



HHS Public Access

Author manuscript

Trends Cogn Sci. Author manuscript; available in PMC 2016 June 15.

Published in final edited form as:

Trends Cogn Sci. 2014 May ; 18(5): 242–250. doi:10.1016/j.tics.2014.02.004.

Attention in the real world: toward understanding its neural basis

Marius V. Peelen¹ and Sabine Kastner^{2,3}

¹ Center for Mind/Brain Sciences (CIMEC), University of Trento, Corso Bettini 31, 38068 Rovereto (TN), Italy

² Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA

³ Department of Psychology, Princeton University, Princeton, NJ 08544, USA

Abstract

The efficient selection of behaviorally relevant objects from cluttered environments supports our everyday goals. Attentional selection has typically been studied in search tasks involving artificial and simplified displays. Although these studies have revealed important basic principles of attention, they do not explain how the brain efficiently selects familiar objects in complex and meaningful real-world scenes. Findings from recent neuroimaging studies indicate that real-world search is mediated by ‘what’ and ‘where’ attentional templates that are implemented in high-level visual cortex. These templates represent target-diagnostic properties and likely target locations, respectively, and are shaped by object familiarity, scene context, and memory. We propose a framework for real-world search that incorporates these recent findings and specifies directions for future study.

Keywords

vision; attention; object perception; visual search; natural scenes

‘It is not true that “the laboratory can never be like life.” The laboratory must be like life!’ [1].

Attentional selection in daily life

The primary goal of selective visual attention is to focus processing resources on behaviorally relevant objects in our visual environment. In our daily lives, we direct attention (and, with it, often our eyes) all the time: when searching for a coffee cup in the cupboard, when looking out for cars while crossing the street, or when trying to find a friend at a conference. Because we perform so many visual searches every day, often for the same objects (e.g., people) and often within the same types of environments (e.g., city streets, living rooms), we are highly experienced in tasks like these. This experience is reflected in exceptionally efficient behavior. For example, most observers will need only a single glance

at scenes like the ones shown in Figure 1A to decide whether objects like people, cars, trees, or houses are present. The efficiency of extracting behaviorally relevant categorical information from real-world scenes is remarkable given the complexity of the task; in daily life, scenes contain many distracter objects that share the visual features of target objects, including their color, size, and orientation. Furthermore, the visual appearances of both target and distracter objects vary greatly across scenes and viewing conditions. How does the brain solve this task?

Classical approaches to the study of visual search

To study the process of selecting behaviorally relevant information from cluttered displays in the laboratory, behavioral and neurophysiological studies have typically greatly reduced the complexity of the real world by having participants perform search tasks in displays comprising simple and well-defined stimuli presented on uniform backgrounds (Figure 1B). For example, in a display comprising red and green lines of different orientation, observers search for the presence of a specific target stimulus defined by a single feature (e.g., red) or a specific combination of features (e.g., a red horizontal line).

Behavioral studies

Using this approach, behavioral studies have measured performance in many different kinds of search [2]. Search efficiency has typically been defined as the relationship between reaction time (RT) and the number of distracter items, or set size. A major goal of several influential attention theories has been to explain why some searches are more efficient than others.

One influential theory, feature-integration theory (FIT) [3], distinguishes between feature and conjunction search. In this account, single features (e.g., red) can be detected in parallel across a visual display, thereby resulting in highly efficient search when the target contains a unique feature (e.g., a red shape in a display of green shapes). By contrast, when targets are defined by a combination of two or more features (conjunction search), focused attention is needed to bind features, leading to inefficient search that reflects serial deployment of attention to the items in the display. Guided-search theory [4] provides an important extension to FIT, proposing that the serial-search process is guided by initial parallel processing of features, thereby accounting for the finding that some conjunction searches are highly efficient [4]. One prominent alternative theory is the attentional engagement theory [5], which postulates that search difficulty is primarily related to the visual similarity of the target to the distracters and to the dissimilarity of the distracters to each other (with more efficient search for displays with homogeneous distracter sets). These and other theories of visual search, based on behavioral findings in artificial multi-element search displays, have provided the foundation for investigations at the neural level.

Neurophysiological studies

Important insights into the neural basis of attentional selection come from monkey electrophysiology studies. These studies have provided evidence for various effects on neural responses when a monkey attends to a visual stimulus relative to when the same

stimulus is ignored: neural responses are enhanced [6] and more reliable [7] and noise correlations in local populations decrease [8]. Together, these effects suggest that attention operates to control neural gain to enhance the neural representation of the attended stimulus. Furthermore, attention increases synchronous firing [9] and this may be a mechanism by which information transfer within an area and across interconnected areas becomes facilitated ([10–12], but see [13]). However, much of this work has focused on the effects of attention on simplified target stimuli presented in isolation, thereby drastically reducing the complexity of naturalistic selection conditions.

Seminal studies probing the interactions of multiple nearby stimuli have shown that two stimuli presented in a neuron's receptive field (RF) interact in mutually suppressive ways such that the neural response reflects a weighted average of the individual responses [6]. Such interactions have been interpreted as a neural competition for limited processing resources at the level of the RF and immediately surrounding regions [14]. Importantly, directing attention to one of two competing stimuli biases neural responses in favor of the attended stimulus, thereby effectively filtering out a nearby distracter stimulus and resolving the neural competition [6]. Such attentional biases operate not only on spatial locations, but also on attended features. Neurons tuned for featural content, such as a specific color or direction of motion, show enhanced responses when the preferred feature is selected for further processing [15]. Whereas spatial attention effects are selective for the attended location, featural effects operate globally across the visual field [16], suggesting that the underlying neural mechanisms for feature- and space-based selections differ [15]. In visual search, spatial and featural attentional-biasing signals interact. When monkeys search for features (e.g., a red circle) in a multi-item array, two processes operate that independently boost neural responses: a parallel process that uses featural information to mark conspicuous stimuli in the search array, thereby reducing the search space, and a serial process that examines the items with spatial specificity to identify matching target stimuli [17].

Importantly, both spatial and featural attentional biases affect neural responses not only in the presence of visual stimuli, and thus during the processing of information, but also in the absence of visual stimulation [18–21], as evidenced by increases in neural baseline responses. For example, neurons in the inferotemporal cortex and prefrontal cortex (PFC) show increased activity during a delay period when a target stimulus has been removed from the visual scene. This feature-specific delay activity is an example of a neural correlate for a 'search template' that aids an upcoming visual search, biasing competition in favor of stimuli matching the template [18]. We elaborate on the concept of attentional templates below.

In sum, neurophysiological studies investigating the neural basis of attention using simplified displays have shown that attention affects neural processing in several ways. When multiple stimuli compete for representation, spatial and featural attention bias competition in favor of task-relevant stimuli. Finally, these top-down biasing signals are instantiated before the onset of a search array. Although we have focused here on findings from monkey physiology, many of the reviewed findings have been confirmed in human neuroimaging studies [22,23].

The next frontier: understanding the neural basis of real-world visual search

In daily life, we select meaningful objects from meaningful scenes. Indeed, we usually do not direct attention to an empty region in space and we rarely decide to detect simple features such as horizontal lines or upward motion. Thus, although studies using simplified displays have been fundamental for our understanding of basic attention mechanisms, their results are not readily applicable to real-world scenarios. For example, what would be the behavioral prediction for detecting people in the scenes of Figure 1A? These scenes contain dozens of distracter objects, there is not one feature that uniquely defines the target, and the distracter set is highly heterogeneous. Classical attention theories would predict an inefficient search, with search times of perhaps several seconds. But clearly it is not; we can detect the person in a single glance. A growing body of literature confirms this intuition, showing that naturalistic visual search is surprisingly efficient (Figure 2).

There are several important ways in which real-world search differs from search in artificial arrays. First, target objects in daily life are typically familiar objects (e.g., ‘my car’) or object categories (e.g., ‘cars’) rather than simple features, colors, or shapes. Familiarity fundamentally alters object representations in the visual system and influences the efficiency of visual search (Box 1). Second, the arrangement of distracters in scenes contains a high degree of regularity; we expect chairs around a table or a mirror above a bathroom sink. Familiarity with the identity and arrangement of distracter sets facilitates visual search (Box 1). Third, a real-world scene gives a powerful context that provides information about likely target locations (Box 2). Scene layout also provides many constraints about the visual properties (e.g., retinal size) of the target object at different locations in the scene (Box 2). Finally, previous experience with specific scenes (e.g., ‘my kitchen’) also facilitates search [24].

Box 1

The role of visual experience in search

Visual search in daily life typically involves objects and environments that we have experienced many times in the past and with which we are thus highly familiar. Visual experience fundamentally shapes the visual system [70]. These effects are particularly pronounced in early stages of development, but visual experience continues to shape the visual cortex throughout the lifespan, as reflected in perceptual learning [71,72]. Understanding the mechanisms of real-world visual search therefore requires taking into account the role of visual experience on perceptual processing and attentional selection. There are at least two ways in which visual experience influences visual search.

First, visual search in daily life usually involves target and distracter objects that are familiar and meaningful to us, such as when we search for people in a street scene also containing trees, houses, and cars. Behavioral studies have shown that visual search is better when target and/or distracter objects are familiar than when they are not [66,73–75]. At the neural level, it has been shown that extensive visual experience alters the representations of objects in high-level visual cortex [72], leading to changes in

distributed activity patterns [76] and in the selectivity of single neurons [77,78]. Furthermore, neural representations of highly familiar object categories such as animals and vehicles may be activated based on rapid feedforward processing [67,79,80], perhaps reflecting a hierarchical convergence of multiple simple visual attributes onto single neurons representing complex category-diagnostic shapes [81,82]. Such category-selective neurons may support real-world search. For example, the body-selective extrastriate body area [83] has been causally implicated in search for people (but not cars) in real-world scenes [84]. Furthermore, visual-search templates may directly modulate category-diagnostic shape representations to support parallel search for familiar object categories [27,28].

Second, repeatedly performing specific search tasks facilitates search performance [34], as for example observed in radiologists [85]. This is due in part to the learning of features that optimally distinguish the target from the distracter, as reflected in efficient neural representations of target-diagnostic features after training [86]. Additionally, targets are easier to find when the identity or spatial arrangement of the distracter set is familiar [87]. This ‘contextual cueing’ is particularly relevant for real-world search, because objects in the real world typically occur in highly regular contexts (Box 2).

In sum, mechanisms supporting real-world visual search are greatly influenced by visual experience. The neural representation of objects changes as a function of experience, which affects how top-down attention modulates neural activity during search for these objects. Additionally, experience in performing specific searches improves search performance due to optimized strategies and implicit learning of distracter sets.

Mechanisms involved in visual search have developed and evolved to function optimally in real-world environments and can be assumed to make optimal use of scene-based constraints and real-world regularities; a lifetime of experience in searching for objects in scenes must have fundamentally shaped neural mechanisms of visual search. To understand fully neural mechanisms of visual search therefore requires studying search in naturalistic situations.

Neural basis of category-level search in real-world scenes

Recent neuroimaging studies have started to investigate the neural basis of visual search in real-world scenes [25–32]. In a series of studies [27,28,30], functional MRI (fMRI) activity was measured while participants detected the presence of objects from a cued category (cars, people) in briefly presented photographs of outdoor scenes like those in Figure 1A. The presented scenes were new to the participants, contained a large number of distracter objects, varied greatly in terms of visual characteristics, were presented briefly, and were followed by visual masks. Importantly, participants did not know the location and visual properties of the target object before the scene was presented; a scene could show a person sitting on a bench but also a group of people walking on a sidewalk (Figure 1). Despite these complexities, participants needed only a single glance at the scene (around 100 ms [33]) to perform this task successfully.

It was found that distributed fMRI activity patterns in object-selective visual cortex (OSC) contained information about object categories that were present in the scene. Importantly, however, this information was strongly modulated by task relevance, with highly reliable information for objects belonging to the target category but much weaker information for objects of other categories [27,28,30]. Interestingly, in addition to the enhanced processing of task-relevant objects, one study showed that the processing of previously (but not currently) relevant objects – known to be particularly distracting in visual search [34] – was actively suppressed relative to never-relevant objects (Figure 3A,B). Another study measured fMRI responses while participants detected people or vehicles in a series of short movie segments [25]. Responses in many parts of the brain increased when the target category, or a semantically similar category, appeared in the movie, suggesting that category-based attention may have widespread influences on brain activity. Together, these results provide neural evidence for an attentional-selection mechanism that biases the processing of scenes at the category level to mediate real-world search.

Box 2

Scene context constrains search

Knowledge about where we are in the world can be quickly established based on various cues such as diagnostic objects in a scene (e.g., stoves are found in kitchens, not bedrooms; [88,89]) or scene layout inferred from global image statistics or ‘gist’ [90]. Scene context has been shown to play an important facilitatory role in visual search and object recognition, with several related factors playing a role. First, scene context creates expectations about likely locations of target objects based on semantic relations [88]; we start looking for a pen on top of a desk, not on the floor. Scene context also provides information about physically impossible target locations, such as a pen floating in the sky [88]. Together, these factors guide search and greatly reduce the effective search space [91,92]. Second, scene context provides information about the visual characteristics (e.g., size) of target and distracter objects at different locations (e.g., depths) within the scene [69]. Third, scene context informs the content of the search template by specifying the distracter set. For example, looking for a person in a forest versus a beach may generate templates that consider the differences in distracter objects. Finally, context also directly facilitates object recognition by creating expectations that help to disambiguate otherwise ambiguous shapes; for example, a road context creates the expectation that an approaching object in the distance is likely a car [47,93].

Intriguingly, category-based attentional biases were found to be spatially global, biasing not just the processing of spatially attended scenes but also the processing of spatially unattended (and totally task-irrelevant) scenes [27]. As reviewed earlier, such obligatory spatially global attention effects have also been observed for attention to simple features like color, orientation, and motion direction [15]. These findings therefore imply an interesting similarity between attention to familiar object categories in scenes and attention to simple features in artificial displays. They are consistent with behavioral findings of rapid object-category detection at spatially unattended locations [35,36]. One explanation for these findings is that our extraordinary visual experience with object categories like cars and

people has resulted in representations of these categories that can be activated using feedforward processing (Box 1), similar to that shown for processing of simple features [37]. Attention might then target category and feature representations in similar ways, despite the fact that these representations are located in different parts of the visual system [38].

An important question concerns the mechanisms that lead to the attentional biases reviewed above. Specifically, are these biases the result of baseline increases that prime representations of task-relevant objects, similar to the spatial- and featural-biasing signals that were reported in studies using simple stimuli, reviewed above? Or do they reflect feedback signals occurring after the scene has been presented and processed? This issue was addressed in an fMRI study in which participants were cued, trial by trial, to detect people or cars in briefly presented scenes [28]. Critically, in a proportion of trials, the symbolic category cue was presented without a subsequent scene to isolate internally driven attention-related fMRI activity (Figure 3C). Results showed that the degree of category selectivity of preparatory activity in high-level visual cortex was strongly negatively correlated with response speed (Figure 3D), whereas a positive correlation with response speed was found for preparatory activity in low-level visual cortex. These results indicate that content-specific attentional templates guide search in scenes and that some templates are more effective than others in doing so (Box 3).

Together, these studies show that the remarkable efficiency of naturalistic category-level search is linked to sophisticated attentional-selection mechanisms that are instantiated in higher levels of visual cortex. A lifetime of experience has enabled us to process rapidly category-diagnostic features in real-world scenes (Box 1). These representations can be effectively targeted by attention, leading to both enhancement of relevant and suppression of irrelevant information. Internally generated category-selective neural activity provides a competitive advantage for objects belonging to the target category, biasing the processing of scenes in parallel across the visual field in favor of the attended category and away from unattended categories.

A framework for real-world search

The studies reviewed in the previous section have provided ample evidence that real-world search is aided by content-specific attentional templates that are implemented in object-selective cortex and carry a neural signature reminiscent of processes that are also used during the feedforward activation of object representations. We refer to these templates as the ‘what’ templates. In addition to the ‘what’ templates, recent studies have shown that spatial biases are generated during template formation that represent scene locations at which targets are expected with high probability based on scene context [29] or episodic memory [31]. We refer to these templates as ‘where’ templates. Figure 4 presents a framework that captures the influence of several variables on the formation of ‘what’ and ‘where’ templates during real-world search. Although this framework is by no means complete, it presents a number of testable hypotheses that may be useful in advancing toward a more theoretical account.

Visual search starts with a behavioral goal that specifies the object or object category that is searched for (e.g., looking for people in a park). The general visual attributes of the target are likely to be stored in systems representing object knowledge, including the ventral temporal cortex [39] and the anterior temporal lobes [40,41]. Importantly, because the ‘what’ template is modeled as an internal representation of attributes that distinguish the target from the distracters [5], it is influenced by both target and distracter properties of the current scene. Accordingly, the template is not only influenced by attributes of the current search target but also by the current scene context (Box 2) and any knowledge of the scene obtained through previous encounters; searching for people on a bridge that occludes the lower part of the body (Figure 4) requires a different template than searching for people in an open field. Scene context and episodic memory provide information about the likely distracter objects to be found in the scene as well as about the likely visual attributes of target and distracter objects (Box 2). fMRI and transcranial magnetic stimulation (TMS) studies have provided evidence that scene-selective regions, including the parahippocampal cortex, retrosplenial cortex, and transverse occipital sulcus, represent global scene properties and scene category [42–45], which may inform the ‘what’ template. In addition to regions of the scene-selective network, the medial PFC has been shown to be involved in context-based predictions regarding object properties [46,47]. Episodic scene memory is thought to involve the medial temporal lobes, particularly the hippocampus [48], and the posterior parietal cortex (PPC) [49]. In this account, the resulting ‘what’ search template is instantiated in regions of occipital and temporal cortex that are best suited to represent the specific contents of the template; for example, regions in early visual cortex for orientation-specific templates [50,51] or regions in high-level visual cortex for category-level templates [28,52].

Box 3

The search template for real-world search

It has been proposed that visual search involves the matching of visual input to an internal representation that approximates the sensory input, an attentional ‘template’ that comprises attributes distinguishing targets from non-targets [5]. When preparing to search for a specific target object defined by specific features (e.g., a particular shape), neurons that are selective for those features are activated before the onset of the search array [18]. This preparatory activity then biases processing in favor of the target object [18].

If the specific features of the target object are known in advance (e.g., red horizontal line), the optimal search template comprises these specific features [94,95]. In most real-world searches, however, the specific features of the target object are not known in advance due to moment-to-moment differences in, for example, viewpoint, lighting, viewing distance, and occlusion. This raises the question of what features real-world search templates comprise; the template must generalize across real-world viewing conditions but at the same time be diagnostic of the target relative to the many possible distracter objects that could be present in the scene. Results of behavioral studies suggest that search for familiar object categories in real-world scenes is mediated by templates comprising a collection of intermediate-level category-diagnostic features [96,97]. This

notion is supported by computational studies showing that objects are best classified by features of intermediate-level complexity, such as the wheel of a car [98].

An important aspect of real-world search is that target-diagnostic features depend on the other objects that are present in the scene. For example, an effective template for finding a person in a desert might comprise vertical features, whereas such a template would be useless when searching for a person in a forest. Thus, search templates for real-world search depend on scene context (Box 2). This context-dependent nature of the search template remains poorly understood, especially at the neural level.

‘Where’ templates are also shaped by scene context and episodic memory. Scene context facilitates search by specifying likely target locations within the scene (Box 2). Context-based spatial expectations are reflected in activity patterns in regions of the occipitotemporal cortex and PPC [29]. Finally, episodic memory-based attentional orienting has been shown to involve the hippocampus [32] and to result in spatially specific anticipatory activity in the occipitotemporal cortex [31].

Both ‘what’ and ‘where’ templates constitute internal representations that are generated based on current behavioral goals, context, and knowledge of the world. Once these templates have been generated, they need to be maintained over multiple fixations to serve a behavioral goal. The maintenance of target templates across intervening stimuli has classically been associated with the PFC [53], with important evidence coming from monkey physiology [20,21]. Recent evidence suggests that, at least in humans, the PPC may also contribute to the maintenance of ‘what’ and ‘where’ templates. The PPC is interconnected with object-selective cortex as well as the PFC and medial temporal lobe long-term memory systems [54]. In humans, the PPC has highly advanced object-vision capabilities [55], is topographically organized [56], and has been implicated in both feature-based and space-based attentional control [57,58]. Simultaneously, the human PPC is an integral part of an extended working-memory system [59]. Thus, the PPC is well positioned to maintain both ‘what’ and ‘where’ attentional search templates. Future work is needed to address the specific roles of, and interactions between, the visual cortex, PFC, and PPC in the representation and maintenance of target templates.

It is important to note that our framework does not consider several variables that will need implementation to develop a more complete account of attentional selection from natural scenes. Such factors include, but are not limited to, low-level visual saliency [60,61] and saliency due to objects’ reward history [62,63], affective properties [64], and social relevance [65].

Concluding remarks and future directions

We have argued that attentional selection under naturalistic conditions uses additional and partly different mechanisms compared with those that have been studied using artificial displays. This provides only one example indicating that a full understanding of cognition will be possible only by considering the complexity of the real world [1,66]. There are many outstanding questions (Box 4). Addressing these questions will require consideration of

findings from multiple fields, including not only classical studies on visual search and attention but also findings from studies investigating the neural basis of scene perception, object perception, perceptual learning, and memory.

Box 4

Outstanding questions

- Are ‘what’ and ‘where’ templates represented in common regions of visual and parietal cortex? How do these templates interact (Figure 4)?
- What are the temporal dynamics of the matching of visual input to ‘what’ and ‘where’ templates? Do top-down templates modulate early sensory processing [99,100]? Does this time course differ for ‘what’ and ‘where’ templates?
- How do neural mechanisms of attention overlap and interact with neural mechanisms of expectation [101]?
- To further approximate the real world, an important avenue for future research will be to study neural mechanisms of visual search in situations that include multiple modalities [102], depth, a realistic temporal structure [61], and active exploration.

References

1. Gibson, JJ. *The Ecological Approach to Visual Perception*. Houghton Mifflin; 1979.
2. Wolfe JM, Horowitz TS. What attributes guide the deployment of visual attention and how do they do it? *Nat. Rev. Neurosci.* 2004; 5:495–501. [PubMed: 15152199]
3. Treisman AM. A feature-integration theory of attention. *Cogn. Psychol.* 1980; 12:97–136. [PubMed: 7351125]
4. Wolfe JM, et al. Guided search: an alternative to the feature integration model for visual search. *J. Exp. Psychol. Hum. Percept. Perform.* 1989; 15:419–433. [PubMed: 2527952]
5. Duncan J, Humphreys GW. Visual search and stimulus similarity. *Psychol. Rev.* 1989; 96:433–458. [PubMed: 2756067]
6. Desimone R, Duncan J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* 1995; 18:193–222. [PubMed: 7605061]
7. Mitchell JF, et al. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron.* 2009; 63:879–888. [PubMed: 19778515]
8. Cohen MR, Maunsell JH. Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* 2009; 12:1594–1600. [PubMed: 19915566]
9. Fries P, et al. Modulation of oscillatory neuronal synchronization by selective visual attention. *Science.* 2001; 291:1560–1563. [PubMed: 11222864]
10. Buschman TJ, Miller EK. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices. *Science.* 2007; 315:1860–1862. [PubMed: 17395832]
11. Gregoriou GG, et al. High-frequency, long-range coupling between prefrontal and visual cortex during attention. *Science.* 2009; 324:1207–1210. [PubMed: 19478185]
12. Saalmann YB, et al. The pulvinar regulates information transmission between cortical areas based on attention demands. *Science.* 2012; 337:753–756. [PubMed: 22879517]
13. Ray S, Maunsell JH. Differences in gamma frequencies across visual cortex restrict their possible use in computation. *Neuron.* 2010; 67:885–896. [PubMed: 20826318]

14. Sundberg KA, et al. Spatial attention modulates center-surround interactions in macaque visual area v4. *Neuron*. 2009; 61:952–963. [PubMed: 19324003]
15. Maunsell JH, Treue S. Feature-based attention in visual cortex. *Trends Neurosci*. 2006; 29:317–322. [PubMed: 16697058]
16. Treue S, Martinez Trujillo JC. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*. 1999; 399:575–579. [PubMed: 10376597]
17. Bichot NP, et al. Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*. 2005; 308:529–534. [PubMed: 15845848]
18. Chelazzi L, et al. A neural basis for visual search in inferior temporal cortex. *Nature*. 1993; 363:345–347. [PubMed: 8497317]
19. Luck SJ, et al. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol*. 1997; 77:24–42. [PubMed: 9120566]
20. Miller EK, et al. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci*. 1996; 16:5154–5167. [PubMed: 8756444]
21. Rao SC, et al. Integration of what and where in the primate prefrontal cortex. *Science*. 1997; 276:821–824. [PubMed: 9115211]
22. Kastner S, Ungerleider LG. Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci*. 2000; 23:315–341. [PubMed: 10845067]
23. Siegel M, et al. Spectral fingerprints of large-scale neuronal interactions. *Nat. Rev. Neurosci*. 2012; 13:121–134. [PubMed: 22233726]
24. Vo ML, Wolfe JM. The interplay of episodic and semantic memory in guiding repeated search in scenes. *Cognition*. 2013; 126:198–212. [PubMed: 23177141]
25. Cukur T, et al. Attention during natural vision warps semantic representation across the human brain. *Nat. Neurosci*. 2013; 16:763–770. [PubMed: 23603707]
26. Guo F, et al. Feature-independent neural coding of target detection during search of natural scenes. *J. Neurosci*. 2012; 32:9499–9510. [PubMed: 22787035]
27. Peelen MV, et al. Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*. 2009; 460:94–97. [PubMed: 19506558]
28. Peelen MV, Kastner S. A neural basis for real-world visual search in human occipitotemporal cortex. *Proc. Natl. Acad. Sci. U.S.A.* 2011; 108:12125–12130. [PubMed: 21730192]
29. Preston TJ, et al. Neural representations of contextual guidance in visual search of real-world scenes. *J. Neurosci*. 2013; 33:7846–7855. [PubMed: 23637176]
30. Seidl KN, et al. Neural evidence for distracter suppression during visual search in real-world scenes. *J. Neurosci*. 2012; 32:11812–11819. [PubMed: 22915122]
31. Stokes MG, et al. Long-term memory prepares neural activity for perception. *Proc. Natl. Acad. Sci. U.S.A.* 2012; 109:E360–E367. [PubMed: 22109554]
32. Summerfield JJ, et al. Orienting attention based on long-term memory experience. *Neuron*. 2006; 49:905–916. [PubMed: 16543137]
33. Potter MC. Short-term conceptual memory for pictures. *J. Exp. Psychol. Hum. Learn.* 1976; 2:509–522. [PubMed: 1003124]
34. Shiffrin RM, Schneider W. Controlled and automatic human information processing: 2. Perceptual learning, automatic attending and a general theory. *Psychol. Rev*. 1977; 84:127–190.
35. Li FF, et al. Rapid natural scene categorization in the near absence of attention. *Proc. Natl. Acad. Sci. U.S.A.* 2002; 99:9596–9601. [PubMed: 12077298]
36. Poncet M, et al. A need for more information uptake but not focused attention to access basic-level representations. *J. Vis*. 2012; 12:15. [PubMed: 22262913]
37. Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol*. 1962; 160:106–154. [PubMed: 14449617]
38. Grill-Spector K, Malach R. The human visual cortex. *Annu. Rev. Neurosci*. 2004; 27:649–677. [PubMed: 15217346]
39. Martin A. The representation of object concepts in the brain. *Annu. Rev. Psychol*. 2007; 58:25–45. [PubMed: 16968210]

40. Patterson K, et al. Where do you know what you know? The representation of semantic knowledge in the human brain. *Nat. Rev. Neurosci.* 2007; 8:976–987. [PubMed: 18026167]
41. Peelen MV, Caramazza A. Conceptual object representations in human anterior temporal cortex. *J. Neurosci.* 2012; 32:15728–15736. [PubMed: 23136412]
42. Dilks DD, et al. The occipital place area is causally and selectively involved in scene perception. *J. Neurosci.* 2013; 33:1331–1336. [PubMed: 23345209]
43. Epstein RA. Parahippocampal and retrosplenial contributions to human spatial navigation. *Trends Cogn. Sci.* 2008; 12:388–396. [PubMed: 18760955]
44. Kravitz DJ, et al. Real-world scene representations in high-level visual cortex: it's the spaces more than the places. *J. Neurosci.* 2011; 31:7322–7333. [PubMed: 21593316]
45. Walther DB, et al. Natural scene categories revealed in distributed patterns of activity in the human brain. *J. Neurosci.* 2009; 29:10573–10581. [PubMed: 19710310]
46. Aminoff EM, et al. The role of the parahippocampal cortex in cognition. *Trends Cogn. Sci.* 2013; 17:379–390. [PubMed: 23850264]
47. Bar M. Visual objects in context. *Nat. Rev. Neurosci.* 2004; 5:617–629. [PubMed: 15263892]
48. Burgess N, et al. The human hippocampus and spatial and episodic memory. *Neuron.* 2002; 35:625–641. [PubMed: 12194864]
49. Wagner AD, et al. Parietal lobe contributions to episodic memory retrieval. *Trends Cogn. Sci.* 2005; 9:445–453. [PubMed: 16054861]
50. Harrison SA, Tong F. Decoding reveals the contents of visual working memory in early visual areas. *Nature.* 2009; 458:632–635. [PubMed: 19225460]
51. Serences JT, et al. Stimulus-specific delay activity in human primary visual cortex. *Psychol. Sci.* 2009; 20:207–214. [PubMed: 19170936]
52. Summerfield C, et al. Predictive codes for forthcoming perception in the frontal cortex. *Science.* 2006; 314:1311–1314. [PubMed: 17124325]
53. Curtis CE, D'Esposito M. Persistent activity in the prefrontal cortex during working memory. *Trends Cogn. Sci.* 2003; 7:415–423. [PubMed: 12963473]
54. Lewis JW, Van Essen DC. Corticocortical connections of visual, sensorimotor, and multimodal processing areas in the parietal lobe of the macaque monkey. *J. Comp. Neurol.* 2000; 428:112–137. [PubMed: 11058227]
55. Konen CS, Kastner S. Two hierarchically organized neural systems for object information in human visual cortex. *Nat. Neurosci.* 2008; 11:224–231. [PubMed: 18193041]
56. Silver MA, Kastner S. Topographic maps in human frontal and parietal cortex. *Trends Cogn. Sci.* 2009; 13:488–495. [PubMed: 19758835]
57. Egner T, et al. Neural integration of top-down spatial and feature-based information in visual search. *J. Neurosci.* 2008; 28:6141–6151. [PubMed: 18550756]
58. Szczepanski SM, et al. Mechanisms of spatial attention control in frontal and parietal cortex. *J. Neurosci.* 2010; 30:148–160. [PubMed: 20053897]
59. Todd JJ, Marois R. Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature.* 2004; 428:751–754. [PubMed: 15085133]
60. Itti L, Koch C. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 2000; 40:1489–1506. [PubMed: 10788654]
61. Nardo D, et al. Stimulus-driven orienting of visuo-spatial attention in complex dynamic environments. *Neuron.* 2011; 69:1015–1028. [PubMed: 21382559]
62. Chelazzi L. Visual selective attention and the effects of monetary rewards. *Psychol. Sci.* 2006; 17:222–227. [PubMed: 16507062]
63. Hickey, C.; Peelen, M. 2013 Annual Meeting Program. Vol. 181. Cognitive Neuroscience Society; 2013. Reward guides attention to object categories in real-world scenes.
64. Ohman A, et al. Emotion drives attention: detecting the snake in the grass. *J. Exp. Psychol. Gen.* 2001; 130:466–478. [PubMed: 11561921]
65. Ro T, et al. Attentional biases for faces and body parts. *Vis. Cogn.* 2007; 15:322–348.
66. Neisser, U. *Cognition and Reality.* W.H. Freeman; 1976.

67. Thorpe S, et al. Speed of processing in the human visual system. *Nature*. 1996; 381:520–522. [PubMed: 8632824]
68. Ehinger KA, et al. Modeling search for people in 900 scenes: a combined source model of eye guidance. *Vis. Cogn.* 2009; 17:945–978. [PubMed: 20011676]
69. Wolfe JM, et al. Visual search for arbitrary objects in real scenes. *Atten. Percept. Psychophys.* 2011; 73:1650–1671. [PubMed: 21671156]
70. Purves D, et al. Understanding vision in wholly empirical terms. *Proc. Natl. Acad. Sci. U.S.A.* 2011; 108(Suppl. 3):15588–15595. [PubMed: 21383192]
71. Gilbert CD, Li W. Adult visual cortical plasticity. *Neuron*. 2012; 75:250–264. [PubMed: 22841310]
72. Baker CI. The neural basis of visual object learning. *Trends Cogn. Sci.* 2010; 14:22–30. [PubMed: 19945336]
73. Hershler O, Hochstein S. The importance of being expert: top-down attentional control in visual search with photographs. *Atten. Percept. Psychophys.* 2009; 71:1478–1486. [PubMed: 19801608]
74. Malinowski P, Hubner R. The effect of familiarity on visual-search performance: evidence for learned basic features. *Percept. Psychophys.* 2001; 63:458–463. [PubMed: 11414133]
75. Wang Q, et al. Familiarity and pop-out in visual search. *Percept. Psychophys.* 1994; 56:495–500. [PubMed: 7991347]
76. Op de Beeck HP, et al. Discrimination training alters object representations in human extrastriate cortex. *J. Neurosci.* 2006; 26:13025–13036. [PubMed: 17167092]
77. Gauthier I, Logothetis NK. Is face recognition not so unique after all? *Cogn. Neuropsychol.* 2000; 17:125–142. [PubMed: 20945176]
78. Logothetis NK, et al. Shape representation in the inferior temporal cortex of monkeys. *Curr. Biol.* 1995; 5:552–563. [PubMed: 7583105]
79. Liu H, et al. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron*. 2009; 62:281–290. [PubMed: 19409272]
80. Serre T, et al. A feedforward architecture accounts for rapid categorization. *Proc. Natl. Acad. Sci. U.S.A.* 2007; 104:6424–6429. [PubMed: 17404214]
81. Riesenhuber M, Poggio T. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* 1999; 2:1019–1025. [PubMed: 10526343]
82. Roelfsema PR. Cortical algorithms for perceptual grouping. *Annu. Rev. Neurosci.* 2006; 29:203–227. [PubMed: 16776584]
83. Downing PE, Peelen MV. The role of occipitotemporal body-selective regions in person perception. *Cogn. Neurosci.* 2011; 2:186–203. [PubMed: 24168534]
84. van Koningsbruggen MG, et al. A causal role for the extrastriate body area in detecting people in real-world scenes. *J. Neurosci.* 2013; 33:7003–7010. [PubMed: 23595757]
85. Evans KK, et al. The gist of the abnormal: above-chance medical decision making in the blink of an eye. *Psychon. Bull. Rev.* 2013; 20:1170–1175. [PubMed: 23771399]
86. Kuai SG, et al. Learning optimizes decision templates in the human visual cortex. *Curr. Biol.* 2013; 23:1799–1804. [PubMed: 24012311]
87. Chun MM. Contextual cueing of visual attention. *Trends Cogn. Sci.* 2000; 4:170–178. [PubMed: 10782102]
88. Biederman I, et al. Scene perception: detecting and judging objects undergoing relational violations. *Cogn. Psychol.* 1982; 14:143–177. [PubMed: 7083801]
89. MacEvoy SP, Epstein RA. Constructing scenes from objects in human occipitotemporal cortex. *Nat. Neurosci.* 2011; 14:1323–1329. [PubMed: 21892156]
90. Greene MR, Oliva A. Recognition of natural scenes from global properties: seeing the forest without representing the trees. *Cogn. Psychol.* 2009; 58:137–176. [PubMed: 18762289]
91. Torralba A, et al. Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychol. Rev.* 2006; 113:766–786. [PubMed: 17014302]
92. Wolfe JM, et al. Visual search in scenes involves selective and nonselective pathways. *Trends Cogn. Sci.* 2011; 15:77–84. [PubMed: 21227734]

93. Oliva A, Torralba A. The role of context in object recognition. *Trends Cogn. Sci.* 2007; 11:520–527. [PubMed: 18024143]
94. Vickery TJ, et al. Setting up the target template in visual search. *J. Vis.* 2005; 5:81–92. [PubMed: 15831069]
95. Wolfe JM, et al. How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Res.* 2004; 44:1411–1426. [PubMed: 15066400]
96. Delorme A, et al. Key visual features for rapid categorization of animals in natural scenes. *Front. Psychol.* 2010; 1:21. [PubMed: 21607075]
97. Reeder RR, Peelen MV. The contents of the search template for category-level search in natural scenes. *J. Vis.* 2013; 13:13. [PubMed: 23750015]
98. Ullman S, et al. Visual features of intermediate complexity and their use in classification. *Nat. Neurosci.* 2002; 5:682–687. [PubMed: 12055634]
99. Mangun GR, Hillyard SA. Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual–spatial priming. *J. Exp. Psychol. Hum. Percept. Perform.* 1991; 17:1057–1074. [PubMed: 1837297]
100. Zhang W, Luck SJ. Feature-based attention modulates feedforward visual processing. *Nat. Neurosci.* 2009; 12:24–25. [PubMed: 19029890]
101. Summerfield C, Egner T. Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* 2009; 13:403–409. [PubMed: 19716752]
102. Nardo, D., et al. Spatial orienting in complex audiovisual environments. *Hum. Brain Mapp.* 2013. <http://dx.doi.org/10.1002/hbm.22276>

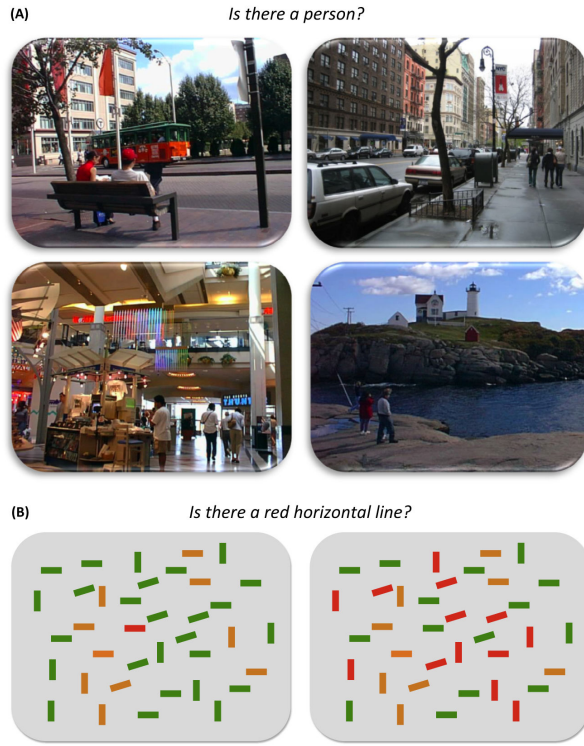


Figure 1. Search in real-world scenes versus search in artificial scenes. **(A)** Detecting the presence of people in these real-world scenes requires only a single glance, despite the large number and variety of distracter objects and the differing appearances of people in the four scenes. **(B)** Detecting the presence of a red horizontal line in the left display (feature search) is highly efficient, such that reaction times (RTs) are largely independent of the number of distracter items [2]. Detecting the same target is more difficult in the right display (conjunction search), with RTs increasing with each additional distracter added to the display [2]. Most experimental work has studied visual search using simplified displays like these.

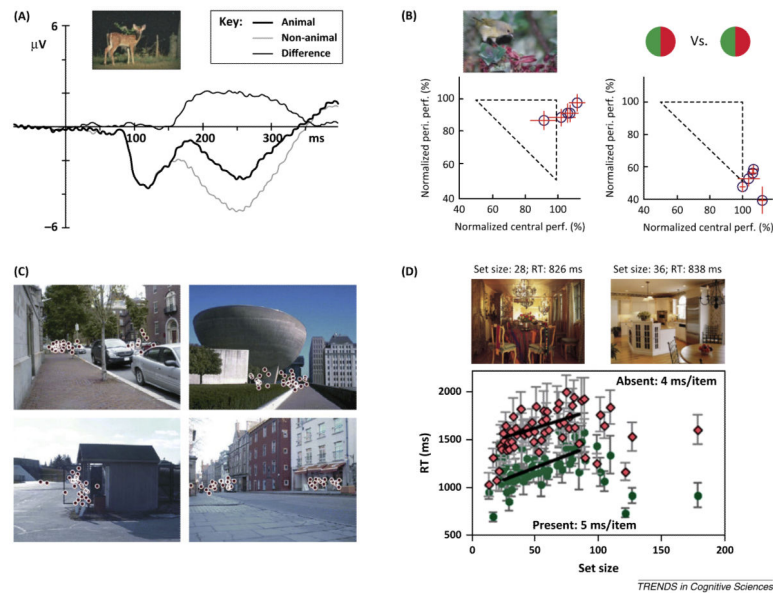


Figure 2.

Naturalistic vision is highly efficient. **(A)** The detection of familiar object categories (e.g., animals, vehicles) in scenes is extremely rapid, with electroencephalography (EEG) measures showing differences between target and no-target scenes starting at around 150 ms [67]. Reproduced, with permission, from [67]. **(B)** Li *et al.* [35] showed that participants could accurately detect animals or vehicles in natural scenes without focused spatial attention. In a dual-task paradigm, participants were instructed to prioritize a demanding visual-discrimination task at fixation (not shown). While they performed the central discrimination task, they additionally detected animals in scenes presented in the periphery. To obtain a baseline measure of performance, each of these tasks was also performed separately under single-task conditions. The graph on the left shows performance for the central and peripheral tasks normalized according to single-task performance. Importantly, performance on the peripheral animal detection task was only mildly impaired by simultaneously performing the central discrimination task. This was in stark contrast to superficially simpler peripheral tasks such as discriminating red–green from green–red disks (right panel); although these tasks were easy to perform on their own, performance dropped to chance level in the dual-task condition [35]. Reproduced, with permission, from [35]. Copyright (2002) National Academy of Sciences, USA. **(C)** Scene context guides search. Dots indicate the first three fixations of each of 14 observers searching for people in photographs of real-world scenes [68]. The target is absent in these scenes. Fixations were restricted to small regions of the scenes and were consistent across observers, reflecting contextual guidance. Reproduced, with permission, from [68]. **(D)** Search in real-world scenes is highly efficient [69]. Participants in this study searched for target objects, indicated by a word cue (e.g., ‘lamp’), in indoor scenes. Reaction time (RT) was plotted as a function of set size, the total number of objects present in a scene. It was found that each additional object in the scene added only about 5 ms to the RT, a highly efficient search. This contrasts with search for objects in artificial arrays, in which each distracter item adds >40 ms to the RT [69]. Reproduced, with permission, from [69].

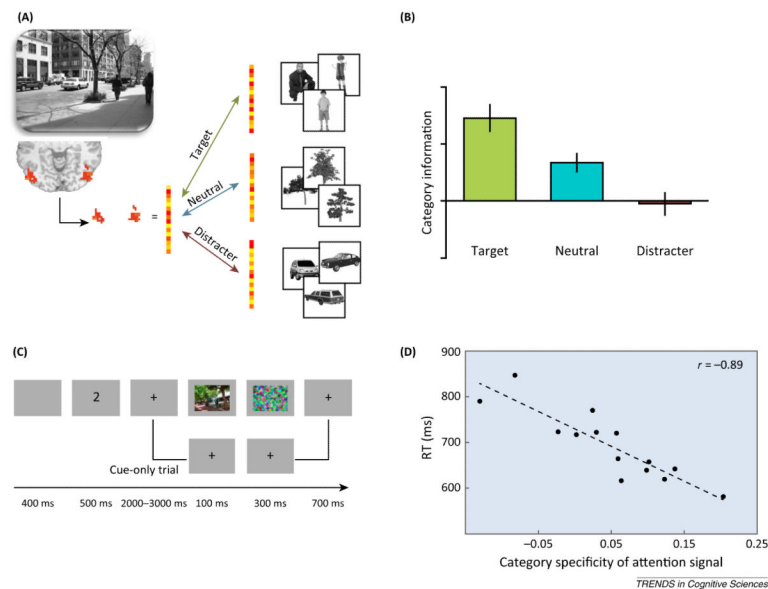


Figure 3.

Neural basis of category-level search in real-world scenes. **(A)** Method used in functional MRI (fMRI) studies to measure category information in the visual cortex during scene perception [27,28,30]. fMRI response patterns evoked by a scene, here schematically illustrated for object-selective cortex, are compared with fMRI response patterns evoked by objects shown in isolation in a separate experiment. The similarity of the scene-evoked patterns and the patterns evoked by each of the object categories (people, trees, cars) is measured as a function of the task relevance of the objects in the scene. In this example, people are the target (task-relevant) category, trees are the neutral (never-relevant) category, and cars are the distracter (previously relevant) category [30]. **(B)** Category information contained in scene-evoked response patterns in object-selective cortex was modulated by task relevance. The scene-evoked response pattern was more similar to the target category than to a neutral category, indicating attentional enhancement. Furthermore, the scene-evoked pattern was less similar to the distracter (previously relevant) category than to a neutral category, indicating attentional suppression [30]. Visual search in real-world scenes thus involves both attentional enhancement of targets and attentional suppression of distracters. Reproduced, with permission, from [30]. **(C)** Paradigm to measure top-down attention signals during search in real-world scenes [28]. On each trial, a symbolic cue (e.g., a digit) indicated the category (people, cars) that participants were instructed to detect in the subsequent scene. Importantly, on 33% of trials the cue was presented but was not followed by a scene, allowing the isolation of purely top-down attention signals. Reproduced, with permission, from [28]. **(D)** Response patterns in object-selective cortex on cue-only trials resembled those evoked by viewing pictures of the cued object category (x-axis). The category specificity of this top-down attention signal was strongly negatively correlated (across participants) with reaction time (RT; y-axis), indicating that category-level search in real-world scenes is guided by top-down search templates [28]. Reproduced, with permission, from [28].

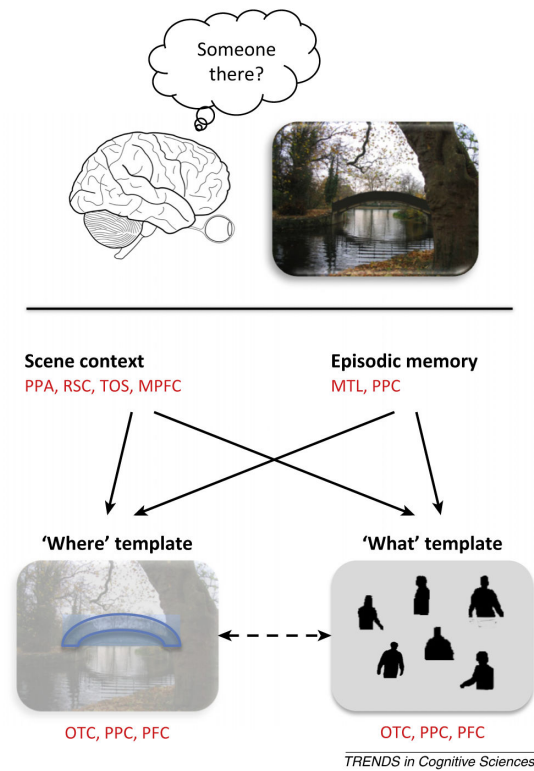


Figure 4.

Framework for top-down attentional influences during search in real-world scenes. In this framework, content-specific ('what') and location-specific ('where') attentional templates guide search in scenes. 'What' templates comprise attributes distinguishing target from distracter objects and thus integrate target and distracter properties of the scene. Representations of target and distracter attributes are influenced by the current scene context and by knowledge of the specific scene being searched (episodic memory). 'Where' templates represent the locations where targets are expected to occur, influenced by current scene context and episodic memory. Each component of the framework is associated with putative brain regions, further discussed in the main text. PPA, parahippocampal place area; RSC, retrosplenial cortex; TOS, transverse occipital sulcus; MPFC, medial prefrontal cortex; MTL, medial temporal lobe; PPC, posterior parietal cortex; OTC, occipitotemporal cortex; PFC, prefrontal cortex.