions, which have been slowly augmented by studies using more ecologically valid approaches. For instance, a study using naturalistic scenes and a delayed forced-choice task asking participants to recall the position of the target found that encoding context-congruent targets activated dorsal frontoparietal regions, while encoding context-incongruent targets deactivated ventral frontoparietal regions (Santangelo et al., 2015). Overall, the inconsistency in the behavioral findings and the limited number of neuroimaging studies that have used natural, realistic scenes to examine the object-context congruency effect make firm conclusions difficult.

In the current study, we investigate the object-context congruency effect using fMRI and a visual search task with naturalistic indoor scenes. Unlike previous studies that primarily focused on passive free exploration, object recognition, or categorization tasks, our approach required participants to actively search for a target object within the scene. Participants were first presented with a word cue (e.g., plate) indicating the target object, followed by a scene where the object appeared in either a congruent (e.g., kitchen) or incongruent (e.g., bathroom) context. By adding a visual search component, we aimed to closely align with real-world search scenarios where individuals often have prior knowledge about the target they are looking for, which in this case is provided by the cue word. Importantly, while previous studies have investigated congruency effects in categorization or recognition tasks, our study is the first to examine them in the context of goal-directed search, where congruency subtly influences search efficiency rather than being explicitly evaluated by participants. Furthermore, we leveraged a large sample size of 71 participants, an uncommon strength in this area of fMRI research, where small sample sizes have historically limited statistical power and replicability (Button et al., 2013; Turner et al., 2018). By addressing these limitations, we aim to enhance the robustness and generalizability of our findings, ensuring they contribute reliably to the broader understanding of object-context integration (Poldrack et al., 2011). Behaviorally, we hypothesized that locating a target object that blends seamlessly into a background scene would be a more difficult visual search, thus requiring more time to fully explore the scene and make a correct decision. Conversely, in incongruent trials, the semantic inconsistency of the target objects would cause them to stand out in the scene, effectively capturing attention and leading to shorter reaction times. Regarding brain activation, we expected that the more demanding visual search needs of the context congruent trials would necessitate stronger engagement of the dorsal frontoparietal

network, which is well known to be involved in top-down attentional control (Corbetta and Shulman, 2002; Corbetta et al., 2008).

2. Materials and methods

2.1. Participants

Seventy-two healthy volunteers participated in the study. One participant was excluded due to an accuracy rate below 70 % in the incongruent condition, resulting in a final sample of 71 participants (33 females; mean age: 32.87 years, range: 25–45 years) for data analysis. Exclusionary criteria included any medical illness affecting the CNS, any neurological or psychiatric disorder, history of head trauma, current substance use, and standard exclusion criteria for undergoing MRI (e.g., ferromagnetic implants). All participants had normal or corrected-to-normal vision and scored in the normal range on measures of cognitive function (i.e., Montreal Cognitive Assessment (MoCA); Nasreddine et al., 2005). All study procedures were approved by the Boys Town National Research Hospital's Institutional Review Board (IRB) and all participants provided written informed consent after a full description of the study.

2.2. Stimuli and procedure

While supine in the scanner, participants viewed a visual display through a mirror mounted on the MRI head coil (27.59° visual angle; screen resolution = 680 × 480; refresh rate = 60 Hz). Stimuli were presented using Psychtoolbox (psychtoolbox.org) running on MATLAB 9.5 (MathWorks, Natick, MA). The experimental stimuli were sourced from a previously published dataset (D'Innocenzo et al., 2022), consisting of 192 images of indoor scenes, which is a subset of a larger collection of normed naturalistic scenes (see the VISIONS dataset; Allegretti et al., 2025). Each scene contains a critical object either consistent or inconsistent with the scene's context and appearing on either the right or left side. Visual saliency across all scenes was balanced across the two conditions and left vs. right side (refer to Allegretti et al., 2025, for details about norms).

For this study, 120 unique pictures were selected (plus four additional practice images) to ensure that each target object and its corresponding cue word appeared only once, thereby preventing any memory-related effects or learned expectations about object-context congruency. Sixty images were used for “congruent” trials (object consistent with the scene background) and sixty for “incongruent” trials (object inconsistent with the scene background; Fig. 1, top panel). Each object was presented in only one scene and did not appear in both conditions across different trials. Additionally, within each congruent and incongruent condition, half of the target objects appeared on the left side of the scene and the other half on the right to ensure location balance. As detailed below, participants were instructed to respond based on the side (left or right) of the target within the scene and not on its congruence or incongruence with the scene. The experiment included two runs, each with 60 images, balanced for object-context congruency (i.e., congruent/incongruent) and object location (i.e., left/right). The order of images was pseudorandomized for each participant to avoid more than three consecutive repetitions of congruency condition and object location. Each trial began with a 500 ms cue word describing the target, followed by

a 1000 ms fixation cross, and then a 4000 ms presentation of the scene containing the target object. Participants were instructed to explore the scene freely and search for the target, responding as quickly and accurately as possible regarding the side (left side: right index finger; right side: right middle finger) where the target object appeared. Participants could respond using a button pad during the presentation and up to 1000 ms after the image offset. A variable inter-trial interval (750 to 1250 ms) followed before the next trial began.

2.3. fMRI data acquisition

All participants underwent structural and functional MRI (fMRI) scans on a 3T Siemens Prisma scanner equipped with a 32-channel head coil. Functional data were acquired using a multiband gradient-echo T2*-weighted echoplanar imaging (EPI) sequence sensitive to blood oxygenation level-dependent (BOLD) contrast. The sequence included 56 slices (TR = 480 ms, TE = 29.20 ms, flip angle = 44°, FOV = 248 mm, slice thickness = 3 mm, voxel size $3 \times 3 \times 3$ mm³ isotropic, 1002 vol, multiband factor = 8), covering the entire cortex. Structural images were collected using a T1-weighted MPRAGE sequence with 192 slices (TR = 2400 ms, TE = 2.05 ms, flip angle = 8°, FOV = 256 mm, slice thickness = 1 mm, voxel size $1 \times 1 \times 1$ mm³ isotropic).

2.4. fMRI data preprocessing and statistical analysis

fMRI data were preprocessed by using the MATLAB-based software CONN (Version 22.a; Whitfield-Gabrieli and Nieto-Castanon, 2012; MathWorks Inc., Natick, MA). Magnetic field inhomogeneities were corrected with fieldmaps processed through FSL's TOPUP toolbox (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/TOPUP>). These fieldmaps were integrated into CONN's indirect normalization pipeline, recommended for susceptibility distortion correction when high-quality fieldmaps are available (Calhoun et al., 2017). Functional data preprocessing steps included realignment and unwrapping to correct susceptibility distortions, slice-timing correction, and identification of outlier frames (flagged if framewise displacement exceeded 0.9 mm or global BOLD signal changes exceeded five standard deviations; Power et al., 2014). A reference BOLD image was computed for each participant by averaging all non-outlier scans. Functional and anatomical data were coregistered and normalized to MNI space, segmented into grey matter, white matter, and CSF tissue classes, and resampled to 2 mm isotropic voxels using the SPM unified segmentation and normalization algorithm (Ashburner and Friston, 2005; Ashburner, 2007) with the IXI-549 template. Functional data were smoothed with an 8 mm FWHM Gaussian kernel.

Statistical analyses were conducted using statistical parametric mapping (SPM12; Wellcome Department of Cognitive Neurology) implemented in MATLAB 9.5 (MathWorks Inc., Natick, MA). The time series at each voxel were high-pass filtered at 128 s and pre-whitened using an AR(1) autoregressive model. Statistical inference followed a two-step random effects approach (Penny and Holmes, 2004). First-level general linear model (GLM) analyses modeled four conditions of interest: incongruent left, incongruent right, congruent left, congruent right. Each condition was represented by delta functions time-locked to half of the scene presentation duration (2000 ms of the 4000 ms scene display) and convolved with SPM12's hemodynamic response function. Only correct trials were included. The single-subject models also included the onsets of missed trials, incorrect trials, cue word

(500 ms), fixation cross (1000 ms), and six motion parameters as covariates of no interest. Following GLM parameter estimation, a linear contrast was created by averaging the four main trial types across the two fMRI runs. This resulted in a contrast image for second-level analyses, which used a one sample t -test for overall task activation.

Since the main aim of the study was to investigate the congruency effect, we created linear contrasts by merging left and right trials into two conditions: incongruent and congruent. The corresponding two contrast images for each participant were entered into a paired t -test to compare the incongruent vs. congruent conditions and vice versa. An initial voxel-level threshold of p -uncorrected = .001 was applied to the output maps and the resulting clusters were corrected for multiple comparisons at the whole brain level for a final threshold of p -FWE = .05.

3. Results

3.1. Behavioral results

Paired t -tests on reaction time (RTs) and accuracy data (Fig. 1, bottom panel) revealed no significant difference in accuracy between congruent and incongruent conditions (Congruent: 97.4 %; Incongruent: 97.9 %; $t(70) = 1.725$, $p > .05$). In contrast, there was a significant RT difference, with participants responding more slowly in the congruent (1402 ms) than in the incongruent condition (1362 ms; $t(70) = 3.388$, $p = .001$).

3.2. fMRI results

The fMRI data showed increased activation during overall task performance within the dorsal frontoparietal attention network, specifically in the bilateral intraparietal sulci (IPS) and left frontal eye field (FEF). Large bilateral activations were also observed in the occipital cortex, encompassing the middle and inferior occipital gyri and the fusiform gyrus. Additionally, increased activation was found in the bilateral anterior insula, right cerebellum, and the left precentral gyrus (see Fig. 2, Table 1).

Regarding the conditional effects, whole-brain paired t -tests revealed a similar pattern, highlighting increased activation for congruent compared to incongruent trials in the dorsal frontoparietal attention network (FEF and IPS, bilaterally) and occipital regions (calcarine cortex and middle occipital gyrus, bilaterally). Furthermore, increased activation was observed in the bilateral precuneus (see Fig. 3, Table 1). Signal plots for all clusters were consistent, showing stronger activation for the congruent relative to incongruent condition (p -FWE = .05), with no voxels showing the opposite effect of significantly stronger activation during incongruent trials.

4. Discussion

The present study aimed to identify the brain regions underlying the object-context congruency effect and thereby to address inconsistencies in the literature regarding the effects of congruence and incongruence on visual processing by implementing a goal-directed visual search paradigm using naturalistic indoor scenes. Leveraging a large sample size and well controlled task design, we provide robust evidence on the neural mechanisms

underlying object-context integration. Participants were asked to locate a target object, either congruent or incongruent within the scene context, after being presented with a word cue describing the target object. We observed behavioral differences suggesting that identifying the target objects was more difficult in the congruent scenes (i.e., prolonged reaction time), which paralleled our observation of stronger activation during congruent trials across a distributed group of brain regions generally linked to visual attention function. These findings were broadly consistent with our hypotheses and below we discuss their implications for understanding the neural regions and processes underlying the widely studied object-context congruency effect.

Behaviorally, we did not observe a significant difference in accuracy between conditions. However, participants were significantly faster in the incongruent condition, a finding that may reflect a perceptual “popout” effect. Incongruent objects, by violating contextual expectations, likely captured attention more rapidly, facilitating their detection. This interpretation aligns with previous findings showing that semantically inconsistent objects can stand out due to their contextual mismatch, thus triggering early attentional engagement (Hollingworth and Henderson, 1998; Pezdek et al., 1989; Underwood et al., 2006; Stirk and Underwood, 2007; Santangelo et al., 2015; Borges et al., 2020; Coco et al., 2020; for a review see Santangelo, 2015). Conversely, the slower responses during congruent trials may reflect the additional attentional effort required to locate objects that blend seamlessly into their context. When objects align with the pre-activated schema (elicited by the word cue), their integration into the scene becomes so seamless that they lack visual distinctiveness, highlighting the well-known consistency effect (Hollingworth and Henderson, 1998; Pezdek et al., 1989).

The neural findings complement this interpretation. Across both task conditions, significant activation of the left FEF and bilateral IPS, key regions of the dorsal frontoparietal attention network, indicated the involvement of top-down, goal-driven attentional control during visual search (Corbetta and Shulman, 2002; Corbetta et al., 2008). These regions are known to be critical for sustaining and directing goal-oriented attention and mediating the deployment of attention based on expectations and schemas (Corbetta and Shulman, 2002; Corbetta et al., 2008). Specifically, the IPS is thought to facilitate attentional shifts within the scene, while the FEF helps coordinate eye movements and maintain attentional focus on task-relevant targets (Corbetta and Shulman, 2002; Corbetta et al., 2008). The bilateral visual occipital areas were also significantly activated, underscoring the importance of detailed visual processing across both conditions. Additionally, the bilateral anterior insula was significantly activated. This was not surprising, as the anterior insula is a hub for the salience network, which detects and processes behaviorally relevant stimuli, integrating attentional demands with task goals (Menon and Uddin, 2010). Its established role in mediating interactions between the dorsal and ventral attention networks further highlight its role in facilitating transitions between top-down and bottom-up attentional processes (Sridharan et al., 2008).

Given the goals of the study, our most important findings were the widespread differences in activation favoring the congruent trials. First, we observed significantly stronger activation in the bilateral FEF and IPS during congruent relative to incongruent trials, which aligns

with the interpretation that congruent trials demand more sustained top-down attentional guidance to locate targets embedded in object-rich complex scenes. Essentially, the congruent objects likely blended into the background of our naturalistic scenes, requiring participants to more strongly engage top-down attentional resources to filter through the scene and pinpoint the target. This “blending-in” would require more deliberate guidance of the visual search, consistent with the higher activation observed in the dorsal frontoparietal attention network. Similar results were found in the study by Santangelo and colleagues (2015), in which participants had to encode and subsequently remember the position of a target object extracted from the initial scene, with the target object being either congruent or incongruent with the context of the scene. Specifically, they found increased activation in the dorsal frontoparietal network, namely IPS, during the encoding of semantically consistent targets in congruent scenes (Santangelo et al., 2015). We also observed greater activation of the bilateral precuneus during congruent trials, which may suggest an additional layer of top-down processing related to the integration of task goals and scene context (Cavanna and Trimble, 2006). In fact, the precuneus has been implicated in schema-based scene understanding, facilitating the alignment of visual information with stored contextual knowledge (Aminoff et al., 2007). Thus, during congruent trials, its activation may reflect the process of matching the target object to its expected location within the scene, helping both attentional allocation and target verification. The precuneus has also been associated with sustained attention and visuospatial imagery during complex search tasks, which could also explain its stronger involvement in the congruent trials (Grosbras et al., 2005).

Another interesting feature of our results was the significant anterior insula activation observed across both task conditions. This region has not been widely reported in object-context congruency studies and this might stem from the nature of the current task. Specifically, in the current study, participants were asked to actively search for and locate targets according to a word cue that appeared before the realistic scene. In this structured, goal-directed task, the anterior insula may play a critical role in detecting and prioritizing semantically relevant stimuli by aligning its processing with the specific demands of the task. In other words, rather than allowing attention to be entirely captured by incongruent objects in a purely bottom-up manner, the anterior insula may help integrate the semantic information into the primary top-down goal of locating the target object described by the word cue. This would support the idea that the anterior insula acts as a mediator, balancing the need to respond to semantic features while ensuring that attentional resources remain focused on the task’s objectives (Menon and Uddin, 2010; Sridharan et al., 2008).

Overall, our findings expand prior research by showing how the naturalistic content of scenes and an ecologically valid search task converge to shape both behavioral and neural responses to object-context congruency. Unlike studies employing rapid recognition or categorization tasks, where congruent objects often facilitate processing, our structured visual search paradigm required sustained top-down attention over longer presentation times, thereby allowing deeper semantic integration of the scene and making congruent objects more difficult to detect. Consequently, incongruent objects stood out due to semantic mismatch, likely facilitating search through a pop-out effect. This adds a valuable piece to earlier work by reconciling previously inconsistent findings and demonstrating that, under more ecologically valid, goal-directed conditions, the attentional demands of detecting

congruent objects can paradoxically increase, engaging dorsal frontoparietal regions more strongly. Such insights underscore the importance of considering real world task demands, highlighting the flexibility of the attentional mechanisms involved when resolving the object-context semantic associations.

Before closing, it is important to note the limitations of this study. First, for neuroimaging, we relied on fMRI alone. While offering great spatial precision, fMRI lacks the temporal resolution needed to capture rapid attentional shifts during visual search. Future research could utilize a combined fMRI/EEG approach, or leverage magnetoencephalography (MEG; Wilson et al., 2016), to enable a more dynamic evaluation of real-time attentional deployment. Such a spatiotemporal mapping approach would enable behavioral and neural responses to be linked with greater precision, shedding light on the temporal evolution of semantically driven mechanisms. Such integrations could provide a more comprehensive understanding of the neural dynamics underlying attentional processing in structured and ecologically valid tasks. A second limitation is that we used only one behavioral task and only naturalistic scenes. Although this approach was intentional and enabled stronger conclusions regarding realistic visual search, future research should systematically vary task difficulty, stimulus complexity, and stimulus duration to investigate how these factors affect attentional processes and associated neural activation. For instance, manipulating object size or contrast relative to the background could reveal how such variations modulate both behavioral performance and activity in key attentional networks. Exploring these elements in a more controlled, multifactorial design would significantly enhance our understanding of attention in real-world contexts. Moreover, while the use of naturalistic indoor scenes increases the ecological validity of this study, it also introduces variability in scene complexity and object distinctiveness, although we controlled for such visual characteristics. These factors may modulate attentional demands differently across trials, potentially interacting with congruence to shape performance. A third limitation is that our study did not include detailed analyses correlating trial-level behavioral performance with neuroimaging data. While such an approach was beyond our current scope, future studies could employ advanced techniques (e.g., single-trial modeling, multivariate pattern analysis, or connectivity methods) to better capture how subtle variations in neural activity relate to behavioral outcomes. A fourth limitation is that the current study focused exclusively on attentional dynamics without considering the potential influence of emotional or motivational salience. Exploring the interplay between semantic and emotional relevance could uncover competitive or additive mechanisms that influence attentional deployment and should be a goal of future research. Specifically, incongruent objects with emotional valence might activate distinct neural pathways, either competing with or amplifying the effects of semantic inconsistency, thereby modulating performance. Investigating these factors could reveal critical interactions between attentional and affective systems, offering a more nuanced understanding of how multiple cognitive dimensions shape attentional processing. Finally, we focused on healthy adults and future work should address how object-context congruency is affected in healthy aging and/or those with psychiatric and neurological conditions. These populations often exhibit changes in attentional networks, and examining whether impairments in dorsal attention network activity are compensated

by increased engagement of salience-processing regions, such as the anterior insula, could provide valuable clinical insights.

In conclusion, these findings provide novel insight on the roles of congruence and incongruence in a goal-driven visual search task using naturalistic scenes. Some studies emphasize the facilitative role of congruence in visual processing by aligning with schemas in order to optimize attention (Bar and Aminoff, 2003; Biederman, 1972), while others have highlighted the attention-grabbing capabilities of incongruent stimuli (Loftus and Mackworth, 1978; Bonitz and Gordon, 2008). The present study builds on this framework by demonstrating that incongruent objects facilitate faster reaction times due to their distinctive nature, while congruent objects require stronger engagement of top-down schema-driven attentional mechanisms likely due to their natural blending into the scene, which ultimately results in slower behavioral responses and stronger activation. Our functional mapping results underscore the robust involvement of the dorsal frontoparietal attention network in congruent trials, likely reflecting the sustained engagement of top-down attentional processes. Additionally, the anterior insula may play a role in aligning or integrating the semantic information content of the stimuli with the current top-down goals related attentional processes. Together, these findings offer insights into the dynamic interplay between top-down and bottom-up mechanisms when processing congruent/incongruent object-scene associations in the context of a goal-driven visual search task, emphasizing the critical role of task structure in determining the relative attentional engagement. Importantly, this work leveraged a well-controlled task design and a large sample size of 71 participants, a critical advantage over traditional fMRI studies in this area, which often suffer from small sample sizes that can limit generalizability. By addressing these limitations, the robustness of our findings offers a more reliable foundation for future research on the neural underpinnings of object-context congruence.

Acknowledgments and Funding Support

The National Institutes of Health supported this study through grants (R01-DA056223, R01-DA059877, R01-DA059542, P20-GM144641, R01-MH118013, and R01-MH116782). The funders had no role in the study design, collection, analysis, or interpretation of data, nor did they influence the writing of the report or the decision to submit this work for publication.

Data availability

Data will be made available on request.

References

- Allegretti E, D'Innocenzo G, Coco MI, 2025. The Visual Integration of Semantic and Spatial Information of Objects in Naturalistic Scenes (VISIONS) database: attentional, conceptual, and perceptual norms. *Behav. Res. Methods* 57 (1). 10.3758/s13428-024-02535-9.
- Aminoff E, Gronau N, Bar M, 2007. The parahippocampal cortex mediates spatial and nonspatial associations. *Cerebral Cortex* (New York, N.Y.: 1991) 17 (7), 1493–1503. 10.1093/cercor/bhl078. [PubMed: 16990438]
- Ashburner J, 2007. A fast diffeomorphic image registration algorithm. *Neuroimage* 38 (1), 95–113. 10.1016/j.neuroimage.2007.07.007. [PubMed: 17761438]

- Ashburner J, Friston KJ, 2005. Unified segmentation. *Neuroimage* 26 (3), 839–851. 10.1016/j.neuroimage.2005.02.018. [PubMed: 15955494]
- Bar M, Ullman S, 1996. Spatial context in recognition. *Perception* 25 (3), 343–352. 10.1068/p250343. [PubMed: 8804097]
- Bar M, 2004. Visual objects in context. *Nat. Rev. Neurosci* 5 (8), 617–629. 10.1038/nrn1476. [PubMed: 15263892]
- Bar M, Aminoff E, 2003. Cortical analysis of visual context. *Neuron* 38 (2), 347–358. 10.1016/s0896-6273(03)00167-3. [PubMed: 12718867]
- Biederman I, 1972. Perceiving real-world scenes. *Science* (1979) 177 (4043), 77–80. 10.1126/science.177.4043.77.
- Biederman I, Mezzanotte RJ, Rabinowitz JC, 1982. Scene perception: detecting and judging objects undergoing relational violations. *Cogn. Psychol* 14 (2), 143–177. 10.1016/0010-0285(82)90007-x. [PubMed: 7083801]
- Bonitz VS, Gordon RD, 2008. Attention to smoking related and incongruous objects during scene viewing. *Acta Psychol* 129, 255–263. 10.1016/j.actpsy.2008.08.006.
- Borges MT, Fernandes EG, Coco MI, 2020. Age related differences during visual search: the role of contextual expectations and cognitive control mechanisms. *Neuropsychology, and Cognition* 10.1080/13825585.2019.1632256.
- Boyce SJ, Pollatsek A, 1992. Identification of objects in scenes: the role of scene background in object naming. *J. Exp. Psychol. Learn. Mem. Cogn* 18 (3), 531–543. 10.1037//0278-7393.18.3.531. [PubMed: 1534354]
- Brandman T, Peelen MV, 2017. Interaction between scene and object processing revealed by human fMRI and MEG decoding. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 37 (32), 7700–7710. 10.1523/jneurosci.0582-17.2017. [PubMed: 28687603]
- Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR, 2013. Power failure: why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci* 14 (5), 365–376. 10.1038/nrn3475. [PubMed: 23571845]
- Calhoun VD, Wager TD, Krishnan A, Rosch KS, Seymour KE, Nebel MB, Mostofsky SH, Nyalakanai P, Kiehl K, 2017. The impact of T1 versus EPI spatial normalization templates for fMRI data analyses. *Hum. Brain Mapp* 38 (11), 5331–5342. 10.1002/hbm.23737. [PubMed: 28745021]
- Caplette L, Gosselin F, Mermillod M, Wicker B, 2020. Real-world expectations and their affective value modulate object processing. *Neuroimage* 213 (116736), 116736. 10.1016/j.neuroimage.2020.116736.
- Cavanna AE, Trimble MR, 2006. The precuneus: a review of its functional anatomy and behavioural correlates. *Brain: A Journal of Neurology* 129 (Pt 3), 564–583. 10.1093/brain/awl004. [PubMed: 16399806]
- Chun MM, 2000. Contextual cueing of visual attention. *Trends Cogn. Sci. (Regul. Ed.)* 4 (5), 170–178. 10.1016/s1364-6613(00)01476-5.
- Coco MI, Araujo S, Petersson KM, 2017. Disentangling stimulus plausibility and contextual congruency: electro-physiological evidence for differential cognitive dynamics. *Neuropsychologia* 96, 150–163. 10.1016/j.neuropsychologia.2016.12.008. [PubMed: 27965026]
- Coco MI, Malcom GL, Keller F, 2014. The interplay of bottom-up and top-down mechanisms in visual guidance during object naming. *Q J Exp* 67, 1096–1120. 10.1080/17470218.2013.844843.
- Coco MI, Nuthmann A, Dimigen O, 2020. Fixation-related brain potentials during semantic integration of object-scene information. *J. Cogn. Neurosci* 32 (4), 571–589. 10.1162/jocn_a_01504. [PubMed: 31765602]
- Corbetta M, Patel G, Shulman GL, 2008. The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58 (3), 306–324. 10.1016/j.neuron.2008.04.017. [PubMed: 18466742]
- Corbetta M, Shulman GL, 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci* 3 (3), 201–215. 10.1038/nrn755. [PubMed: 11994752]
- Davenport JL, Potter MC, 2004. Scene consistency in object and background perception. *Psychol. Sci* 15 (8), 559–564. 10.1111/j.0956-7976.2004.00719.x. [PubMed: 15271002]

- De Graef P, Christiaens D, d'Ydewalle G, 1990. Perceptual effects of scene context on object identification. *Psychol. Res* 52, 317–329. 10.1007/BF00868064. [PubMed: 2287695]
- Demiral SB, Malcolm GL, Henderson JM, 2012. ERP correlates of spatially incongruent object identification during scene viewing: contextual expectancy versus simultaneous processing. *Neuropsychologia* 50 (7), 1271–1285. 10.1016/j.neuropsychologia.2012.02.011. [PubMed: 22391475]
- D'Innocenzo G, Della Sala S, Coco MI, 2022. Similar mechanisms of temporary bindings for identity and location of objects in healthy ageing: an eye-tracking study with naturalistic scenes. *Sci. Rep* 12 (1), 11163. 10.1038/s41598-022-13559-6.
- Fenske MJ, Aminoff E, Gronau N, Bar M, 2006. Top down facilitation of visual object recognition: object-based and context-based contributions. *Prog. Brain Res* 155, 3–21. 10.1016/S0079-6123(06)55001-0. [PubMed: 17027376]
- Ganis G, Kutas M, 2003. An electrophysiological study of scene effects on object identification. *Brain Res. Cogn. Brain Res* 16 (2), 123–144. 10.1016/S0926-6410(02)00244-6. [PubMed: 12668221]
- Goh JOS, Siong SC, Park D, Gutchess A, Hebrank A, Chee MWL, 2004. Cortical areas involved in object, background, and object-background processing revealed with functional magnetic resonance adaptation. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 24 (45), 10223–10228. 10.1523/JNEUROSCI.3373-04.2004. [PubMed: 15537894]
- Grosbras MH, Laird AR, Paus T, 2005. Cortical regions involved in eye movements, shifts of attention, and gaze perception. *Hum. Brain Mapp* 25 (1), 140–154. 10.1002/hbm.20145. [PubMed: 15846814]
- Hamm JP, Johnson BW, Kirk IJ, 2002. Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology* 113 (8), 1339–1350. 10.1016/S1388-2457(02)00161-X. [PubMed: 12140015]
- Henderson JM, Hollingworth A, 1999. High-level scene perception. *Annu. Rev. Psychol* 50 (1), 243–271. 10.1146/annurev.psych.50.1.243. [PubMed: 10074679]
- Henderson John M., Weeks PA, Hollingworth A, 1999. The effects of semantic consistency on eye movements during complex scene viewing. *J. Exp. Psychol. Hum. Percept. Perform* 25 (1), 210–228. 10.1037/0096-1523.25.1.210.
- Hollingworth A, Henderson JM, 1998. Does consistent scene context facilitate object perception? *J. Exp. Psychol. Gen* 127 (4), 398–415. 10.1037/0096-3445.127.4.398. [PubMed: 9857494]
- Jenkins LJ, Yang YJ, Goh J, Hong YY, Park DC, 2010. Cultural differences in the lateral occipital complex while viewing incongruent scenes. *Soc. Cogn. Affect. Neurosci* 5 (2–3), 236–241. 10.1093/scan/nsp056. [PubMed: 20083532]
- Kaiser D, Quek GL, Cichy RM, Peelen MV, 2019. Object vision in a structured world. *Trends Cogn. Sci. (Regul. Ed.)* 23 (8), 672–685. 10.1016/j.tics.2019.04.013.
- LaPointe MRP, Milliken B, 2016. Semantically incongruent objects attract eye gaze when viewing scenes for change. *Vis. cogn* 24 (1), 63–77. 10.1080/13506285.2016.1185070.
- Li C, Ficco L, Trapp S, Rostalski SM, Korn L, Kovács G, 2023. The effect of context congruency on fMRI repetition suppression for objects. *Neuropsychologia* 188 (108603), 108603. 10.1016/j.neuropsychologia.2023.108603.
- Loftus GR, Mackworth NH, 1978. Cognitive determinants of fixation location during picture viewing. *J. Exp. Psychol. Hum. Percept. Perform* 4 (4), 565–572. 10.1037/0096-1523.4.4.565. [PubMed: 722248]
- McAndrews MP, Girard TA, Wilkins LK, McCormick C, 2016. Semantic congruence affects hippocampal response to repetition of visual associations. *Neuropsychologia* 90, 235–242. 10.1016/j.neuropsychologia.2016.07.026. [PubMed: 27449709]
- McLean D, Nuthmann A, Renoult L, Malcolm GL, 2023. Expectation-based gist facilitation: rapid scene understanding and the role of top-down information. *J. Exp. Psychol. Gen* 152 (7), 1907–1936. 10.1037/xge0001363. [PubMed: 37126050]
- Menon V, Uddin LQ, 2010. Saliency, switching, attention, and control: a network model of insula function. *Brain Structure and Function* 214 (5–6), 655–667. 10.1007/s00429-010-0262-0. [PubMed: 20512370]

- Mudrik L, Shalgi S, Lamy D, Deouell LY, 2014. Synchronous contextual irregularities affect early scene processing: replication and extension. *Neuropsychologia* 56, 447–458. 10.1016/j.neuropsychologia.2014.02.020. [PubMed: 24593900]
- Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H, 2005. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment: moca: a brief screening tool for MCI. *J. Am. Geriatr. Soc* 53 (4), 695–699. 10.1111/j.1532-5415.2005.53221.x. [PubMed: 15817019]
- Oliva A, Torralba A, 2007. The role of context in object recognition. *Trends Cogn. Sci. (Regul. Ed.)* 11 (12), 520–527. 10.1016/j.tics.2007.09.009.
- Penny W, Holmes A, 2004. Random-effects analysis. *Human Brain Function: Second Edition*. Elsevier Inc, pp. 843–850. 10.1016/B978-012264841-0/50044-5.
- Pezdek K, Whetstone T, Reynolds K, Askari N, Dougherty T, 1989. Memory for real-world scenes: the role of consistency with schema expectation. *J. Exp. Psychol. Learn. Mem. Cogn* 15 (4), 587–595. 10.1037/0278-7393.15.4.587.
- Pinker S, 2007. *The Stuff of Thought: Language as a Window into Human Nature*. Penguin Books.
- Poldrack RA, Mumford JA, Nichols TE, 2011. *Handbook of Functional MRI Data Analysis*. Cambridge University Press.
- Power JD, Mitra A, Laumann TO, Snyder AZ, Schlaggar BL, Petersen SE, 2014. Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage* 84, 320–341. 10.1016/j.neuroimage.2013.08.048. [PubMed: 23994314]
- Rémy F, Saint-Aubert L, Bacon-Macé N, Vayssière N, Barbeau E, Fabre-Thorpe M, 2013. Object recognition in congruent and incongruent natural scenes: a life-span study. *Vision. Res* 91, 36–44. 10.1016/j.visres.2013.07.006. [PubMed: 23891714]
- Rémy F, Vayssière N, Pins D, Boucart M, Fabre-Thorpe M, 2014. Incongruent object/context relationships in visual scenes: where are they processed in the brain? *Brain Cogn* 84 (1), 34–43. 10.1016/j.bandc.2013.10.008. [PubMed: 24280445]
- Rémy F, Vayssière N, Saint-Aubert L, Bacon-Macé N, Pariente J, Barbeau E, Fabre-Thorpe M, 2020. Age effects on the neural processing of object-context associations in briefly flashed natural scenes. *Neuropsychologia* 136 (107264), 107264. 10.1016/j.neuropsychologia.2019.107264.
- Santangelo V, 2015. Forced to remember: when memory is biased by salient information. *Behav. Brain Res* 283, 1–10. 10.1016/j.bbr.2015.01.013. [PubMed: 25595422]
- Santangelo V, Di Francesco SA, Mastroberardino S, Macaluso E, 2015. Parietal cortex integrates contextual and saliency signals during the encoding of natural scenes in working memory: encoding of Natural Scenes in Working Memory. *Hum. Brain Mapp* 36 (12), 5003–5017. 10.1002/hbm.22984. [PubMed: 26333392]
- Silva MM, Groeger JA, Bradshaw MF, 2006. Attention-memory interactions in scene perception. *Spat. Vis* 19 (1), 9–19. 10.1163/156856806775009223. [PubMed: 16411480]
- Spotorno S, Tatler BW, Faure S, 2013. Semantic versus perceptual salience in visual scenes: findings from change detection. *Acta Psychol* 142 (2), 168–176. 10.1016/j.actpsy.2012.12.009.
- Sridharan D, Levitin DJ, Menon V, 2008. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proc. Natl. Acad. Sci. U.S.A* 105 (34), 12569–12574. 10.1073/pnas.0800005105. [PubMed: 18723676]
- Stirk JA, Underwood G, 2007. Low-level visual saliency does not predict change detection in natural scenes. *J. Vis* 7 (10), 1–10. 10.1167/7.10.3, 3.
- Turner BO, Paul EJ, Miller MB, Barbey AK, 2018. Small sample sizes reduce the replicability of task-based fMRI studies. *Commun. Biol* 1 (1), 62. 10.1038/s42003-018-0073-z. [PubMed: 30271944]
- Underwood G, Foulsham T, van Loon E, Humphreys L, Bloyce J, 2006. Eye movements during scene inspection: a test of the saliency map hypothesis. *Eur. J. Cogn. Psychol* 18 (3), 321–342. 10.1080/09541440500236661.
- Underwood G, Templeman E, Lamming L, Foulsham T, 2008. Is attention necessary for object identification? Evidence from eye movements during the inspection of real-world scenes. *Conscious. Cogn* 17 (1), 159–170. 10.1016/j.concog.2006.11.008. [PubMed: 17222564]
- Võ MLH, Wolfe JM, 2013. Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychol. Sci* 24 (9), 1816–1823. 10.1177/0956797613476955. [PubMed: 23842954]

- Whitfield-Gabrieli S, Nieto-Castanon A, 2012. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. *Brain Connect* 2 (3), 125–141. 10.1089/brain.2012.0073. [PubMed: 22642651]
- Wilson TW, Heinrichs-Graham E, Proskovec AL, McDermott TJ, 2016. Neuroimaging with magnetoencephalography: a dynamic view of brain pathophysiology. *Translational Research: The Journal of Laboratory and Clinical Medicine* 175, 17–36. 10.1016/j.trsl.2016.01.007. [PubMed: 26874219]
- Wolfe JM, 1994. Guided search 2.0. A revised model of visual search. *Psychon. Bull. Rev* 1, 202–238. 10.3758/BF03200774. [PubMed: 24203471]
- Wolfe JM, Horowitz TS, 2004. What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci* 5 (6), 495–501. 10.1038/nrn1411. [PubMed: 15152199]
- Wolfe JM, Võ MLH, Evans KK, Greene MR, 2011. Visual search in scenes involves selective and nonselective pathways. *Trends Cogn. Sci. (Regul. Ed.)* 15 (2), 77–84. 10.1016/j.tics.2010.12.001.

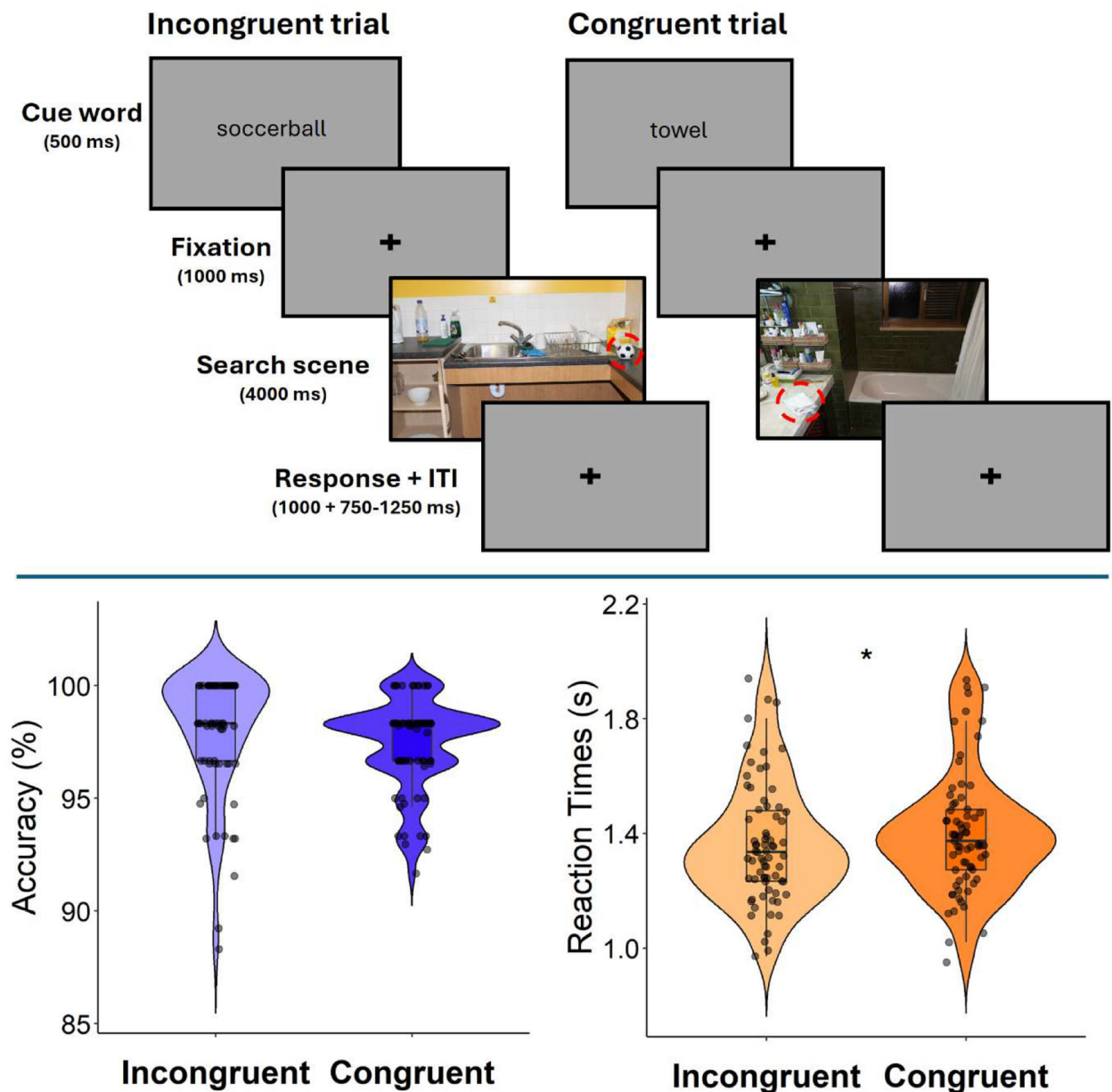


Fig. 1. Task Design & Behavioral Results.

(*Top panel*): Each trial began with a 500 ms cue word indicating the target item, followed by a 1000 ms fixation cross. The target object was then presented within an indoor scene for 4000 ms, during which participants freely searched for the target. Participants responded as to the side of the target object (left or right) using a button pad and responses were recorded during the scene presentation and within 1000 ms after its offset. A variable inter-trial interval of 750 to 1250 ms followed. “Congruent trials” featured objects consistent with the scene, while “incongruent trials” featured objects inconsistent with the scene. (*Bottom panel*): While the accuracy data did not show any conditional differences, there was a significant reaction time (RT) difference, with participants being slower during

congruent compared to incongruent trials. Asterisks indicate significant differences between conditions. $*p = .001$.

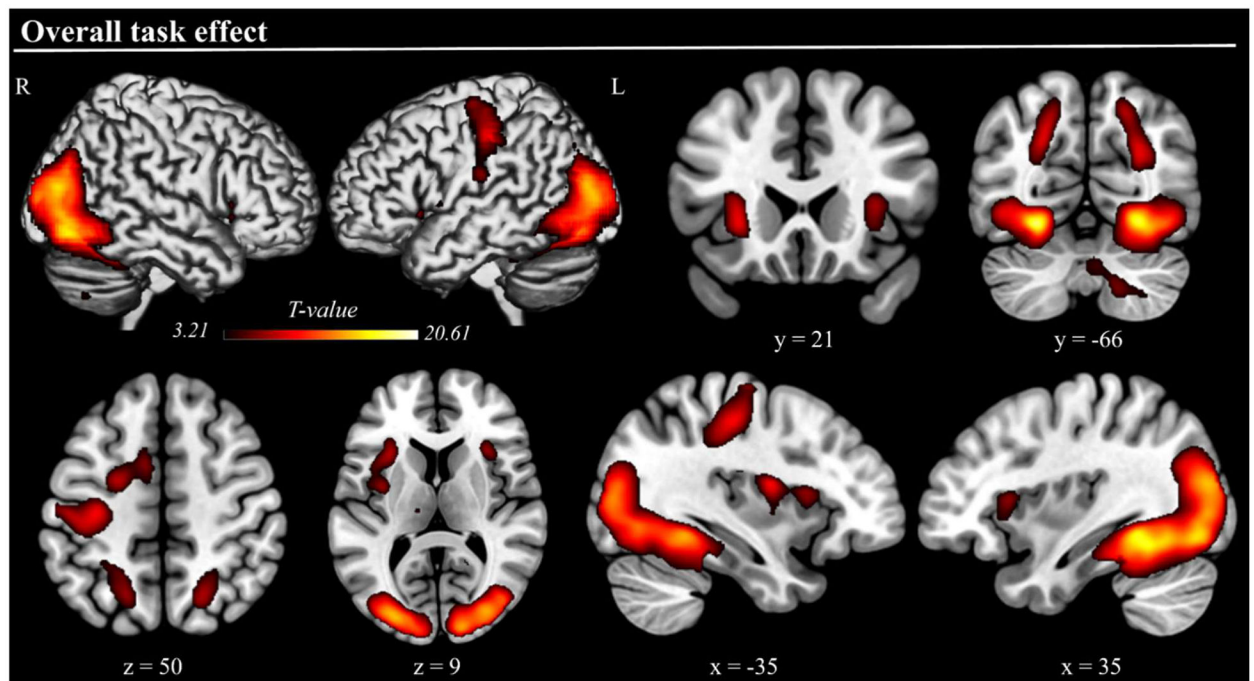


Fig. 2. Overall task effect.

Three-dimensional rendered projections and cross-sections showing dorsal frontoparietal activations, as well as activations in the bilateral anterior insula, occipital visual cortices, and other regions, elicited by the overall effect of task execution. The task execution contrast averaged activity across conditions (correct trials only). Results are displayed at a cluster-level corrected p -FWE = .05, with a minimum cluster size of 200 used for visualization only.

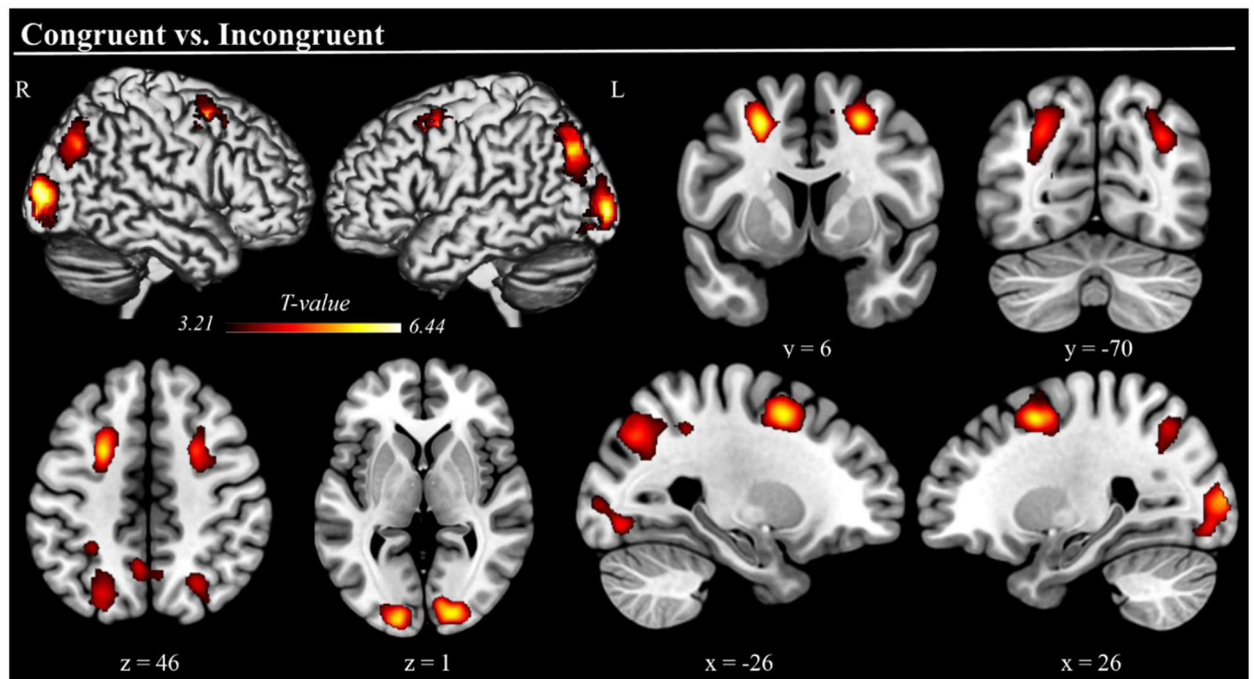


Fig. 3. Brain Regions Exhibiting Stronger Activation in the Congruent Condition.

Three-dimensional rendered projections and cross-sections showing dorsal frontoparietal (FEF and IPS bilaterally), precuneus, and visual occipital cortical regions where activation was stronger during congruent compared to incongruent trials. No brain regions exhibited stronger activation during incongruent trials. Results are displayed at a cluster-level corrected p -FWE = .05, with a minimum cluster size of 200 used for visualization only.

Table 1
MNI coordinates and statistical values for task and conditional effects.

Contrast	Region	Cluster		Peak	
		<i>p-FWE-corr</i>	<i>k</i>	<i>t-value</i>	<i>x, y, z</i>
<i>Overall task effect</i>	R Fusiform Gyrus	<.001	21,699	20.55	30, -48, -10
	L Fusiform Gyrus			19.63	-30, -50, -10
	R Inferior occipital gyrus			15.06	40, -78, -6
	L Inferior occipital gyrus			14.75	-28, -88, 10
	R Middle occipital gyrus			15.52	32, -84, 16
	L Middle occipital gyrus			14.70	-30, -84, 14
	R Intraparietal sulcus			7.77	22, -64, 48
	L Intraparietal sulcus			5.60	-26, -48, 46
	L Precentral gyrus			9.94	-34, -22, 48
	L Anterior insula			9.78	-28, 22, 0
	L Frontal eye field			7.94	-22, -4, 50
	R Anterior insula			7.33	32, 20, 8
	R Cerebellum			7.62	16, -60, -44
<i>Congruent vs. Incongruent</i>	R Calcarine cortex	<.001	1237	6.44	12, -86, -4
	L Calcarine cortex			5.65	-14, -92, -2
	R Middle occipital gyrus			4.63	32, -70, 36
	R Intraparietal sulcus			4.26	26, -66, 46
	R Precuneus			3.68	6, -60, 46
	L Precuneus			4.21	-4, -58, 46
	L Middle occipital gyrus			5.05	-30, -78, 38
	L Intraparietal sulcus			4.36	-22, -62, 40
	R Frontal eye field			5.68	26, 6, 50
	L Frontal eye field			6.02	-24, 4, 50

L/R: left/right hemisphere; *k*: cluster size; coordinates are in MNI space and reflect peak voxel.