The search template for object detection in natural scenes:

Contents, characteristics, neural underpinnings, and individual differences

Reshanne Reeder

Center for Mind/Brain Sciences, University of Trento

Adviser: Dr. Marius Peelen

Acknowledgements

I planted the corn, I harvested the corn, I baked the corn, and now I've got a big heaping plateful of cornbread. But unlike the unfortunate Little Hen, I get to thank some people who helped me arrive at this thesis feast. First, Marius Peelen for his supervision on these projects and specifically for refining my experimental design and writing skills and preparing me to be an independent researcher: because of this I am now ready to step off into the land of postdoctoral research. I also thank the many members of the Peelen lab who have offered their comments and criticisms of my studies at all stages of development, from design to interpretation of results: they have undoubtedly made those studies better. I also thank Wieske van Zoest for her work as a member of my advisory committee: her comments on my presentations and thesis, help and encouragement with applications, and her collaboration have been invaluable. I must give great thanks to Rakesh Sengupta, who has seen me through it all: who got excited with me over crazy theories, listened to my woes, read my papers, and promised to be a future collaborator. Finally, I must thank two others responsible for my continued psychological well-being: Philipp Ruhnau, for his untiring reinforcement and advice (having done it all before) and Jana, for making me laugh about this whole thing even if she had no idea what I was doing.

Table of Contents

Abstract

The work presented here is at the meeting point of two branches of visual search research, one of which focuses on the proposition that visual search is guided by preparatory internal representations of targets (i.e., search templates: e.g., Bravo & Farid, 2009; 2012; Castelhano & Heaven, 2010; Duncan & Humphreys, 1989; Malcolm & Henderson 2009; 2010; Schmidt & Zelinsky, 2009; Vickery, King, & Jiang, 2005; Wolfe, 2007; Wolfe, Cave, & Franzel, 1989; Yang & Zelinsky, 2009), and the other of which focuses on investigating target detection in naturalistic search environments (e.g., Delorme, Richard, & Fabre-Thorpe, 2010; Delorme, Rousselet, Macé, & Fabre-Thorpe, 2004; Li, VanRullen, Koch, & Perona, 2002; Peelen, Fei-Fei, & Kastner, 2009; Peelen & Kastner, 2011; Thorpe, Fize, & Marlot, 1996; VanRullen & Thorpe, 2001). The search template for objects presented in naturalistic scenes is relatively unknown in terms of its content and characteristics, neural underpinnings, and individual differences in its representation. This thesis explores these topics in depth using behavioral and neurostimulation methods in four experimental chapters.

Chapter one: Introduction

This chapter will begin with a review of the relevant literature on the aforementioned topics, including a short history of selective attention and the current models of visual search, what we have learned so far about the preparatory template for object categories and naturalistic search targets, the cortical regions presumably involved in search template activity, and the individual differences associated with category-level search in naturalistic scenes. I will then summarize the four experimental chapters of this thesis.

1.1 Major models of visual search

In the natural world, there are millions of stimuli to which one can attend. The visual system is limited in its capacity to process items in the visual field (Duncan, 1980; Kastner, Pinsk, De Weerd, Desimone & Ungerleider, 1999; Kastner, De Weerd, Pinsk, Elizondo, Desimone & Ungerleider, 2001), but we are very good at selecting and filtering visual input to perform efficient visual processing. For example, when you are looking for cars while crossing a busy street, your attention is focused on the vehicles around you and not on the buildings or trees nearby.

Early studies of visual selective attention suggested that we filter our attention to items based on their association to a task-relevant conceptual category.

Detecting an item based on its category (e.g., a number among a display of letters) was considered “partial processing”, where detailed information about a target (such as the identity of the number) was ignored, resulting in more rapid target detection. Processing the identity of a target was considered “full processing”, which was thought to be less efficient than partial processing. In support of this hypothesis, Gleitman and Jonides (1978) asked subjects to detect a target surrounded by distractors from the same category (e.g., the number 2 in a display composed of numbers) or distractors from a different category (e.g., the number 2 in a display composed of letters). Within-category search showed significantly steeper search slopes with an increase in set size compared to between-category search, which the authors took as evidence that attention can be allocated more efficiently when targets and distractors are semantically dissimilar. In further support of this claim, the authors found that when subjects were cued to detect the number “zero” among letters, they showed shallower search slopes than when they were cued to detect the letter “oh” among letters. This was referred to as the “oh-zero” effect, in which the ambiguous character was detected with different efficiency depending on its semantic association, suggesting that targets were optimally detected if they were conceptually dissimilar from distractors. However, later studies could not replicate the “oh-zero” effect (Duncan, 1983), and instead posited that efficient detection was more dependent on physical, rather than conceptual, target-distractor dissimilarity. This claim was further verified by Krueger (1984), who replicated the task by Gleitman and Jonides (1978) using hand-drawn letters and numbers that controlled for the physical dissimilarities between these types of characters (e.g., the number 3 and a stylistically similar letter B would be presented in the same display). With the addition of this control, the between-category advantage was eliminated.

These studies led to a seminal paper by Duncan and Humphreys (1989), who suggested that visual search efficiency is dependent on the physical discriminability between targets and distractors. According to this model, in the earliest stages of search, all items in a display are rapidly processed in parallel and segmented into groups based on feature similarity and proximity. Limited attention resources are spread among these items in the visual field, equally distributed at first, but later in different quantities depending on the search goals. Greater resources directed to an item leads to a decrease in the resources directed to other items in a competitive fashion, and this increases the chances that the item will be preferentially selected and represented in visual short term memory. An item may be pre-assigned a particular weight, or bias, of resources if it is the target of goal-driven search. In this way, attention can be directed by a preparatory "template" of target information to items in the visual field that have the highest "match" of features with the template (i.e., they receive the greatest bias of resources).

The idea that attention resources are distributed by competitive bias among items in a display led to the advent of the biased competition model of selective attention (Desimone & Duncan, 1995). The biased competition model suggests, as in the earlier visual search model, that we have a limited capacity to process items in the visual field, which results in selective attention to certain items at the expense of other items. Resources are distributed to whole objects ("object" meaning anything with a collection of features separable from other items in the display); for examples, search for two properties of the same object is easier than search for a different property of two different objects. Competition is a neural process: objects in a neuron's receptive field (RF) are mutually suppressive, which means that when there are two items in a display, each suppresses the excitatory neural response of the other, resulting in a weaker neural response to either object compared to a single object in isolation. Increasing the number of objects in the RF leads to more competition between them for limited resources and a weaker neural response to items that would show a strong response when presented alone. Objects outside of a neuron's RF do not have a suppressive effect on items within the RF, so objects presented far apart from each other (e.g., on opposite sides of a computer screen) will not compete for resources. All stimuli in a neuron's RF initially compete for attention, but if a stimulus has high physical salience or if it is a predetermined target, attention can be biased to that stimulus preferentially over other items in the display, thus eliminating the suppressive effect on that object by the other objects. In other words, the neural competition for selection can be biased toward a particular stimulus, which is the basis for the biased competition model. This model further integrates the idea that an attentional template can influence the competitive bias among stimuli in the visual field. Stimuli that are poor matches to the attentional template receive a weak competitive bias, lifting their suppressive effect from items that match the template.

Other models have adopted the idea that goal-driven search is directed by "search templates". According to the guided search model, features in the environment reach the visual system in parallel in the initial stage of processing, but selective attention for items matching the search template enables a filter for further processing referred to as a “bottleneck” of attention. Features that match the active search template may pass through the bottleneck, whereas those that do not match the template may not be processed further than the initial stage. Template-matching features are accumulated until a threshold is reached at which a target can be identified, while other information in the visual field is ignored, thus optimizing search performance (Wolfe et al., 1989). Nowadays, many researchers argue that a search template is critical in guiding attention in the environment, and thus heavily influences visual perception (Kuo, Stokes, & Nobre, 2012; Malcolm & Henderson 2010).

1.2 Characteristics of the search template

Previous behavioral studies suggest that the search template can only represent a single item at one time (Houtkamp & Roelfsema, 2009), but multiple features of that item may be activated simultaneously (Malcolm & Henderson, 2010), and different features may be selectively attended for different tasks (Folk, Remington, & Johnston, 1992). Different templates can also be activated depending on the amount of informative content provided by a search cue. Search performance incrementally improves with the amount of visual information provided (Hwang, Higgins, & Pomplun, 2009; Schmidt & Zelinsky, 2009; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004); an image cue that matches the target precisely is most effective in guiding search (Delorme et al., 2004; Malcolm & Henderson, 2009; 2010), whereas an image representative of a target category, or a word cue with a visually descriptive adjective (e.g., “red purse”), are both less effective than an exact image but more effective than a single word cue (e.g., “purse”; Castelhano & Heaven, 2010; Malcolm & Henderson 2009; 2010; Schmidt & Zelinsky, 2009; Vickery et al., 2005).

Search templates approximate, rather than perfectly represent, the features of an object, which allows for a certain amount of flexibility between the cue and the target image (Bravo & Farid, 2009; Vickery et al., 2005). Templates activated by a specific image cue can still guide attention to targets that do not exactly match the cue (Bravo & Farid, 2009; Li & DiCarlo, 2008; Vuilleumier, Henson, Driver, & Dolan, 2002), perhaps because we expect certain changes in the appearance of an object (in lighting, distance, occlusion, etc.) in the dynamic natural world. Evidence for this comes from a study by Bravo and Farid (2009), which showed that a specific image could efficiently cue targets that varied in size or orientation. Additional evidence comes from a study by Ghose and Liu (2013); in one experiment, subjects were instructed to sort images as “new” or “old”, in which repetitions of a specific image were considered “old” and different viewpoints of the same image were considered “new”. The authors found that subjects responded “old” more often than “new” to different viewpoints of the specific image. This suggests that subjects involuntarily activated a view-invariant template for the specific targets.

1.3 Category-level search templates

We can process objects at subordinate (e.g., “Collie), basic (e.g., “dog”), and superordinate (e.g., “animal”) category levels depending on task goals, and previous behavioral studies have found that processing speed and detection accuracy depend on the level at which a target must be classified (Delorme et al., 2004; Large, Kiss, & McMullen, 2004). Processing efficiency also depends on the category level of non-targets in the environment: subjects are fastest at identifying targets if distractors are from a different superordinate category (e.g., search for a target dog among furniture results in faster RT than search for a target dog among other animals; Macé, Joubert, Nespoulous, & Fabre-Thorpe, 2009), presumably because the features of targets and distractors of different superordinate categories are maximally discriminable from one another, whereas items of the same superordinate category share more features (Markman & Wisniewski, 1997), and so require greater attention resources for efficient detection.

It therefore comes as no surprise that the target features activated in the search template can change with different task demands (Bravo & Farid, 2012; Foulsham & Underwood, 2007; Large et al., 2004; Underwood, Foulsham, van Loon, & Underwood, 2005; Vickery et al., 2005; Wolfe et al., 2004). For example, Yang and Zelinsky (2009) found that both specific (picture) and general (name) cues were effective for target detection if subjects were instructed to look for various targets at the basic category level, but if subjects were required to search for a target at the subordinate level, only the specific cue was effective, suggestive of a less flexible template. Searchers can rely on the visual features of a single exemplar to find a specific target, but in the case of basic category level search, the visual features in the template must be able to generalize across many exemplars of the same basic category (Bravo & Farid, 2012; Yang & Zelinsky, 2009).

In line with this, there is evidence that people attend to different visual features of an object if it must be identified on a basic or subordinate level. One behavioral study found that high spatial frequency information (such as texture patterns) is necessary to classify objects at the subordinate level (e.g., Labrador), whereas low spatial frequency information (such as general object shape) is sufficient to classify images at the basic level (e.g., dog); this suggest a qualitative difference in the processing of objects at subordinate and basic category levels (Collin & McMullen, 2005).

Basic-level (from now on simply referred to as "category-level") detection requires the activation of a template that can generalize across both typical and atypical category exemplars (Castelhano, Pollatsek, & Cave, 2008), and how we prepare for those exemplars is not well understood. For example, preparing for a particular layout of category-diagnostic features would work if targets were typical (e.g., chairs with a high back and four legs), but then atypical exemplars would be overlooked (e.g., Papasan chairs, which share very few features with wooden chairs); nevertheless, searchers are quick to fixate both typical and atypical exemplars from a target category (Castelhano et al., 2008). What features must the category-level search template be composed of that would allow searchers to adequately prepare for such varied exemplars of a single object category while simultaneously excluding all other categories? This is not a trivial question, and it becomes even more complicated when object categories must be detected in realistic settings.

1.4 Templates for naturalistic search

An effective search template must consist of features that optimally distinguish targets from non-targets, but features of even a single object in the natural world may vary in perspective, size, and location, and changes appearance constantly due to motion, lighting, distance, viewpoint, position, and occlusion; different objects within the same category (e.g., cars) additionally vary in color pattern, texture, size, and shape (e.g., a sports car versus a truck). Furthermore, non-targets in the natural world normally share many low-level features (e.g., colors, orientations) with targets. Despite these many variations between targets and the feature overlaps between targets and non-targets, humans are surprisingly efficient at detecting many different kinds of objects in naturalistic contexts, even at the category level (Li et al., 2002; Peelen & Kastner, 2011; Thorpe et al., 1996; VanRullen & Thorpe, 2001). A large portion of this thesis is devoted to resolving how this is achieved. In particular, I used behavioral methods to investigates the contents (Chapter 2) and characteristics (Chapter 3) of the search template for familiar object categories presented in naturalistic contexts, and neurostimulation methods (Chapter 4) to explore the neural underpinnings of this search template to gain a broader perspective of the resources it recruits.

1.5 Neural correlates of the search template

Early neural evidence for the search template comes from single cell recording studies in monkeys. In delayed match-to-sample tasks, neurons in inferior temporal cortex (IT) show task-relevant activity throughout the delay period between cue and target presentation, which is indicative of sustained preparatory attention for the target (Chelazzi, Duncan, Miller, & Desimone, 1998; Chelazzi, Miller, Duncan, & Desimone, 1993; Fuster & Jervey, 1982). Another defining feature of the search template is that it biases top-down attention to relevant stimuli and away from task-irrelevant stimuli, which is reflected in enhanced neural activation to task-relevant features and suppressed activation to task-irrelevant features primarily in IT cortex (Chelazzi et al., 1998; Chelazzi, Miller, Duncan, & Desimone, 2001; De Weerd, Peralta, Desimone, & Ungerleider, 1999; Moran & Desimone, 1985). Finally, several previous studies have found evidence that single neurons in IT can activate to multiple properties of the same object (Booth & Rolls, 1998; Fuster, 1990; Li & Dicarlo, 2008; Tanaka, 1996), as well as variations in those properties (David, Hayden, Mazer, & Gallant, 2008; Fuster & Jervey, 1982; Miyashita & Chang, 1988), suggesting that IT represents flexible, view-invariant features of targets. These studies provide evidence for a complex preparatory mechanism in monkey IT cortex that facilitates the selection of task-relevant stimuli throughout the search process.

Neuroimaging studies in humans also provide evidence that object information activated in high-level visual cortex results in more efficient detection of targets (e.g., Puri, Wojciulik, & Ranganath, 2009). Comparable to non-human primate studies, humans show selective neural enhancement for task-relevant features and suppression of task-irrelevant features in target-selective regions such as the Parahippocampal Place Area (Gazzaley, Cooney, McEvoy, Knight, & D'Esposito, 2005; O'Craven, Downing, & Kanwisher, 1999), feature-selective visual areas such as V4 for color (Corbetta, Miezin, Dobmeyer, Shulman, & Petersen, 1990), as well as prefrontal cortex (Barceló, Suwazono, & Knight, 2000; Bar et al., 2006). However, the neural mechanisms involved in naturalistic, category-level search have only recently been explored in humans.

Two recent studies used fMRI methods to explore the neural correlates of our nearly automatic ability to detect object categories (people and cars) in diverse images of real-world scenes (Peelen et al., 2009; Peelen & Kastner, 2011). Peelen et al., (2009) found evidence for car and person-specific patterns of activity in object-selective cortex (OSC) even when cued categories were task-irrelevant and presented in task-irrelevant locations, suggesting that real-world category search is nearly attention free. More recently, Peelen and Kastner (2011) found that category-specific patterns of activity in an area of the right posterior temporal cortex (pTC), overlapping both object- and scene-selective regions, is positively correlated with behavioral performance; conversely, category-specific patterns in early visual cortex (EVC) was negatively correlated with behavioral performance in this study. These results suggest that naturalistic, category-level search is an efficient process based on a high-level search template in high-level visual cortex, particularly pTC.

Peelen & Kastner (2011) found evidence for a correlation between category-specific preparatory activity and search performance in pTC, but there are a few open questions related to this study: first, because fMRI can only be used to explore correlational relationships, the causal role of pTC in search template activity remains to be explored. Additionally, it is unclear whether this area is preferentially involved in category-level search, or whether it is generally recruited for search at different levels of specificity. Trans-cranial magnetic stimulation (TMS) is an adequate tool to address these questions, as it can be used to selectively disrupt activation in small areas of cortex during a task to explore causal relationships between a cortical region and behavioral performance.

TMS is placed on the head with the center of the coil over the targeted stimulation area. The coil then generates a magnetic field, which, when placed on the scalp, painlessly penetrates the cortex and creates an electrical pulse that disrupts normal neural function for about 100 ms. Subjects often do not even notice the disruption of function in such a brief amount of time, but it is enough to measure accuracy and reaction time (RT) differences on behavioral tasks. TMS is highly localized, in that very small areas of cortex can be selectively stimulated. An area of cortex is thought to be critically involved in a particular behavioral task if an impairment of performance is observed following TMS to that area. Chapter 4 of this thesis reports investigations of the causal role of pTC in the search template for naturalistic, category-level search using TMS.

1.6 Individual differences in the search template

Top-down search requires locating predetermined targets amongst distractors, so the optimal strategy relies on heightened attention to target features and suppression of task-irrelevant information during the search process; however, there is evidence that the ability to do this varies across individuals. Target enhancement and distractor suppression is correlated with working memory (WM) capacity (Kane, Bleckley, Conway, & Engle, 2001), and searchers with higher WM capacity perform better during goal-driven search than those with lower WM capacity (Sobel, Gerrie, Poole, & Kane, 2007). Additionally, a more effective preparatory representation of target features (i.e., search template) is correlated with enhanced attention orienting capabilities (Giesbrecht, Weissman, Woldorff, & Mangun, 2006), but not all searchers activate effective templates; for example, previous studies have found that preparatory activity in high-level object selective cortex such as the lateral occipital complex (LOC) and an area of the right posterior temporal cortex (pTC) correlates positively with category-level search performance in naturalistic scenes (Peelen & Kastner, 2011; Soon, Namburi, & Chee, 2012); conversely, preparatory activity in early visual cortex (EVC) correlates negatively with category-level search performance (Peelen & Kastner, 2011) and only correlates positively with performance on low-level visual search tasks (Ress, Backus, & Heeger, 2000). In other words, activating high-level visual cortex for category-level search is beneficial to performance, whereas activating low-level visual cortex reduces performance except during low-level feature search tasks.

All of these studies have found that individual differences in search performance correlates with BOLD response in different regions. Therefore, using TMS, we should find that individuals use different cognitive and neural strategies for category-level search (e.g., high-level preparatory strategy, low-level post-stimulus strategy) that should present as beneficial or detrimental to performance. Furthermore, searchers should be able to change their search strategies if they are required to switch between high-level (i.e., category-level) and low-level (i.e., feature-level) detection tasks. Because these tasks require attention to different features, they should also rely on different neural regions for successful completion. In order to establish the causal involvement of pTC in preparing for naturalistic object categories, we focused on the extent to which natural search strategies (general or specific) and preparing for different search tasks (high-level or low-level) rely on pTC. We hypothesized that pTC would be critically involved in general strategies and search templates for high-level search tasks, but not specific strategies and search templates for low-level search tasks; this is explored in Chapter 4 of this thesis.

1.7 Expertise and visual search

Thus far in the discussion of individual differences in search strategies, I have speculated on the characteristics of good and poor search strategies, but what about the strategies of people who have an expert level of experience with discriminating individuals from a particular object category? Can perceptual expertise somehow facilitate category-level detection? Despite extensive literature on work-related search expertise such as lifeguarding, airport security, and radiology (e.g., Koller, Hardmeier, Michel, and Schwaninger, 2008; Manning, Ethell, Donovan, & Crawford, 2006; McCarley, Kramer, Wickens, Vidoni, & Boot, 2004; Nodine, Kundel, Lauver, & Toto, 1996), previous studies have only scratched the surface on the effects of perceptual, or discrimination, expertise on visual search ability (Hershler & Hochstein, 2009). There is already evidence that experts process categories of perceptual expertise differently from novices; for example, they show just as high subordinate-level categorization as basic-level categorization for the expert object category (e.g., Bukach, Phillips, & Gauthier, 2010; Gauthier, Williams, Tarr, & Tanaka, 1998; Tanaka & Curran, 2001; Tanaka, Curran, & Sheinberg, 2005; Tanaka & Taylor, 1991) and detection abilities on par with faces (e.g., Gauthier & Tarr, 2002; Myles-Worsley, Johnston, & Simons, 1988); therefore, it will be a useful endeavor to explore the search abilities associated with perceptual expertise, despite a lack of search training. Specifically, it is important to investigate the relationship between individuation ability (perceptual expertise) and detection. Whether these processes rely on similar or different representations is highly disputed (Delorme, Rousselet, Macé, & Fabre-Thorpe, 2004; Grill-Spector & Kanwisher, 2005; Large, Kiss, & McMullen, 2004; Mack, Gauthier, Sadr, & Palmeri, 2008; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Thorpe, Fize, & Marlot, 1996), and facilitated detection following individuation training would suggest a link between these processes. In Chapter 5 of this dissertation, I explore the effects of perceptual expertise on category-level detection performance.

1.8 Summary of thesis chapters

1.8.1 Chapter two: The contents of the search template for category-level search in natural scenes

In this chapter, we investigated the contents of the search template during visual search for familiar object categories in real-world scenes. Visual search involves matching features in the environment to a search template, an internal representation of target information. The category-level template must contain features that can generalize to most members of an object category and exclude features that overlap with other object categories. In this study, we sought to determine the features that are most diagnostic for the categories “cars” and “people”.

Subjects were cued to detect people and cars in photographs of real-world scenes. On a subset of trials, task-irrelevant stimuli appeared instead of scenes, directly followed by a dot that subjects were instructed to detect. We hypothesized that task-irrelevant stimuli that matched the active search template would capture attention, resulting in faster detection of the dot probe when it was presented in the same location as a template-matching stimulus (consistent trials) than when it was presented in the opposite location (inconsistent trials).

Silhouettes of cars and people presented in various orientations (0°, 90°, or 180°) resulted in faster detection of the dot probe on consistent trials compared to inconsistent trials. Consistency effects were also observed for silhouettes of category-diagnostic parts, such as the arm and torso of a person. Finally, we found consistency effects on trials in which silhouettes were presented in locations that were irrelevant to the search task. We argue that search for object categories in real-world scenes is mediated by a search template composed of a set of view-invariant shapes of diagnostic object parts represented globally across the visual field.

1.8.2 Chapter three: Involuntary attentional capture by task-irrelevant objects that match the search template for category detection in natural scenes

In this chapter, we investigated various characteristics of the search template for category-level search in natural scenes using a paradigm we developed in our first behavioral study (Reeder & Peelen, 2013). After five experiments, we concluded that the search template is composed of view-invariant shapes of diagnostic object parts represented globally across the visual field. One assumption of the first study was that results reflected template-guided search in our paradigm, but they could just as easily have been due to trivial influences of the semantic cue (e.g., priming or a passive preference to attend to the cued category) or other top-down factors not accounted for.

In Experiment 1, we developed a variation of our previous paradigm and compared consistency effects on dot probe detection during two search tasks: one in which subjects could activate a visual search template, and one in which they could not. No cue-consistent effects were found in the probe task when subjects could not activate a visual search template, suggesting that attentional capture in our paradigm is contingent on an active search template.

In Experiment 2, we addressed whether the consistency effects observed in our previous paradigm are modulated by other top-down strategies than contingent attentional capture. We changed the paradigm so that attending to cued stimuli in the probe task would be detrimental to performance. Subjects were told that attending to cued silhouettes would impair performance (Experiment 2a) or that attending to uncued silhouettes would improve performance (Experiment 2b); in either case, subjects continued to orient to template-matching stimuli. These results support our hypothesis that the search template guides attention automatically and independently of other top-down strategies.

1.8.4 Chapter four: TMS reveals efficient category detection relies on preparatory activity in object-selective cortex prior to visual stimulation

In this chapter, we explored the extent to which category-level search in naturalistic contexts relies on the activation of search templates in object selective cortex. Detecting the presence of an object category in real-world scenes requires a relatively abstract search template that is flexible enough to account for the many ways in which object category members can appear. A previous fMRI study investigating the neural basis of category-level search in real-world scenes (Peelen & Kastner, 2011) found that category-specific preparatory activity patterns in a region in right posterior temporal cortex (pTC) was positively correlated with category-level search performance.

In two experiments, subjects were cued to detect cars and people in photographs of naturalistic scenes. In Experiment 1, we used TMS to disrupt an area thought to be involved in high-level search templates (pTC) and a control area (vertex) at 4 different time points before (-200 ms, -100 ms) and after scene onset (+100 ms, +200 ms). Results indicated that subjects who reported to use a more high-level search strategy showed a larger decrement in performance under stimulation to pTC compared to vertex in the pre-scene time window only. Additionally, these subjects showed superior search performance under vertex stimulation compared to subjects who reported to use a more low-level search strategy. These results suggest a causal role for pTC in efficient category-level detection in real-world scenes.

In Experiment 2, we applied TMS to pTC and early visual cortex (EVC) while participants searched for objects in real-world scenes at the category level (people or cars) or at the individual level (a specific person or car). We found a significant interaction between TMS region and task: TMS over pTC significantly impaired category-level search relative to TMS over EVC, whereas a trend in the opposite direction was observed for the individual-level search task. These results provide causal evidence for a neural distinction between search for categories and search for individuals in real-world scenes, suggesting that mechanisms mediating search in real-world scenes might flexibly adapt to meet current task demands; importantly, our results suggest that pTC is specifically involved in high-level attentional templates.

1.8.5 Chapter five: Behavioral effects of perceptual expertise on category-level detection in natural scenes

In this chapter, we were interested in researching the effects of perceptual expertise on category-level detection in naturalistic scenes, for which the ability to discriminate objects from the expert category is seemingly irrelevant. Perceptual expertise for an object category results in many processing advantages over other categories, and a detection advantage would suggest a link between the templates activated for individuation (expertise) and detection, a topic that is currently highly debated.

In this study we used a visual search paradigm to test the degree to which car expertise affects detection performance during car search compared to person search. Self-proclaimed car experts completed a car discrimination task that assessed their level of expertise, and then they performed a real-world detection task in which they were cued to search for cars and people on separate trials. Results revealed that car experts who scored higher on the discrimination task showed greater detection efficiency during car search relative to person search than experts who scored lower on the discrimination task. This indicates that perceptual expertise for cars significantly improves search performance for cars compared to people in naturalistic scenes, indicative of a link in the cognitive resources, and perhaps the template, recruited for individuation and detection.

Chapter two: The contents of the search template for category-level search in natural scenes

Reshanne R. Reeder[1], Marius V. Peelen[1]

[1]Center for Mind/Brain Sciences, University of Trento, 38068 Rovereto, Italy

The work reported in the following chapter is published as Reeder, R. R., & Peelen, M. V. (2013). The contents of the search template for category-level search in natural scenes. Journal of vision, 13(3), 13.

Acknowledgements:

We thank Timo Stein and Wieske van Zoest for helpful comments and discussions. M.V.P. was supported in part by the fondazione Cassa di Risparmio di Trento e Rovereto.

2.1 Abstract

Visual search involves the matching of visual input to a “search template” – an internal representation of task-relevant information. The present study investigated the contents of the search template during visual search for object categories in natural scenes, for which low-level features do not reliably distinguish targets from non-targets. Subjects were cued to detect people or cars in diverse photographs of real-world scenes. On a subset of trials, the cue was followed by task-irrelevant stimuli instead of scenes, directly followed by a dot that subjects were instructed to detect. We hypothesized that stimuli that matched the active search template would capture attention, resulting in faster detection of the dot when presented at the location of a template-matching stimulus. Results revealed that silhouettes of cars and people captured attention irrespective of their orientation (0°, 90°, or 180°). Interestingly, strong capture was observed for silhouettes of category-diagnostic object parts, such as the wheel of a car. Finally, attentional capture was also observed for silhouettes presented at locations that were irrelevant to the search task. Together, these results indicate that search for object categories in real-world scenes is mediated by spatially global search templates that consist of view-invariant shape representations of category-diagnostic object parts.

Keywords: natural scenes, visual search, object category, top-down attention

2.2 Introduction

Real-world visual search involves the selection of target objects among complex and diverse non-targets. In daily life, this selection often operates at the category level (e.g., looking out for cars when crossing a road, or for pedestrians when driving). Considering the infinite number of ways in which objects can appear in the real world, humans are remarkably good at selecting behaviorally relevant object categories in natural scenes (Li, VanRullen, Koch, & Perona, 2002; Thorpe, Fize, & Marlot, 1996)

It has been hypothesized that the attentional selection of a target object among distractors is achieved through the matching of incoming visual input to a top-down attentional set that guides visual search to items containing task-relevant features (Duncan & Humphreys, 1989; Wolfe, Cave, & Franzel, 1989). There is currently great interest in how visual features activated in the attentional set, or "search template", guide search. Studies of eye movements have found that searchers may mistakenly fixate distractors that share visual features with targets, which increases search times (Castelhano & Heaven, 2010; Castelhano, Pollatsek, & Cave, 2008; Pomplun, 2006). The more visual information provided by a cue, the better the search performance (Hwang, Higgins, & Pomplun, 2009; Schmidt & Zelinsky, 2009; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004); therefore, an image cue that matches the target exactly is most effective in guiding search (providing the most visual information relevant to the target), while an image representing the target category or a feature-and-word cue (e.g., "blue car") are both more effective than a word cue alone (e.g., "car"; Castelhano & Heaven, 2010; Malcolm &

Henderson 2009; 2010; Schmidt & Zelinsky, 2009; Vickery, King, & Jiang, 2005). At the same time, a search template that approximates, rather than perfectly replicates, an image allows for a certain amount of flexibility between the visual features of the cue and the target (Bravo & Farid, 2009; Vickery, et al., 2005). Some visual features may be preferentially represented in the search template. A previous study of feature search found that searchers are faster to detect targets based on color than orientation (Hannus, van den Berg, Bekkering, Roerdink, & Cornelissen, 2006), suggesting that some features are naturally weighted higher than others in the visual/attention system, thus biasing the search template. The contents of the search template may also be determined by task demands (Bravo & Farid, 2012; Foulsham & Underwood, 2007; Underwood, Foulsham, van Loon, & Underwood, 2005; Vickery, et al., 2005; Wolfe, et al., 2004). For example, the search template may represent different visual information for targets at different category levels. Searchers can rely on specific visual features of a single exemplar to find a target in individual-level search, but in the case of category-level search, activated visual features in the template must generalize across exemplars (Bravo & Farid, 2012; Yang & Zelinsky, 2009). Category-level search in real-world scenes adds even more complexity to the equation. An effective search template must consist of features that optimally distinguish targets from non-targets, but features of targets in real-world scenes may vary in perspective, size, and location, among other aspects; furthermore, non-targets in these scenes typically share many low-level features with targets. Currently little is known about the contents of the search template for category-level visual search in real-world scenes.

What might the template consist of for search tasks in which certain physical properties (size, perspective) and low-level features (color, orientation) are unlikely to be informative of the presence of a target category? One possibility is a holistic representation of the target category, one that contains information about both the shape and typical configuration of object parts. Previous research has proposed that such a prototype template may underlie face and body detection (Lewis & Ellis, 2003; Stein, Sterzer, & Peelen, 2012), which would explain the drop in detection performance when faces and bodies are inverted. Alternatively, the template may consist of view-invariant category-diagnostic shape features of intermediate complexity, such as a car's wheel or a person's leg (Evans & Treisman, 2005; Treisman, 2006). Activating various diagnostic features of an object category in parallel (e.g., arms, legs, or a torso) may be a more effective preparatory strategy than activating a holistic layout of features when the specific target exemplar is unknown prior to detection. In agreement with this idea, a computational study (Ullman, Vidal-Naquet, & Sali, 2002) showed that features of "intermediate complexity", such as a portion of a face that reveals the mouth and nose area, are optimally suited for classifying novel images, presumably because these features are consistent across variable exemplars (also see Yang & Zelinsky, 2009). Complementing these findings, Delorme, Richard, and Fabre-Thorpe (2010) found that the detection of animals in natural scenes is significantly impaired in the absence of diagnostic features (such as limbs), further supporting the idea that such feature information is critical to rapid and accurate object detection.

To better investigate the contents of the search template for naturalistic visual search, we developed a variant of the contingent attentional capture paradigm (Folk, Leber, & Egeth, 2002; Folk, Remington, & Johnston, 1992). Contingent attentional capture refers to the orienting of attention toward task-irrelevant stimuli that contain task-relevant features. For example, when subjects are instructed to attend to red items in a central rapid serial visual presentation stream, the appearance of an irrelevant red item in the periphery captures attention, as indicated by a decrement in central target identification (Folk, et al., 2002). A related study (Downing, 2000) showed that an irrelevant object that matches a target held in working memory captures spatial attention, leading to better discrimination of an immediately subsequent probe presented at the same location as the memory-matching object. Combining these approaches, we used a modified dot-probe paradigm (MacLeod, Mathews, & Tata, 1986) to measure the degree to which particular stimuli capture attention when subjects are prepared to detect real-world object categories.

In the current study, subjects were cued to detect people or cars in diverse photographs of natural scenes. Of interest was a subset of trials (25%) in which task-irrelevant stimuli without scene background appeared instead of scenes. Subjects were instructed to ignore these stimuli and simply indicate the location of a subsequently presented dot that could appear on the left or right of a central fixation cross. We hypothesized that stimuli that matched the active search template would capture attention, resulting in faster detection of the probe when presented at the same location as the template-matching stimulus. In 5 experiments, we systematically varied the properties of the task-irrelevant stimuli to reveal the contents of the search template during visual search for object categories in real-world scenes.

2.3 Methods

2.3.1 Subjects

Sixty-six undergraduate and graduate students from the University of Trento (53 females) participated in the experiments for course credit or payment: 11 in Experiment 1, 17 in Experiment 2, 13 in Experiment 3, 14 in Experiment 4, and 11 in Experiment 5. All subjects had normal or corrected-to-normal visual acuity and were between the ages of 18 and 38 years (mean=23.1 years). Four subjects participated in more than one experiment separated by at least one month. The research protocol of all experiments adhered to the tenets of the Declaration of Helsinki.

2.3.2 Stimuli

All stimuli were presented on a 19-inch Dell 1905 FP monitor with a screen resolution of 1280x1024 pixels and 60Hz refresh frequency. A fixation cross and letter cues appeared centered on the screen. Letter cues were uppercased with 70-point “strong” Times New Roman font. The fixation cross had dimensions of 31x31 pixels subtending 0.92 degrees in height and width, and letters had dimensions of 70x70 pixels subtending 2.1 degrees in height and width. Stimuli were presented using A Simple Framework (Schwarzbach, 2011), based on the Psychophysics Toolbox for MATLAB.

2.3.3 Natural scene stimuli

Stimuli presented in the search trials (75% of trials) were 864 color photographs of real-world scenes obtained from the LabelMe online database (Russell, Torralba, Murphy, & Freeman, 2008; see Figure 1a for some examples) and were divided into scenes containing cars (216), people (216), both cars and people (216), or neither cars nor people (216). Two scenes appeared on every trial and no scene was repeated within an experiment.

Scenes were scaled to 548x411 pixel resolution, subtending a visual angle of 16.07° x 12.1°. In Experiments 1-4, scenes were presented at a visual angle of 8.96° from the center of the screen to the center of the image, to the left and right of fixation. In Experiment 5, scenes were presented at a visual angle of 6.95° from the center of the screen to the center of the image, above and below fixation.

2.3.4 Attentional capture stimuli

Stimuli presented in the dot-probe trials (25% of trials) showed aspects of 192 photographs of cars (96) and people (96) without scene background (see Figure 1b). Most images were obtained from free-access online image sources, and were chosen to encompass a variety of viewpoints and features (e.g. a crouching child, a man standing; a truck as seen from the side, a sports car as seen from behind). Heads were removed from all images of people to be consistent with previous imaging studies that used headless bodies as reference stimuli to investigate neural correlates of the search template for the same real-world search task as in the current experiment (Peelen, Fei Fei, & Kastner, 2009; Peelen & Kastner, 2011; Seidl, Peelen, & Kastner, 2012). None of the stimuli presented in the dot-probe trials were shown in the scenes presented in the search trials.

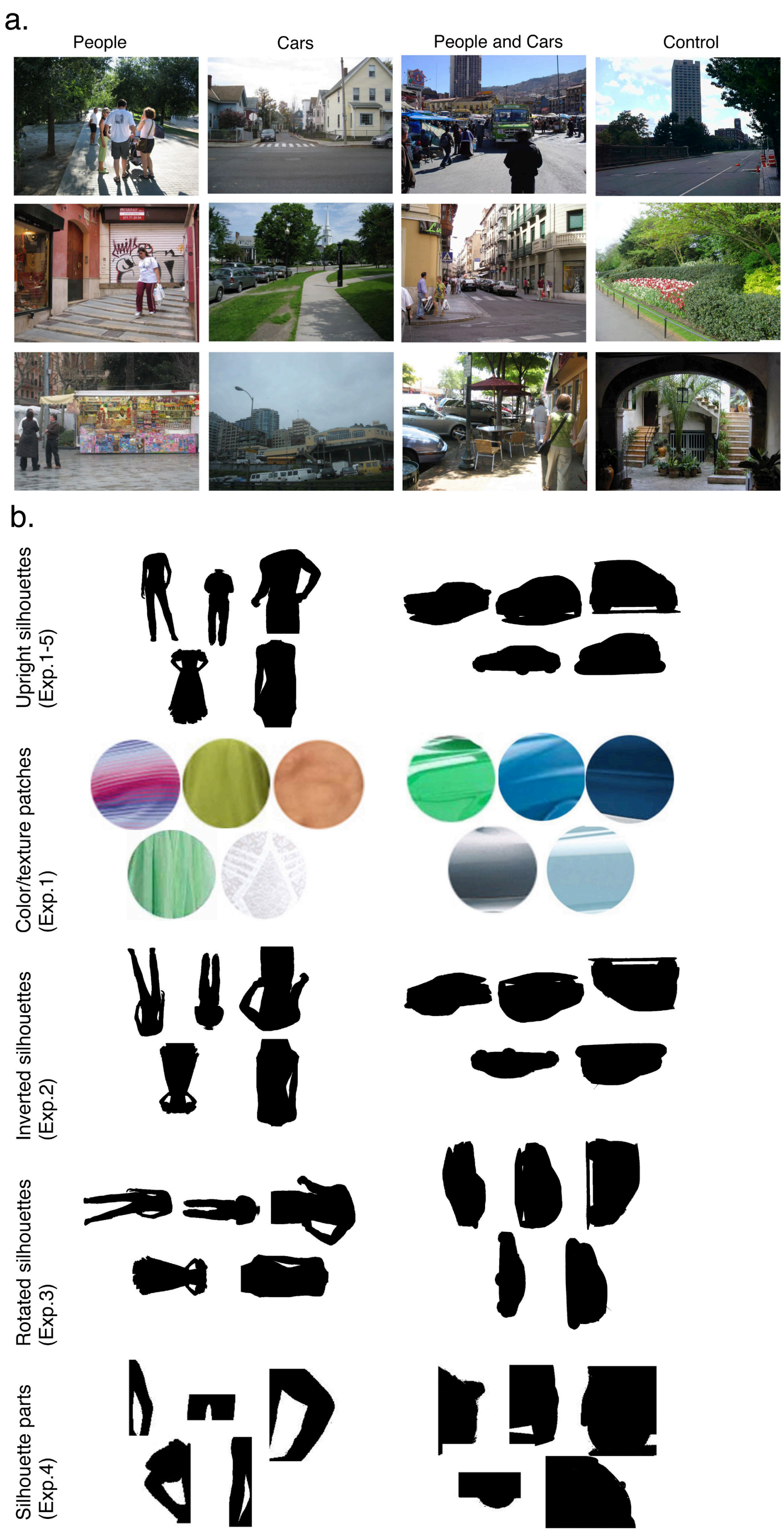

a.
People
Cars
People and Cars
Control
b.
Upright silhouettes (Exp.1-5)
Color/texture patches (Exp.1)
Inverted silhouettes (Exp.2)
Rotated silhouettes (Exp.3)
Silhouette parts (Exp.4)

Figure 1. Examples of stimuli used in 1a) the search task and 1b) the prime task.

Upright silhouettes of the person and car photographs were presented in every experiment. Additional transformations of the photographs were presented in Experiments 1-4. In Experiment 1, textures were cut from sections of the photographs. For person stimuli, textures consisted mainly of clothing patterns, although skin was used if the original photograph showed large amounts of exposed skin. Car textures were cut from the bodies of cars, which could include streaks, shines, the lines around doors, and the area above the wheel. Color/texture patches were based on the largest surface area occupied by a given color/texture in the original color photograph, with the additional constraint that the patches did not reveal shape features. All textures were adjusted to a constant radius of 150 pixels subtending a visual angle of 4.43°, located 8.52° from the center of the screen. In Experiment 2, silhouettes were rotated 180 degrees to create inverted images. In Experiment 3, silhouettes were rotated clockwise by 90 degrees. In Experiment 4, small diagnostic parts were taken from the upright silhouettes (e.g., an arm, a pair of feet, or the wheel of a car). The size of each part (based on the number of black pixels) never exceeded 25% of the size of the whole silhouette (range: 4.61-24.19%, mean=14.66% for cars and range: 7.68-23%, mean=14.27% for people). The parts were scaled such that they could appear in the same three possible sizes and locations as whole silhouettes during the experiment. In Experiment 5, upright silhouettes remained unchanged. See Figure 1b for examples of each of the stimuli described above.

Stimuli could appear in three possible sizes (100x100, 180x180, or 200x200 pixels, or 2.95° x 2.95°, 5.31° x 5.31°, or 5.90° x 5.90° of visual angle, respectively) and at three different screen locations along the X-axis, subtending 6.46°, 7.99°, or 10.75° of visual angle. Size and location values were chosen randomly on each trial and independently for the left and right stimulus. On each prime trial, a single aspect of a car and person appeared to the left and right of fixation. Aspects of cars and people appeared on the left and right an equal number of times. Each image was repeated once per experiment under two different transformations (e.g. 192 upright silhouettes and 192 rotated silhouettes). Transformations were randomly intermixed within the experiment but were not mixed within a trial, that is, an upright silhouette of a person always appeared with an upright silhouette of a car, and never with a rotated silhouette of a car.

2.3.5 Procedure

All subjects completed one practice block followed by nine blocks of 64 trials each. Each block was made up of two tasks as illustrated in Figure 2. The search task made up 75% of trials in a block to ensure that subjects actively prepared to detect the cued object category. Trials were randomized so subjects did not know whether they would perform the search task or the prime task on any given trial.

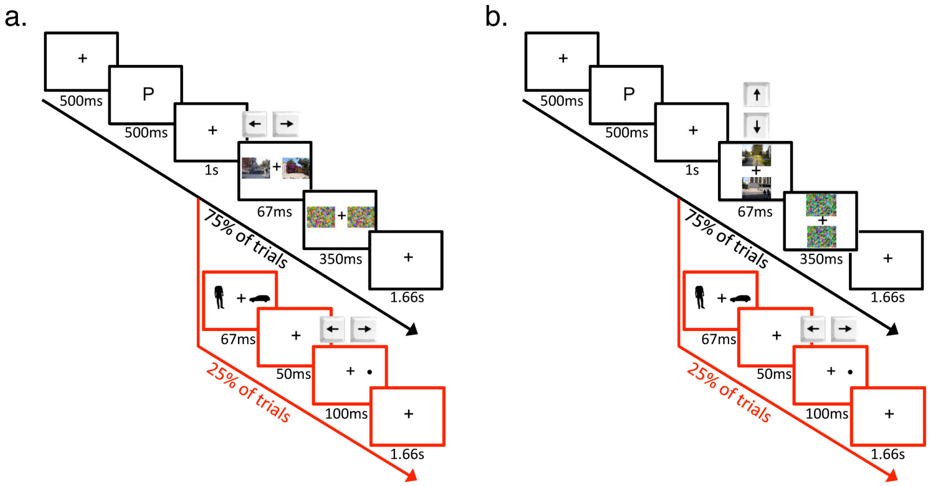


Figure 2. Schematic outline of the experimental paradigm for 2a) Experiments 1-4 and 2b) Experiment 5. In the prime task (25% of trials), subjects were required to respond whether the dot probe appeared on the left or right of fixation. In the search task (75% of trials), subjects were required to respond whether the cued object appeared in the left or right scene (Experiments 1-4) or in the top or bottom scene (Experiment 5).

For both tasks, a trial began with the presentation of a fixation cross for 500ms, followed by a single letter for 500ms: "P" for "persona" or "M" for "macchina" (the Italian words for person and car, respectively). After the letter, another fixation cross appeared for 1s. On search trials, subjects would then see two photographs of real-world scenes for 67ms, followed by a 350ms mask. On prime trials, subjects would instead see the primes representing a person and a car for 67ms, then a fixation cross for 50ms, followed by a dot probe that would appear for

100ms, 8.52° from the center of the screen on the left or right. The trial sequence for both tasks ended with a 1.66s fixation.

For the search task in Experiments 1-4, subjects were instructed to respond whether a cued object category (person or car) appeared in the scene on the left or right using the left and right arrow keys, respectively (Figure 2a). In Experiment 5, subjects were instructed to respond whether the cued object category appeared in the scene above or below fixation using the up and down arrow keys, respectively (Figure 2b). The cued object category always appeared in one of the two scenes. The two scenes that appeared could either be one containing cars and the other containing people, or one containing both cars and people and the other containing no cars or people. This structure allowed us to present cars and people on every trial without making the presence of one category informative of the absence of the other category. Each of the four scene types appeared in each possible location an equal number of times (left and right for Experiments 1-4, or up and down for Experiment 5).

For the prime task, subjects were instructed to respond using the arrow keys whether the dot probe appeared to the left or right of fixation. Subjects were instructed to ignore the prime images (i.e. the various transformations of silhouettes or color/texture patches) that appeared prior to the probe. Prime images did not predict the probe's location, and the probe appeared on the left and right an equal number of times.

2.3.6 Analysis

For the prime task, we analyzed accuracy and reaction time (RT) for consistent and inconsistent trials. Consistent trials were those in which the cued prime (e.g., the person prime following the "P" cue) appeared on the same side of fixation as the dot probe. Inconsistent trials were those in which the cued prime appeared on the opposite side of fixation. Only correct trials were included in the RT analysis. Subjects were excluded from analysis if their mean prime task accuracy fell 2.5 standard deviations below the group mean for the experiment. Three subjects were excluded based on this criterion (one each from Experiments 1, 2, and 4).

For the search task, we analyzed accuracy and RT. Only correct trials were included in the RT analysis. Results of the search task are reported in Table 1. The Results section reports the results of the prime task only.

Table 1. Mean reaction time (ms) and accuracy (% correct), with standard deviation, in the search task for Experiments 1-5.

	Reaction time (ms)	Accuracy (% correct)
Exp.1	613±96	80.4±5.6
Exp.2	606±124	82.3±5.4
Exp.3	664±127	83.6±6
Exp.4	655±123	80.4±8.5
Exp.5	737±76	76.4±5

2.4 Results

2.4.1 Experiment 1: Upright silhouettes vs. color/texture patches

Experiment 1 was conducted to test whether the search template for category-level search consists of object shape and/or surface features (texture and color; see Figure 1b). In the prime trials, subjects had to ignore the prime images (half upright silhouettes and half color/texture patches) and respond whether a dot probe appeared on the left or right.

RT

Figure 3 illustrates mean RT for consistent and inconsistent trials in the upright silhouette and color/texture conditions. A repeated-measures ANOVA with prime type (silhouette, color/texture) and consistency (consistent, inconsistent) as factors revealed a significant interaction ($F(1,9)=7.18$, $p=0.025$, $\eta_p^2=0.444$), reflecting a larger consistency effect for the upright silhouette condition than for the color/texture condition. Paired-samples t-tests revealed a significant consistency effect for the upright silhouette condition ($T(9)=3.82$, $p=0.004$, $d=0.44$) but not for the color/texture condition ($T(9)=1.05$, $p=0.32$, $d=0.06$). These results indicate that subjects' attention was captured by consistent prime images in the upright silhouette condition only.

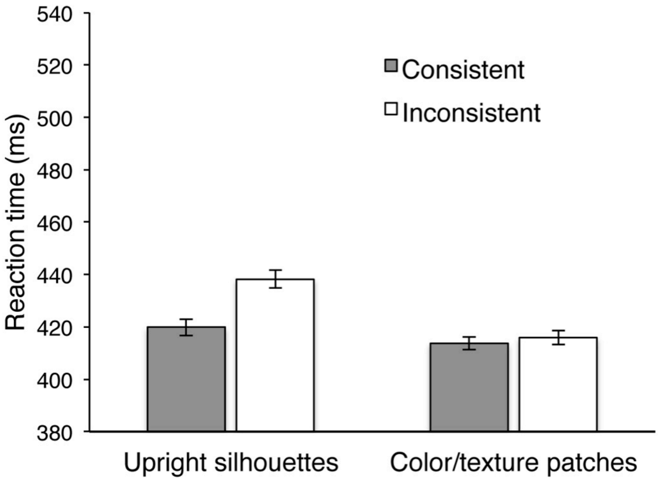


Figure 3. Reaction time (RT) for consistent and inconsistent trials in the upright silhouette and color/texture conditions of Experiment 1. Error bars represent the standard error of the mean after adjusting for between-subjects variance (Loftus & Masson, 1994).

Accuracy

A repeated-measures ANOVA with prime type (silhouette, color/texture) and consistency (consistent, inconsistent) as factors did not reveal a significant interaction, $F(1,9)=1.55$, $p=0.244$, $\eta_p^2=0.147$. There was a main effect of prime type: subjects performed the prime task with higher accuracy in the color/texture condition (98.6%) than in the upright silhouette condition (96.5%; $F(1,9)=12.27$, $p=0.007$, $\eta_p^2=0.577$). A main effect of consistency did not reach significance, $F(1,9)=3.61$, $p=0.09$, $\eta_p^2=0.286$.

Recognition of color/texture patches

To ensure that the color/texture patches we used could be recognized as belonging to either cars or people, a subset of subjects (N=8) performed a color/texture discrimination task after the main experiment. Subjects were shown 96 pairs of color/texture patches that appeared in the prime task, and were required to respond on each trial whether the patch associated with a car (N=4) or person (N=4) appeared on the left or right, using the left and right arrow keys. The position and size of the patches were identical to those used in the prime task. Color/texture patches remained on screen until subjects made a decision, which triggered the onset of the next pair of images. Accuracy on this task ranged from 85.4-94.8% with a mean of 90.9%, which was significantly higher than chance (50%; $T(7)=30.37$, $p=0.0001$, $d=1.92$).

Discussion

Results from Experiment 1 indicate that the current paradigm can successfully reveal the attentional capture effect of task-irrelevant stimuli when those stimuli contain features that match the search template. Subjects experienced a significantly stronger capture effect by upright silhouettes compared to color/texture patches, as indicated by significant RT differences between consistent and inconsistent trials in the silhouette condition only. These results suggest that the search template may be dominated by object shape information rather than color and texture, although it cannot be ruled out that color and texture may be activated in the template when manipulated in other ways.

Thus far, it is unclear whether the shape representations active in the template are restricted to canonical, upright orientations (as used in Experiment 1), or whether they are view-invariant such that unexpected or unnatural orientations of objects (e.g., inverted images) may still capture attention. The generalization of object shape to unexpected orientations would suggest that the search template is composed of combinations of local features (e.g., arms, legs) that are not canonically grounded. We compared prime task performance between upright and inverted silhouettes in Experiment 2 to explore this possibility.

2.4.2 Experiment 2: Upright vs. inverted (180°) silhouettes

The task in Experiment 2 was the same as in Experiment 1. Half of prime trials showed upright silhouettes and the other half showed silhouettes rotated 180 degrees (Figure 1b).

RT

Figure 4 depicts mean RT for consistent and inconsistent trials in the upright and inverted silhouette conditions. A repeated-measures ANOVA with prime type (upright, inverted) and consistency (consistent, inconsistent) as factors did not reveal a significant interaction ($F(1,15)=2.47$, $p=0.137$, $\eta_p^2=0.141$), indicating a comparable consistency effect for upright and inverted silhouette conditions. There was a main effect of consistency ($F(1,15)=16.1$, $p=0.001$, $\eta_p^2=0.518$), with faster responses on consistent trials compared to inconsistent trials. There was no main effect of prime type, $F(1,15)=1.41$, $p=0.254$, $\eta_p^2=0.086$. These results suggest that attention was captured by consistent prime images regardless of image orientation (upright or inverted).

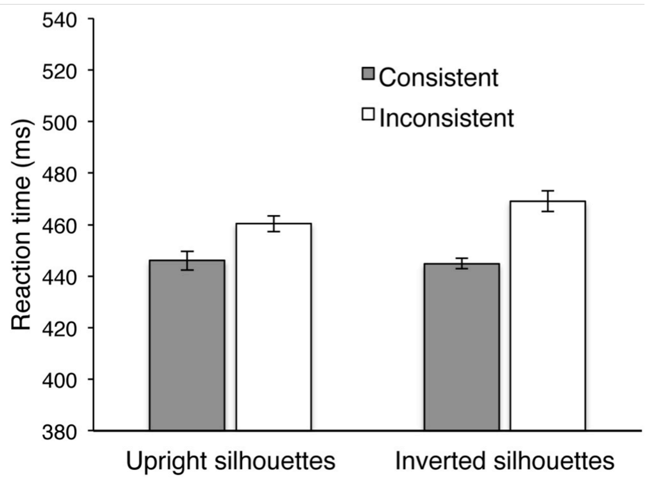


Figure 4. Reaction time (RT) for consistent and inconsistent trials in the upright and inverted silhouette conditions of Experiment 2. Error bars represent the standard error of the mean after adjusting for between-subjects variance.

Accuracy

A repeated-measures ANOVA with prime type (upright, inverted) and consistency (consistent, inconsistent) as factors revealed a significant interaction ($F(1,15)=9.5$, $p=0.008$, $\eta_p^2=0.388$), reflecting a larger consistency effect in the upright silhouette condition (98.8% for consistent trials vs. 89.6 % for inconsistent trials) than the inverted silhouette condition (98.3% for consistent trials vs. 92.9% for inconsistent trials). Paired-samples t-tests revealed a significant consistency effect for both the upright silhouette condition ($T(15)=3.9$, $p=0.001$, $d=1$) and the inverted silhouette condition ($T(15)=2.85$, $p=0.012$, $d=0.73$).

Discussion

Results from Experiment 2 suggest that object shape is part of the active search template regardless of canonical orientation. This may indicate that the template for natural search is, to an extent, composed of view- and orientation-invariant shape features. Alternatively, it is possible that the search template consists mainly of simple orientation features: cars typically have many horizontally oriented features while people typically have many vertically oriented features. These category-related orientation features were largely maintained in the inverted silhouette condition of Experiment 2. Thus, in Experiment 3, we presented 90°-

rotated silhouettes on half of prime trials so that people appeared along a horizontal plane and cars appeared along a vertical plane.

2.4.3 Experiment 3: Upright vs. rotated (90°) silhouettes

The task in Experiment 3 was the same as in Experiments 1 and 2. Half of prime trials showed upright silhouettes and the other half showed silhouettes rotated clockwise by 90 degrees (Figure 1b).

RT

Figure 5 depicts mean RT for consistent and inconsistent trials in the upright and rotated silhouette conditions. A repeated-measures ANOVA with prime type (upright, rotated) and consistency (consistent, inconsistent) as factors did not reveal a significant interaction ($F(1,12)=0.53$, $p=0.482$, $\eta_p^2=0.042$), indicating a comparable consistency effect for upright and rotated silhouette conditions. There was a main effect of consistency ($F(1,12)=16.18$, $p=0.002$, $\eta_p^2=0.574$), with faster responses on consistent trials than inconsistent trials. Additionally, there was a main effect of prime type ($F(1,12)=7.31$, $p=0.019$, $\eta_p^2=0.379$), with faster responses in the rotated silhouette condition compared to the upright silhouette condition.

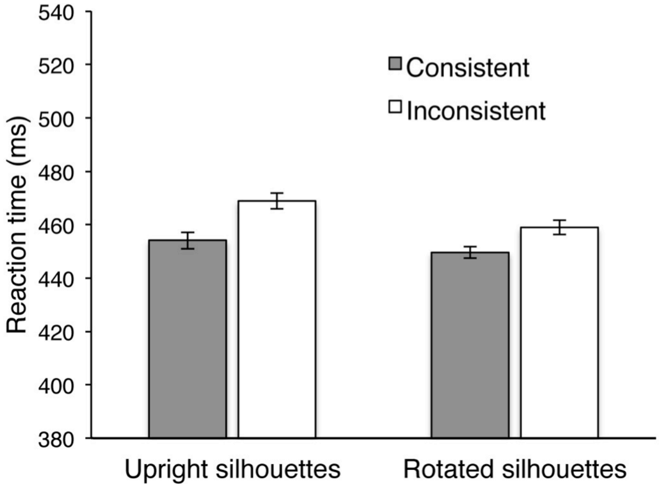


Figure 5. Reaction time (RT) for consistent and inconsistent trials in the upright and rotated silhouette conditions of Experiment 3. Error bars represent the standard error of the mean after adjusting for between-subjects variance.

Accuracy

A repeated-measures ANOVA with prime type (upright, rotated) and consistency (consistent, inconsistent) as factors revealed no significant interaction, $F(1,12)=0.0$, $p=0.995$, $\eta_p^2=0.00$. There was a main effect of consistency ($F(1,12)=8.5$, $p=0.013$, $\eta_p^2=0.415$), with higher accuracy on consistent trials (98.9%) than inconsistent trials (97.2%). There was no main effect of prime type, $F(1,12)=0.0$, $p=0.997$, $\eta_p^2=0.00$.

Discussion

Results of Experiment 3 suggest that the search template is made up of object shape information that is not dependent on simple vertical and horizontal discrimination when searching for people and cars, respectively. This is further evidence that the search template consists of view- and orientation-invariant representations of object shape. However, we have not yet determined whether that shape information is based on the whole object or on object parts. The involuntary capture of objects regardless of orientation suggests that searchers activate representations of diagnostic object parts independent of their global layout. We conducted Experiment 4 to directly address the possibility of a part-based template.

2.4.4 Experiment 4: Whole silhouettes vs. silhouette parts

The task in Experiment 4 was the same as in Experiments 1-3. Half of prime trials showed whole upright silhouettes and the other half showed diagnostic parts of those silhouettes (see Figure 1b). Object parts, on average, were made up of about 15% of the whole silhouettes (see Methods).

RT

Figure 6 depicts mean RT for consistent and inconsistent trials in the whole silhouette and silhouette parts conditions. A repeated-measures ANOVA with prime type (whole silhouettes, silhouette parts) and consistency (consistent, inconsistent) as factors did not reveal a significant interaction ($F(1,12)=0.012$, $p=0.915$, $\eta_p^2=0.001$), suggesting a comparable consistency effect for whole silhouettes and silhouette parts. There was a main effect of consistency, with significantly faster responses on consistent trials than on inconsistent trials, $F(1,12)=6.39$, $p=0.027$, $\eta_p^2=0.347$. There was no main effect of prime type, $F(1,12)=0.485$, $p=0.499$, $\eta_p^2=0.039$.

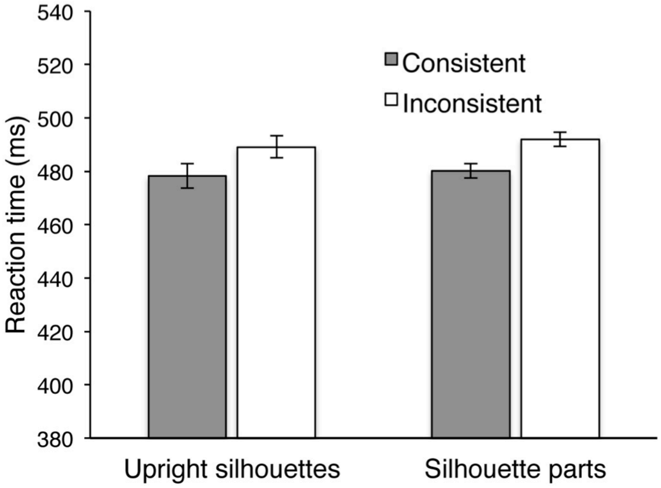


Figure 6. Reaction time (RT) for consistent and inconsistent trials in the upright silhouette and silhouette parts conditions of Experiment 4. Error bars represent the standard error of the mean after adjusting for between-subjects variance.

Accuracy

A repeated-measures ANOVA with prime type (whole silhouettes, silhouette parts) and consistency (consistent, inconsistent) as factors revealed a significant interaction ($F(1,12)=5.21$, $p=0.042$, $\eta_p^2=0.303$), with a larger consistency effect in the whole silhouette condition (98.5% for consistent trials vs. 93.6% for inconsistent trials) compared to the silhouette parts condition (98.9% for consistent trials vs. 98.3% for inconsistent trials). Paired-samples t-tests revealed a significant consistency effect for the whole silhouette condition ($T(12)=2.76$, $p=0.017$, $d=0.96$) but not for the silhouette parts condition ($T(12)=0.74$, $p=0.47$, $d=0.35$).

Recognition of object parts

In addition to the main experiment, a subset of subjects (N=8) performed a discrimination task after the experiment to ensure that the object parts were recognized as belonging to cars or people. This task was identical to that described for color/texture patches in Experiment 1, except that subjects were shown 96 pairs of object parts instead of color/texture patches. Accuracy on this task ranged from 95.8-99% with a mean of 97.1%, which was significantly higher than chance (50%, $T(7)=109.86$, $p<0.0001$, $d=1.94$). These results indicate that the object parts we used in this experiment could be reliably recognized as belonging to people or cars.

Discussion

Results from Experiment 4 suggest that the search template is composed of a flexible layout of diagnostic object parts (e.g., an arm connected to a torso) rather than global shape (e.g., a person standing in the center of the frame), a finding that complements the results of Experiments 2 and 3.

2.4.5 Experiment 5: Differing locations of scenes and primes

Previous research on contingent attentional capture has found evidence that attending to low-level features can result in the capture of attention by these features even if presented at task-irrelevant locations (Folk, et al., 2002). We conducted Experiment 5 to examine whether this was similarly the case for the capture effects observed in our paradigm. Such a result would provide evidence that the search template is spatially global in its representation of category-level object parts.

The search task differed from Experiments 1-4 in that subjects now had to respond whether a car or person appeared in a scene above or below fixation using the up and down arrow keys. The prime task remained the same as in the previous experiments, in which subjects were required to ignore prime images presented to the left and right of fixation and to respond whether a dot probe appeared on the left or right using the left and right arrow keys. Prime images were all upright silhouettes, with each image repeated once in the experiment. Because only one type of prime was presented, paired samples t-tests were conducted for consistent versus inconsistent trials on RT and accuracy results.

RT

Figure 7 depicts mean RT for consistent and inconsistent trials in the prime task. A paired-samples t-test on consistency revealed that subjects responded significantly faster on consistent trials than inconsistent trials, $T(10)$=4.76, p=0.001, d=0.26.

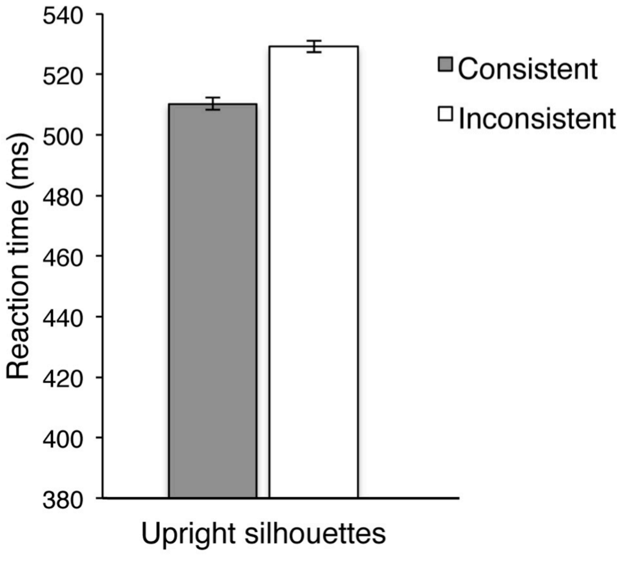


Figure 7. Reaction time (RT) for consistent and inconsistent trials in the upright silhouette condition of Experiment 5, in which silhouettes were presented at search-task-irrelevant locations. Error bars represent the standard error of the mean after adjusting for between-subjects variance.

Accuracy

A paired-samples t-test revealed that subjects performed the prime task with similar accuracy on consistent (99.4%) and inconsistent trials (98.6%), $T(10)= 1.58$, $p=0.145$, $d=0.62$.

Discussion

The results of Experiment 5 suggest that the template activated for our search task was spatially global, modulating the processing of visual input across the visual field even at locations that were known to be irrelevant for the search task.

Attentional capture for car vs. person primes

To investigate whether the reported consistency effects differed for car and person primes, we categorized trials as belonging to “person” or “car” based on the correspondence in the location of the prime category and the subsequent dot probe. For example, trials in which the dot probe followed the car prime were labeled as car trials. Cue consistency was then determined by the category of the preceding cue. For example, an inconsistent car trial would be a trial with a person cue and with the dot probe appearing at the location of the car prime. Using this method, we split capture trials into car and person categories and explored the consistency effects for upright silhouettes combined across Experiment 1-5 (N=59; RT was averaged across experiments for subjects who participated in two experiments (N=4)). A repeated-measures ANOVA with category (cars, people) and consistency (consistent, inconsistent) as factors revealed, as expected, a highly significant main effect of consistency ($F(1,58)=26.4$, $p<.001$, $\eta_p^2=.313$). There was also a significant interaction between category and consistency ($F(1,58)=6.92$, $p=.011$, $\eta_p^2=.107$), with somewhat stronger consistency effects for person than car trials. Importantly, however, the consistency effect was significant for both person ($T(58)=5.85$, $p<.001$, $d=1.54$) and car ($T(58)=2.70$, $p=.009$, $d=.71$) conditions separately, indicating that both categories contributed to the overall consistency effects observed.

2.5 General Discussion

What do people look for when searching for an object category in a natural scene? We developed a novel attentional capture paradigm to explore this question. On the majority of trials, subjects searched for people or cars in real-world scenes. On a subset of trials, the search cue was followed by task-irrelevant stimuli instead of scenes, directly followed by a dot that subjects were instructed to detect. Attentional capture was defined as the RT difference in detecting a dot presented at the location of the consistent, putatively template-matching stimulus, versus the location of the inconsistent stimulus. In Experiment 1, there was a strong capture effect for upright silhouettes of people and cars, but not for color/texture patches extracted from those same objects. This is evidence that the search template in our study was predominantly composed of object shape. Experiment 2 was conducted to test the necessity for canonical orientation of whole object shape, and we found that objects can be inverted and still elicit a comparable capture effect to upright images, suggesting that the representations activated in the search template are orientation-invariant. In Experiment 3, we presented cars and people rotated by 90 degrees to rule out the possibility that searchers may prepare for targets based on low-level orientation features (i.e., preparing for cars and people by looking for horizontally oriented and vertically oriented objects, respectively). Results indicated that cars presented along a vertical plane and people presented along a horizontal plane nevertheless captured attention, providing further evidence for an orientation-invariant search template for object form. Following from these findings, we hypothesized that the search template likely consists of a collection of diagnostic object parts rather than representations of whole objects, as this would allow searchers to prepare more flexibly for varied category exemplars in complex scenes. Experiment 4 confirmed this hypothesis, showing significant capture effects by various object parts (e.g., arms, feet, a car tire) that consisted of only about 15% of the pixels of the whole silhouette. Finally, in Experiment 5 we showed that silhouettes capture attention even when they are presented at locations that are irrelevant to the search task, indicating that the search template for this task is spatially global.

Results of Experiment 5 showed that attentional capture occurred even at locations that were fully irrelevant for the search task (i.e., subjects never had to detect objects at these locations). Despite this, we found strong category-specific attentional capture at these locations. This is evidence that the search template spreads to locations outside of where search is performed, suggesting a category-specific attentional bias across the visual field, as previously observed using neuroimaging methods (Peelen, et al., 2009). This is reminiscent of the effects observed for feature-based attention (Serences & Boynton, 2007). Indeed, behavioral work has shown that familiar object categories can be detected in the absence of focal attention (Li, et al., 2002), similar to low-level features (Treisman & Gelade, 1980). To account for these behavioral results, Treisman (2006) suggested that subjects "may be set to sense, in parallel, a highly overlearned vocabulary of features that characterize a particular semantic category" when searching for familiar object categories in natural scenes. In support of this account, a recent study (Korjoukov et al., 2012) found that subjects were capable of detecting object categories in briefly presented natural scenes, while they were much worse at grouping sets of features into a coherent whole, which suggests that more focused attention is needed to perform perceptual grouping than to detect familiar object categories.

Altogether, the results of the current study suggest that the template for object category search in natural scenes is made up of view-invariant, category-diagnostic shapes of object parts represented globally across the visual field. Such a template likely reflects the ways in which the visual and attentional systems optimally deal with the complexities of search in the real world; natural scenes are cluttered and highly detailed, targets may share many low-level features with non-targets, and features may vary widely even within the target category. The key to accurate detection is forming a search template that is flexible enough to account for versatile targets, but specific enough to eliminate similar non-targets. For example, forming a representation that encompasses all vertically oriented objects may lead to incorrect detection of a lamp post as a person, while forming a holistic representation of a person in a prototypical standing posture is only diagnostic of a subset of people, in which case people shown in other stances may be overlooked. But a template composed of spatially non-specific shapes of body parts is sufficiently diagnostic of people in many different views and locations, and at the same time eliminates the risk of lower-level (e.g., orientation-based) identification errors. This is consistent with the computational results of Ullman et al. (2002) and behavioral results of Delorme et al. (2010), which both concluded that variable objects are optimally classified by such diagnostic part features.

The current study explored the components of the search template for the detection of people and cars; however, in daily life we also search for objects at more specific or general levels, such as looking for your own car versus looking for any kind of vehicle, respectively. It would be interesting for future studies to use the current paradigm to directly compare the contents of the search template for these different levels of search. We expect that if the exact identity of the target is known prior to search (e.g., a red scarf), or if a certain feature dimension reliably discriminates the target from distractors (e.g., the color red), then search based on low-level features (e.g., color) may be most efficient for that task (Pomplun, 2006; Vickery, et al., 2005; Wolfe, et al., 2004). Similarly, for object categories for which surface features are diagnostic of their presence (e.g., trees, for which the color green is a consistent feature across exemplars), such features are likely part of the search template for these categories. For search tasks that are broader than those in the current study (e.g., detecting vehicles rather than cars), a shape-based template may not always be advantageous for target detection, as members of the same superordinate category may share few shape features. In this case, search preparation could be specified at higher levels of the visual processing hierarchy, possibly at the conceptual level (Wyble, Folk, & Potter, 2012). Collectively, these findings provide evidence for a "flexible template", which may change depending on task demands (Bravo & Farid, 2009, 2012). That the template may depend on the specific task demands also highlights the value of using natural scenes, as it is hard to mimic the particularities of the real world using artificial search arrays (Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011). It may therefore also be beneficial for future studies to investigate whether the current results extend to active search tasks that involve eye movements (Findlay & Gilchrist, 2003).

Finally, an interesting avenue for future research is to explore individual differences in the contents of the search template, and how these relate to differences in search performance. There are considerable individual differences in the ability to detect object categories in briefly presented real-world scenes (e.g., Peelen & Kastner, 2011), which are stable over time, and are only partly explained by general traits such as intelligence (Huang, Mo, & Li, 2012). These differences have been found to relate to differences in self-reported search strategy, with good searchers reporting to use a more general strategy (e.g., preparing for cars and people at multiple angles) and poor searchers reporting to use a more specific strategy (e.g., preparing for cars and people in prototypical orientations; Peelen & Kastner, 2011). It may be of interest to apply the current paradigm to studies of individual differences in the template, revealing the most effective strategy for a given search task by comparing the templates of good and poor searchers. This avenue may be particularly useful when applied to professions in which search is of high practical relevance, such as radiology, airport security, emergency services, and the military.

2.5.1 General Conclusions

Here we presented a novel attentional capture paradigm to explore the contents of the search template for familiar object categories in real-world scenes. The results of the current study indicate that such a template is made up of spatially global, view-invariant shapes of diagnostic object parts. Our paradigm can be adopted to explore the templates for various search tasks and individual strategies. These investigations could be of use to professionals whose search performance and efficiency are of high practical importance.

Chapter three: Involuntary attentional capture by task-irrelevant objects that match the search template for category detection in natural scenes

Reshanne R. Reeder[a], Wieske van Zoest[a], Marius V. Peelen[a]

[a]Center for Mind/Brain Sciences (CIMeC), University of Trento, 38068 Rovereto (TN), Italy

Acknowledgments

The research was funded by the Autonomous Province of Trento, Call "Grandi Progetti 2012", project "Characterizing and improving brain mechanisms of attention - ATTEND". No conflicts of interest, financial or otherwise, are declared by the authors.

3.1 Abstract

Visual search is thought to be guided by top-down attentional sets, or "search templates". We recently studied the contents of the search template for category-level search in natural scenes (Reeder & Peelen, 2013). On most trials, subjects searched for categorical targets (cars or people) in natural scenes, but on a subset of trials, a dot-probe detection task was used to measure attentional orienting to task-irrelevant stimuli. The results showed a consistency effect on dot-probe trials, such that probe detection was faster for probes preceded by features of search targets relative to not-target features. Here we test whether this consistency effect was due to attentional capture by stimuli matching the active visual search template, or alternatively whether it reflected voluntary orienting to objects semantically related to the cued category. Experiment 1 showed that the consistency effect on dot-probe trials was specific to conditions in which subjects actively maintained a category-diagnostic search template, indicating that it did not reflect general effects of semantically encoding the category cues. In Experiment 2, the paradigm was modified so that attending to the cued category in the probe task would be detrimental to performance. Although subjects were made aware of this, they continued to orient to task-irrelevant but template-matching stimuli, which is evidence for automatic attentional capture contingent on the search template. Together, these results provide evidence for involuntary attentional capture by stimuli that visually match the active search template for category-level search in real-world scenes.

3.2 Introduction

The selection of a target amongst distractors involves the matching of incoming visual input to a top-down attentional set or "search template". It has been hypothesized that the search template guides visual attention to items in the environment that contain task-relevant features (Duncan & Humphreys, 1989; Wolfe, Cave, & Franzel, 1989), thereby facilitating target detection. Although it is widely accepted that top-down templates influence search behavior, there are many unanswered questions about the characteristics of such templates. Recently, studies have started to investigate the functional role and neural basis of search templates in naturalistic category-level search (for a review, see Peelen & Kastner, 2014). Nevertheless, little is known about the contents of the category-level search template for naturalistic targets, which must be composed of features that generalize across varied category exemplars (Bravo & Farid, 2012; Yang & Zelinsky, 2009) and at the same time exclude non-target objects that may share many low-level features with the target category.

Several previous studies have found that a search template can increase the priority of template-matching features in the environment, which can result in attentional capture by those features (often termed "contingent attentional capture"; Ariga & Yokosawa, 2008; Downing, 2000; Folk, Remington, & Johnston, 1992; Lien, Ruthruff, Goodin, & Remington, 2008). Following from these findings, we recently designed a contingent attentional capture paradigm to investigate the features that capture attention during naturalistic, category-level visual search (Reeder & Peelen, 2013). In this study, subjects were cued to detect people and cars in photographs of real-world scenes. On a subset of trials (dot-probe trials), instead of scenes, task-irrelevant features of people and cars appeared to the left and right of a centrally presented fixation cross. Subjects were instructed to ignore these stimuli and indicate the location of a subsequent dot probe that could appear to the left or right. This brief (100 ms) dot probe appeared quickly (117 ms) after the onset of the people and car features. We hypothesized that task-irrelevant features that matched the active search template would draw attention, resulting in faster detection of the dot probe when it was presented in the same location as a template-matching stimulus (consistent trials) than when it was presented in the opposite location (inconsistent trials).

We found consistency effects for a range of object features presented on dot-probe trials (e.g., car silhouettes during car search). Assuming that these consistency effects reflected attentional capture by features matching the active search template (Downing, 2000; Folk et al., 1992), we varied the object features on dot-probe trials to gain insight into the contents of the search template. However, we did not directly test whether the observed consistency effects truly reflected attentional capture by stimuli matching the search template. This was the goal of the present study.

We tested two predictions that were implied in the interpretation of our previous results (Reeder & Peelen, 2013). In Experiment 1 of the current study, we tested whether the consistency effect on dot-probe trials reflects orienting to stimuli that match the active search template or alternatively reflects more general effects of encoding and maintaining the preceding category cue; for example, it cannot be excluded that subjects attended to the car silhouettes simply because they held the word “car” in mind. Experiment 1 of the current study was conducted to investigate whether contingent attentional capture would also occur when subjects were instructed to perform a car- or person-related task without the need for a car- or person-diagnostic search template. In some blocks, instead of searching for cars or people, subjects were instructed to search for objects semantically related to cars or people. This “related object” search condition required preparation for a variety of objects that are united by their semantic association with the cued category but have virtually no overlapping diagnostic shape features; in other words, a visual template could not be activated in this condition. Importantly, the cueing procedure and the dot-probe task was the same for both search conditions. If the consistency effect can be fully explained by general effects of the cue (i.e., the semantic encoding of the car or person cue), we would expect similar consistency effects when subjects searched for category-related objects as when they searched for category exemplars.

A second prediction tested in the current study is that the consistency effect on dot-probe trials reflects contingent attentional capture rather than voluntary orienting to cue-matching stimuli. Experiment 2 tested this prediction by making the target-matching stimuli on dot-probe trials counter-predictive of the location of the dot probe. Subjects were informed that attending to the cued category on dot-probe trials would impair task performance. Furthermore, the search task stimuli and the dot-probe task stimuli were always and consistently presented in different, non-overlapping, locations. If the consistency effect reflects voluntary orienting rather than contingent attentional capture, subjects should be able to ignore the dot-probe task stimuli when this impairs performance. If, however, we continue to observe a consistency effect on dot-probe trials following these instructions, it would provide evidence that consistency effects reflect contingent attentional *capture*.

3.3 Experiment 1

Experiment 1 was conducted to determine whether attentional capture by silhouettes of a cued category requires a category-diagnostic search template. Results of our previous study (Reeder & Peelen, 2013) suggested that the search template is composed of diagnostic shape features of targets, but we could not rule out the possibility that the consistency effects observed reflected a more conceptual preference for the target category, perhaps as the result of semantically encoding the cue unrelated to search-specific effects. To disentangle specific search template effects from other cue-based priming effects, subjects in Experiment 1 were cued to search for objects conceptually related to people and cars (related object search condition) or for people and cars in natural scenes (natural scene search condition, i.e., a replication of Experiment 5 in Reeder & Peelen, 2013), in separate blocks. Stimuli in the related object search condition comprised objects such as wristwatches, handbags, and sunglasses for people, or gas cans, engines, and seatbelts for cars (Figure 1). Dot-probe trials with car and person silhouettes made up 25% of trials in both search conditions (Figure 2).

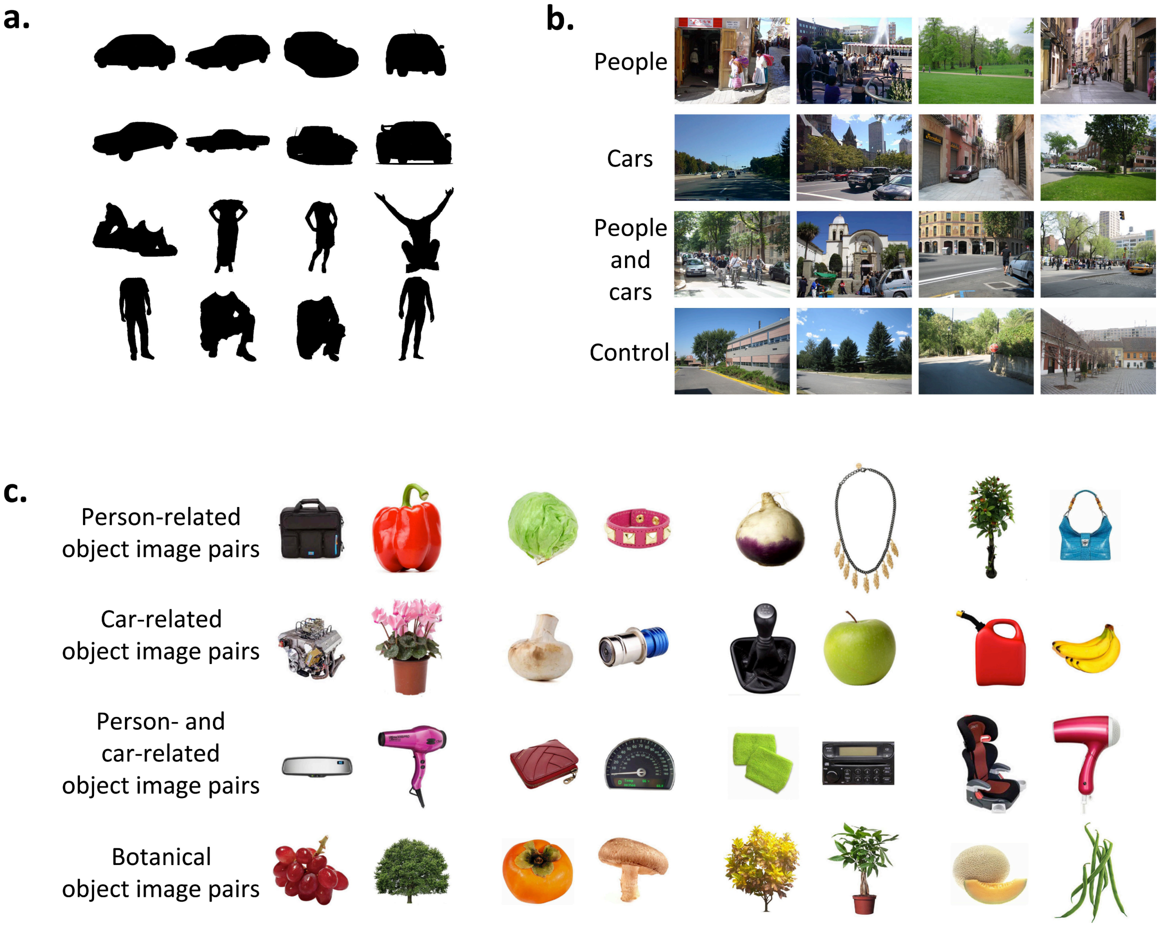


Figure 1. a) Examples of attentional capture stimuli in all experiments (upright silhouettes). b) Examples of natural scene search stimuli in Experiments 1 and 2. c) Examples of related-object search stimuli in Experiment 1.

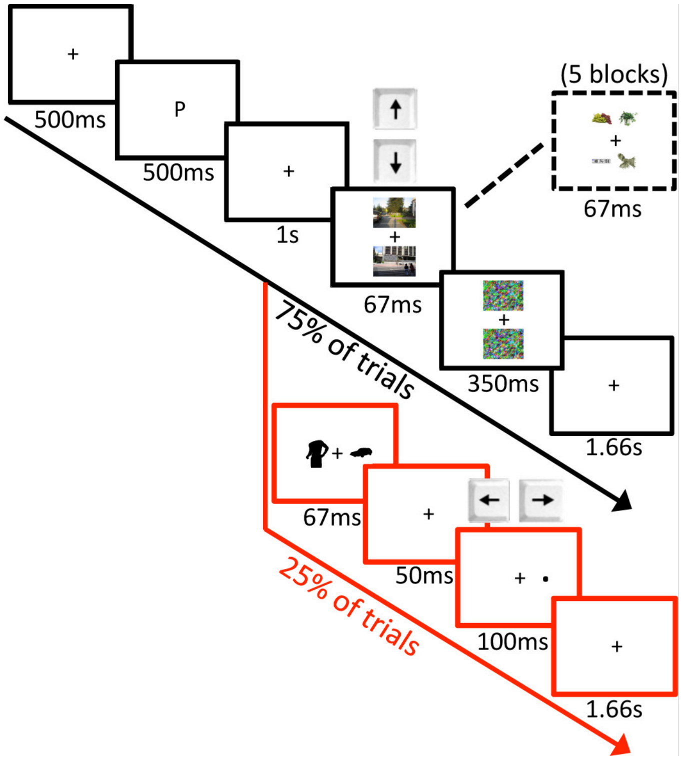


Figure 2. A schematic of the paradigms used in Experiments 1 and 2; in Experiment 1, 5 blocks of trials corresponded to natural scene search, while the other 5 blocks corresponded to related-objects search. In Experiment 2a, dot-probe trials were inconsistent 68.75% of the time and subject were told that attending to the silhouette on dot-probe trials would harm performance. In Experiment 2b, subjects were given the same probability of inconsistent trials as in Experiment 2a, except now they were explicitly told that the dot probe would most likely appear on the same side of fixation as the uncued category. Additionally, category cues were orange and blue circles instead of letters in Experiment 2b. Otherwise, the experimental paradigm was the same as in the natural scene search condition of Experiment 1.

If general processes related to the encoding of the category cues account for the consistency effect on dot-probe trials, a consistency effect should occur independently of the specific search task and should therefore also be observed in the related object search condition. If, however, only active search with a diagnostic search template drives contingent attentional capture, we would expect to find a consistency effect in the natural scene search condition only.

3.4 Methods

3.4.1 Subjects

13 undergraduate and graduate students from the University of Trento (12 women) participated in the experiments for payment. All subjects had normal or corrected-to-normal visual acuity and were between the ages of 19 and 25 years (mean age = 21.7 years). The research protocols in all experiments adhered to the ethical principles of the Declaration of Helsinki.

3.4.2 Stimuli

All stimuli were presented on a 19-inch Dell 1905 FP monitor with a screen resolution of 1280 x 1024 pixels and 60Hz refresh frequency (Dell Inc., Round Rock, TX). Subjects sat 57cm from the screen. Stimuli were presented using A Simple Framework (Schwarzbach, 2011), a toolbox based on the Psychophysics Toolbox for MATLAB (The MathWorks, Inc., Natick, MA).

A fixation cross and uppercased letter cues appeared centered on the screen in 70-point “strong” Times New Roman font. The fixation cross had dimensions of

31 x 31 pixels subtending 0.92° in height and width, and letters had dimensions of 70 x 70 pixels subtending 2.1° in height and width.

Probe task stimuli

Stimuli presented in the dot-probe trials (25% of trials) were composed of 160 upright silhouettes made from photographs of cars (80) and people (80) without scene background (see Figure 1a). Most images were obtained from free-access online image sources and were chosen to encompass a variety of viewpoints and features (e.g., a crouching child, a woman standing, a pickup truck as seen from behind, a sedan as seen from the side). Heads were removed from all images of people to be consistent with the previous behavioral study on which these experiments are based (Reeder & Peelen, 2013).

None of the stimuli presented in the dot-probe trials were shown in the search task. Stimuli could appear in three possible sizes (100 x 100, 180 x 180, or 200 x 200 pixels, or 2.96 x 2.96°, 5.32 x 5.32°, or 5.91 x 5.91° of visual angle, respectively) and in three different screen locations along the X-axis, subtending 6.47°, 7.99°, or 10.76° of visual angle. Size and location values were chosen randomly on each trial and independently for the left and right stimulus. On each dot-probe trial, a single car and person silhouette appeared to the left and right of fixation, and cars and people appeared an equal number of times on either side. Each image was repeated once in an experiment.

Natural scene stimuli

Stimuli presented in the search trials (75% of trials) were 240 color photographs of real-world scenes obtained from the LabelMe online database (Russell, Torralba, Murphy, & Freeman, 2008; see Figure 1b for some examples) and were divided into scenes containing cars (120), people (120), both cars and people (120), or neither cars nor people (120). Two scenes appeared on every trial and no scene was repeated within the experiment.

Scenes were scaled to 548 x 411 pixel resolution, subtending a visual angle of 16.08 x 12.18°. Scenes were presented 7.41° from the center of the screen to the center of the image, above and below fixation.

Car- and person-related object stimuli

Stimuli presented in the search trials (75% of trials) were color photographs of person-related objects (e.g., wristwatches, backpacks, sunglasses, hats), car-related objects (e.g., gas cans, GPS, air fresheners, seatbelts), and botanical objects (e.g., fruits, vegetables, trees, flowers; see Figure 1c for some examples). Each display was made up of two image pairs that each consisted of one car-related object and one person-related object (120), one car-related object and one botanical object (120), one person-related object and one botanical object (120), or two botanical objects (120). No image pair was repeated within the experiment.

Image pairs were scaled to fit within a boundary that was 548 x 411 pixel resolution subtending a visual angle of 16.08 x 12.18°, and were presented 7.41° from the center of the screen to the center of the image pair, above and below fixation.

3.4.3 Procedure

All subjects (N=13) took part in one practice block followed by 10 blocks of 64 trials each. Each block was made up of two tasks as illustrated in Figure 2: a search task on 75% of trials and a dot-probe task on 25% of trials. In 5 blocks the search task was made up of natural scene stimuli (natural scene search condition), and in the other 5 blocks the search task was composed of car- and person-related object stimuli (related object search condition). Each search condition was performed in 5 consecutive blocks of trials, with the order of the search conditions counterbalanced across subjects. Trials were randomized within a block so subjects could not predict whether the search task or dot-probe task would appear on any given trial.

The search task and dot-probe task both started with the presentation of a fixation cross for 500 ms, followed by a single letter for 500 ms: “P” for “persona” or “M” for “macchina” (the Italian words for person and car, respectively). After the letter, another fixation cross appeared for 1 s.

On dot-probe trials, following the fixation cross, two images (upright silhouettes of a person and a car) appeared to the left and right of fixation for 67 ms. People and cars were equally likely to appear on the left or right. A 50-ms fixation succeeded the silhouettes, after which a dot probe appeared for 100 ms 8.52° from the center of the screen on the left or right. Subjects were required to respond whether the dot probe appeared on the left or right of fixation using the “left” and “right” arrow keys, respectively. They were instructed to ignore the silhouettes that appeared prior to the probe. Silhouettes did not predict the dot-probe's location, and the dot appeared on the left and right an equal number of times.

On search trials, if the blocks corresponded to the natural scene search condition, subjects would see two photographs of natural scenes for 67 ms, followed by a 350-ms mask. Subjects were required to respond whether the cued object category appeared above or below fixation using the "up" and "down" arrow keys, respectively. The two scenes that appeared could either be one containing cars and the other containing people, or one containing both cars and people and the other containing no cars or people. This structure allowed us to present people and cars on every trial without making the presence of one category informative of the absence of the other. Each of the four scene types appeared in each possible location an equal number of times above and below fixation.

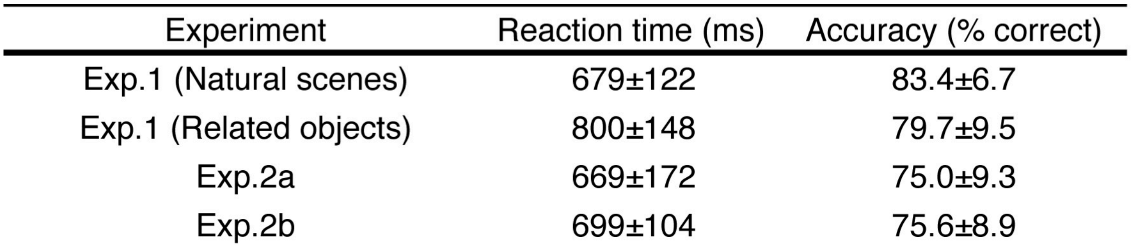
If the blocks corresponded to the related object search condition, subjects would see two image pairs instead of natural scenes for 67 ms, followed by a 350-ms mask. Subjects were required to respond whether the object conceptually related to the cued object category appeared above or below fixation using the "up" and "down" arrow keys, respectively. The two image pairs that appeared could either be one containing a car-related object presented with a botanical object and the other containing a person-related object with a botanical object, or one containing both car- and person-related objects and the other containing two botanical objects. Each of the four image pair types appeared in each possible location an equal number of times above and below fixation.

3.4.4 Analysis

We analyzed accuracy and reaction time (RT) for consistent and inconsistent trials in the dot-probe task. Only correct trials were included in the RT analysis. Subjects were excluded from analysis if their mean dot-probe task accuracy fell 2 standard deviations below the group mean for the experiment. One subject was excluded from Experiment 1 based on this criterion.

For the search task, we analyzed accuracy and RT. Only correct trials were included in the RT analysis. Results of the search task are reported in Table 1. The Results section reports the results of the dot-probe task only.

Table 1. Mean reaction time (ms) and accuracy (% correct), with standard deviation, in the search task for Experiments 1 and 2.



Experiment	Reaction time (ms)	Accuracy (% correct)
Exp.1 (Natural scenes)	679±122	83.4±6.7
Exp.1 (Related objects)	800±148	79.7±9.5
Exp.2a	669±172	75.0±9.3
Exp.2b	699±104	75.6±8.9

3.5 Results

3.5.1 Reaction time

Figure 3 illustrates mean RT for consistent and inconsistent trials in the dot-probe tasks associated with the natural scene search condition and the related-object search condition. A repeated-measures ANOVA with search condition (natural scene, related object) and consistency as factors revealed a significant interaction in which the consistency effect associated with the natural scene search condition was significantly greater than the consistency effect in the related object search condition, $F(1,11)=8.45$, $p=0.014$, $\eta_p^2=0.434$. Furthermore, there was a main effect of consistency, $F(1,11)=12.21$, $p=0.005$, $\eta_p^2=0.526$, driven by the consistency effect in the natural scene search condition, $t(11)=-3.64$, $p=0.004$, $d=-.261$, whereas there was no such effect in the dot-probe task associated with the related object search condition, $t(11)=-.31$, $p=0.76$, $d=-.012$.

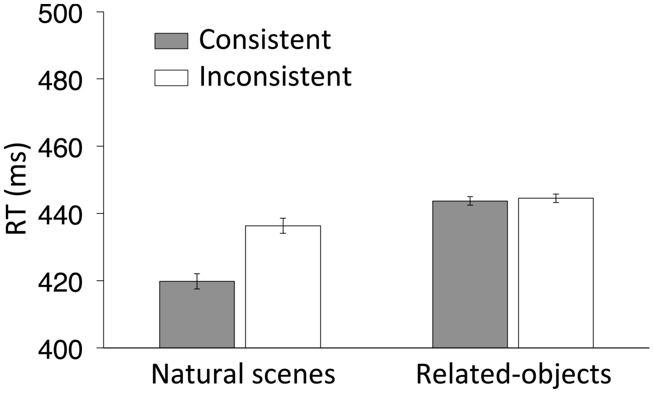


Figure 3. Reaction time (RT) for consistent and inconsistent trials in the dot-probe tasks of the natural scene and related-object search conditions of Experiment 1. Error bars represent the standard error of the mean after adjusting for between-subjects variance (Loftus & Masson, 1994).

3.5.2 Accuracy

A repeated-measures ANOVA with search condition and consistency as factors revealed no significant interaction for accuracy, $F(1,11)=0.371$, $p=0.555$, $\eta_p^2=0.033$, and no significant main effects of either search condition, $F(1,11)=0.084$, $p=0.777$, $\eta_p^2=0.008$, or consistency, $F(1,11)=0.805$, $p=0.389$, $\eta_p^2=0.068$. Subjects showed consistently high accuracy in all conditions.

3.5.3 Discussion

In two separate tasks in Experiment 1, subjects were required to search either for cars and people (natural scene search) or for objects conceptually related to cars and people (related object search). Reaction time analysis revealed a significant consistency effect for dot-probe trials intermixed in the natural scene search condition but not for (otherwise identical) dot-probe trials intermixed in the semantically related object search condition. These results indicate that the consistency effect on dot-probe trials likely reflects orienting to stimuli that match the active search template rather than more general effects of encoding the category cues.

The results of Experiment 1 further suggest that searching for car- or person-related objects does not necessarily lead to the orienting of visual attention to silhouettes of cars or people in the dot-probe task. This is in line with previous research showing that greater *visual* similarities between the cue and the target (and by association, the template and the target) lead to more efficient search performance (Schmidt & Zelinsky, 2009; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004). Similarly, in studies of overt selection, search-irrelevant distractors that look more similar to a target are more disruptive for efficient target selection compared to irrelevant distractors that are less similar to the target (van Zoest & Donk, 2006).

3.6 Experiment 2

The second experiment was conducted to explore whether the consistency effects on dot-probe trials reflect attentional capture by stimuli matching the template (i.e., contingent attentional capture) or instead reflect voluntary orienting to these stimuli. Specifically, if attention is fully under top-down control on dot-probe trials, subjects should be able to voluntarily guide their attention to uncued stimuli (e.g., a person silhouette on a car search trial) if this becomes beneficial for the task. In Experiment 2, subjects were required to perform a similar task to that of Experiment 5 of the previous study (Reeder & Peelen, 2013), with the exception that inconsistent trials occurred on 68.75% of dot-probe trials rather than 50%. In Experiment 2a, subjects were told prior to experimentation that paying attention to the cued category on dot-probe trials would harm performance; in Experiment 2b, subjects were explicitly informed of the higher probability of a dot probe appearing on the same side of fixation as the uncued category. It is possible that one set of instructions could more effectively induce subjects to change their preferences than the other; alternatively, a reliable consistency effect observed across both parts of Experiment 2 would boost experimental power and increase our confidence that the consistency effects represent an automatic capture effect contingent on the search template. As an additional control, the category cues in Experiment 2b were changed from letters to arbitrary colors, the former of which could potentially semantically prime attention to the cued category. If the consistency effect is due to voluntary allocation of attention, and if attending to the cued category were known to be disadvantageous to performance, then subjects should preferentially attend to the uncued category.

3.7 Methods

3.7.1 Subjects

26 undergraduate and graduate students from the University of Trento (23 women) participated in the experiments for payment: 12 in Experiment 2a and 14 in Experiment 2b. All subjects had normal or corrected-to-normal visual acuity and were between the ages of 18 and 33 years (mean age = 23.7 years). The research protocols in all experiments adhered to the ethical principles of the Declaration of Helsinki.

3.7.2 Stimuli

All stimuli were presented using the same hardware and software as in Experiment 1. A fixation cross (Experiment 2a and 2b) and uppercased letter cues (Experiment 2a) appeared with the same dimensions as in Experiment 1. Orange and blue circle cues (Experiment 2b) appeared centered on the screen, subtending 148 pixels (4.37°) in diameter.

Probe task stimuli

The same probe task stimuli were used in Experiments 1 and 2.

Natural scene stimuli

Natural scene stimuli were taken from the same database and were presented using the same dimensions as in Experiment 1. Twice as many images (480) were presented in Experiment 2 and could contain cars (240), people (240), both cars and people (240), and neither cars nor people (240).

3.7.3 Procedure

Experiment 2a

All subjects (N=12) except two in Experiment 2a were different from those that participated in Experiment 1. All subjects took part in one practice block followed by 9 blocks of 64 trials each. Each block was made up of two tasks: a search task on 75% of trials and a dot-probe task on 25% of trials. All trials in the search task corresponded to the natural scene search condition of Experiment 1. The dot-probe task followed the same procedure as in the previous experiment with one exception: the proportion of consistent dot-probe trials was 5/16 instead of 8/16. Consistent trials were those in which the cued category (e.g., the person silhouette following the ''P'' cue) appeared on the same side of fixation as the dot probe; inconsistent trials were those in which the cued category appeared on the opposite side of fixation. Therefore, the dot probe was inconsistent with the cued category on 68.75% of trials. Prior to experimentation, subjects were told that paying attention to the cued category on dot-probe trials would be detrimental to performance.

Experiment 2b

All subjects (N=14) except one were different from those that participated in Experiments 1 and 2a. Experiment 2b was the same as Experiment 2a with two exceptions: first, subjects were cued to search for cars and people with orange and blue colored circles; an orange circle cued cars for half of subjects and cued people for the other half of subjects. Second, prior to experimentation, subjects were explicitly informed that the dot on dot-probe trials would most likely appear on the same side of fixation as the uncued category.

Analysis

We analyzed accuracy and reaction time (RT) for consistent and inconsistent trials in the dot-probe task. Only correct trials were included in the RT analysis. Subjects were excluded from analysis if their mean dot-probe task accuracy fell 2 standard deviations below the group mean for the experiment. One subject was excluded from Experiment 2 based on this criterion.

For the search task, we analyzed accuracy and RT. Only correct trials were included in the RT analysis. Results of the search task are reported in Table 1. The Results section reports the results of the dot-probe task only.

3.8 Results

3.8.1 Experiment 2a: Letter cues and nonspecific instructions

Reaction time

Figure 4a illustrates mean RT for consistent and inconsistent dot-probe trials of Experiment 2a. A paired-samples t-test with consistency as the within-subjects factor revealed a strong consistency effect, $t(10)=-3.43$, $p=0.006$, $d=-.185$, in which subjects responded significantly faster on consistent trials compared to inconsistent trials.

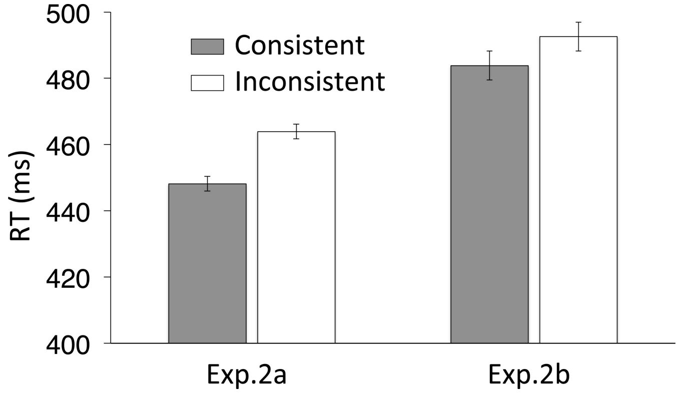


Figure 4. Reaction time (RT) for consistent and inconsistent trials in the dot-probe task for which 68.75% of trials were inconsistent and subjects were told that attending to the cued silhouette would be disadvantageous to performance (Experiment 2a) or that it would benefit performance to attend to the uncued category (Experiment 2b). Error bars represent the standard error of the mean after adjusting for between-subjects variance.

Accuracy

A paired-samples t-test with consistency as the within-subjects factor revealed that accuracy was not driven by consistency, *t*(10)=0.355, *p*=0.729, *d*=0.098.

3.8.2 Experiment 2b: Color cues and explicit instructions

Reaction time

Figure 4b illustrates mean RT for consistent and inconsistent trials on dot-probe trials of Experiment 2b (N=14). A paired-samples t-test with consistency as the within-subjects factor revealed a significant consistency effect, *t*(13)=-3.012, *p*=0.010, *d*=0.102, such that subjects responded faster to dot probes that appeared on the same side of fixation as the cued stimulus.

Accuracy

A paired-samples t-test with consistency as the within-subjects factor revealed that accuracy was not driven by an effect of consistency, *t*(13)=0.969, *p*=0.350, *d*=0.254.

3.8.3 Discussion

In Experiment 2, the location of the cued category exemplar was inconsistent with the location of the dot probe on 68.75% of trials, and subjects were either told that attending to the silhouettes would be detrimental to performance (Experiment 2a), or that attending to the uncued category would improve performance on dot-probe trials (Experiment 2b). Both versions of experimental instructions and both category cues (letters, colors) revealed significant consistency effects on dot-probe trials. This suggests that the consistency effects are driven by contingent capture by features that match the search template and are not due to voluntary allocation of attention to the silhouettes.

3.9 General Discussion

We conducted two experiments to investigate attentional capture by task-irrelevant stimuli that match the search template for familiar object categories in naturalistic scenes. The results of Experiment 1 showed that this attentional capture was specific to conditions in which subjects actively maintained a category-diagnostic search template, suggesting that it did not reflect general effects of encoding the category cues. Experiment 2 revealed that attentional orienting to template-matching stimuli is not fully under top-down control.

In Experiment 1 subjects searched for exemplars of cued categories (natural scene search condition) or objects semantically related to the cued categories (related object search condition). Consistency effects on dot-probe trials were only observed in the natural scene search condition, suggesting that contingent attentional capture is induced by stimuli in the environment that match a target-diagnostic visual search template. Similar effects were not observed during search for semantically related objects. Within the semantic priming literature, conceptual similarities between words have been reported to affect performance in lexical decision tasks. These studies show for example that response times to determine whether a stimulus is a word or a non-word varies as function of whether a prime (i.e., a word cue) is related to the probe. This semantic priming effect has been reported for probe-words that are conceptually as well as perceptually related to the prime (Schreuder, d'Arcais, & Glazenborg 1984). The present findings suggest that the potential effects of conceptual similarity in semantic priming do not extend in the same way to the selection of visual information, since no contingent capture was found for conceptually related search templates; visual selection in our experiment was influenced only by stimuli that were visually similar to the active search template.

Experiment 2 tested whether the consistency effects found on dot-probe trials occurred automatically or instead reflected voluntary orienting. In this experiment, the proportion of consistent trials was reduced and subjects were told that attending to silhouettes of cued categories would be detrimental to performance (2a) or that attending to silhouettes of uncued categories would improve performance (2b). If observers are able to use this information in a strategic manner, they should be able to prioritize processing of the uncued element, which would result in faster response times to the uncued compared to the cued item. Nevertheless, the results continued to reveal consistency effects in both parts of Experiment 2. This suggests that subjects oriented to the silhouettes on dot-probe trials even when it was disadvantageous to performance, providing evidence that these consistency effects reflected involuntary attentional capture. These findings are reminiscent of previous studies that have used simple feature arrays, demonstrating that observers are limited in their ability to strategically modify effects of attentional capture, for example, due to abrupt onsets (e.g., Yantis & Jonides, 1990) or bottom-up feature priming (Theeuwes, Reimann, & Mortier, 2006).

An important aspect of the current study was that search targets and dot-probe targets were always presented in different locations with no overlap between the scene and probe task stimuli. The finding that contingent attentional capture occurs for the probe task stimuli – presented in locations that were always irrelevant for the search task – therefore indicates that the search template affects visual processing globally across the visual field. This is in line with the findings of a previous fMRI study, showing category-based attentional modulation of scenes presented in task-irrelevant locations (Peelen, Fei-Fei, & Kastner, 2009).

Interestingly, the results of Experiment 2a indicate that subjects were unable to inhibit orienting to template-matching stimuli in locations that were irrelevant for the search task, suggesting that the spatially global nature of the category-level template is, to some extent, obligatory. Results of Experiment 2b further show that subjects were unable to activate different templates for different spatial locations (e.g., attending to cars in the scene task locations and to people in the dot-probe task locations), in line with previous findings that only one template may be active at a time (e.g., Houtkamp & Roelfsema, 2009). Together, these results suggest that, at least for category-level search, a single search template represented globally across the visual field guides attention.

3.9.1 Conclusion

The search template for category-level search in naturalistic scenes is likely composed of a spatially global collection of diagnostic object parts of a single category actively attended in working memory. Attentional capture by stimuli matching the search template occurs in an involuntary manner, as evidenced by the finding that observers are unable to strategically modulate the contingency effect when it would be beneficial to do so.

Chapter four: TMS reveals a causal role of object-selective cortex in preparation for real-world category detection

Reshanne R. Reeder[a] & Marius V. Peelen[a]

[a]Center for Mind/Brain Sciences (CIMeC), University of Trento, 38068 Rovereto (TN), Italy

4.1 Abstract

There is much debate about the neural mechanisms that bias attention to predetermined targets in a search display. We propose this mechanism is part of a feedforward process driven by feature-based preparatory attention, rather than a feedback process limited by the serial allocation of attention. In two experiments, we used TMS to test whether object-selective cortex (OSC) is causally involved in the preparatory phase of real-world search. Subjects were cued to detect people or cars in diverse natural scenes. In Experiment 1, following the cue and before scene onset, single-pulse TMS was delivered over OSC or a control region (vertex). After experimentation, subjects indicated the strategy they used to prepare for target detection (high-level, low-level). The strength of the performance impairment following TMS to OSC relative to vertex was predicted by the degree to which subjects attended to high-level target features. This finding indicates that TMS to OSC specifically impaired preparation for high-level features of objects. In Experiment 2, subjects searched at the category level (e.g., any car) or for a specific target (e.g., black minivan) in separate tasks, and received double-pulse pre-stimulus TMS to OSC or Early Visual Cortex (EVC). Category-level detection was impaired following stimulation to OSC but not EVC, whereas this distinction was not present during search for a specific target. This suggests OSC is particularly involved in preparatory activity when the activation of high-level visual features is required to perform search. We conclude that preparatory activity in high-level visual cortex causally contributes to real-world visual perception.

4.2 Introduction

In everyday life, there are millions of stimuli to which one can attend. We have a limited capacity to process items in the visual environment (Duncan, 1980; Kastner, De Weerd, Desimone & Ungerleider, 1999; Kastner, De Weerd, Pinsk, Elizondo, Desimone & Ungerleider, 2001), so we must select and filter incoming information to make sense of the visual world. One of the dominant models of how we select and filter visual stimuli is the biased competition theory, which proposes that all stimuli in the visual field initially compete for selection (Duncan & Humphreys 1989; Wolfe, Cave, & Franzel, 1989), but internally generated information about a target can bias attention to relevant features in a top-down manner (Desimone & Duncan, 1995). The neural mechanism of this top-down bias is highly debated. One theory proposes that visual processing is limited by spatial attention, in that detailed information about target features can only be collected after attention is allocated to a region of the visual field that contains task-relevant features (Olshausen, Anderson, & Van Essen, 1993; Riesenhuber & Poggio, 2000; Serre, Oliva, & Poggio, 2007; VanRullen & Thorpe, 2002). Alternatively, attention may be biased to predetermined target features *prior* to visual processing (Deco & Rolls, 2004; Kastner et al., 1999; Peelen & Kastner, 2011; Reeder & Peelen, 2013; Zhang & Luck, 2009).

The first account dominates the literature currently (Beck & Kastner, 2009; Luck et al, 1997; Motter, 1993), but there is some evidence that has brought the second account to the fore. Several studies have found feature-selective (color, shape) neural activity in monkey inferior temporal (IT) cortex during the delay period between a search cue and the onset of a search display (Chelazzi, Duncan, Miller, & Desimone, 1998; Chelazzi, Miller, Duncan, & Desimone, 1993; Fuster & Jervey, 1982). Zhang and Luck (2009) found in humans that preparatory attention to a color can bias attention to that color within 100 ms of display onset. It is easy to imagine that we use visual cortex to prepare for an impending color, but it is unclear whether visual cortex is involved in feature-based preparatory attention when it is difficult to activate specific visual features in advance, such as during search for object categories in real-world environments. It is the goal of the current study to determine whether preparatory activity in visual cortex is causally involved in category-level target detection in naturalistic scenes.

A previous fMRI study provides a starting point for this investigation (Peelen & Kastner, 2011). Subjects in that study were cued to detect people or cars in photographs of real-world scenes. Subjects who performed well on the task showed category-specific preparatory activity in right posterior temporal cortex (pTC), overlapping object-selective visual cortex, whereas subjects who performed more poorly showed preparatory activity in early visual cortex (EVC). We propose that efficient category-level search relies on category-specific, feature-based preparatory activity in pTC. In the current study, we used TMS in two experiments to investigate the causal role of pTC in preparing to detect real-world object categories.

4.3 Experiment 1

4.3.1 Introduction

We conducted Experiment 1 to investigate whether preparatory activity in pTC causally contributes to object detection in naturalistic environments. Although object selective regions like pTC are typically recruited once an object appears in a display, there is some evidence that the extent to which these regions are recruited during the preparatory phase of search is highly variable between subjects (Peelen & Kastner, 2011). Previous studies have found that preparatory activity in object selective cortex correlates positively with real-world category detection performance (Peelen & Kastner, 2011; Soon, Namburi, & Chee, 2013); conversely, preparatory activity in EVC correlates negatively with category-level search performance (Peelen & Kastner, 2011). This difference at the neural level was linked to different search strategies, with "good" searchers adopting a more general strategy (e.g., searching for categories in various locations and orientations) and "poor" searchers adopting a more specific strategy (e.g., searching for a particular category exemplar). By adding search strategy as a factor in our analyses, we were able to investigate the correlation between the strength of the effect of TMS to pTC and subjects' search strategies for category detection (as reported in Peelen & Kastner, 2011). This provided a well-controlled measure of the degree to which efficient category detection relies on preparatory activity in pTC.

In this experiment, subjects were cued to detect cars and people in diverse photographs containing cars, people, both cars and people, or other objects in a natural scene context. We used single-pulse TMS to stimulate pTC and vertex at two experimental time points before (-200 ms, -100 ms) scene onset, and two control time points after scene onset (+100 ms, +200 ms). We hypothesized that an efficient strategy for category-level search would involve pre-stimulus activity in pTC. This would manifest as a greater decrement in performance under stimulation to pTC compared to vertex during the pre-stimulus time window in subjects who report to use general search strategies compared to subjects who report to use more specific search strategies. This pattern of results would suggest that pTC is critically involved in preparing for high-level features of objects.

4.4 Methods

4.4.1 Subjects

We recruited 16 healthy undergraduate and graduate students from the University of Trento (8 female, all right-handed) aged 21-31 years (mean=25.5 years). All subjects completed a screening questionnaire to ensure that they met the published safety criteria to undergo TMS experimentation (Rossi, Hallett, Rossini, Pascual-Leone, & The Safety of TMS Consensus Group, 2009). Subjects received stimulation to the two target areas prior to the task to expose them to the sensation of TMS and to ensure that stimulation was comfortable for all stimulation sites. All subjects were comfortable with the TMS procedure and received monetary compensation for their participation. All subjects provided written, informed consent prior to taking part in the experiment. The study was approved by the human research ethics committee of the University of Trento and adhered to the tenets of the Declaration of Helsinki.

4.4.2 TMS methods

A Magstim Rapid stimulator with a 75mm MCF-B65 Butterfly coil was used for the TMS. The hand area of the motor cortex was first localized in the left hemisphere. We then determined visual resting motor threshold (MT) for each subject based on the minimum stimulation intensity needed to produce a visible right hand twitch in at least 5/10 pulses. Stimulation intensity for experimentation was set at 120% of each subject's MT.

For each subject, pTC was localized using anatomical brain scans acquired for previous (unrelated) fMRI experiments. Anatomical data were normalized to Talairach space to localize pTC based on the Talairach coordinates reported in Peelen & Kastner (2011: peak: *xyz* = 46, -58, 8; see Figure 1c for an illustration of the pTC coordinates mapped on a sample brain). Once pTC was located for each individual, brain images were transformed back to native space and used to position the coil during the experiment. pTC was located for each subject using Zebris neuronavigation software, and its position was marked on the scalp with a permanent marker. The coil was placed over pTC with the handle pointed toward the back of the head. The coil was turned slightly clockwise or counterclockwise to produce inferior-to-superior current flow to find the best stimulation angle that minimized muscle twitching and eye blinking.

The vertex was located by placing the coil centered between the two cerebral hemispheres on the top of the scalp halfway between the inion and the nasion. During the experiment, subjects sat in a straight-backed chair and rested their heads on a chinrest to minimalize body and head movement during experimentation. Once the stimulation site was located, the TMS coil was stabilized against the subject's head using a metal arm.

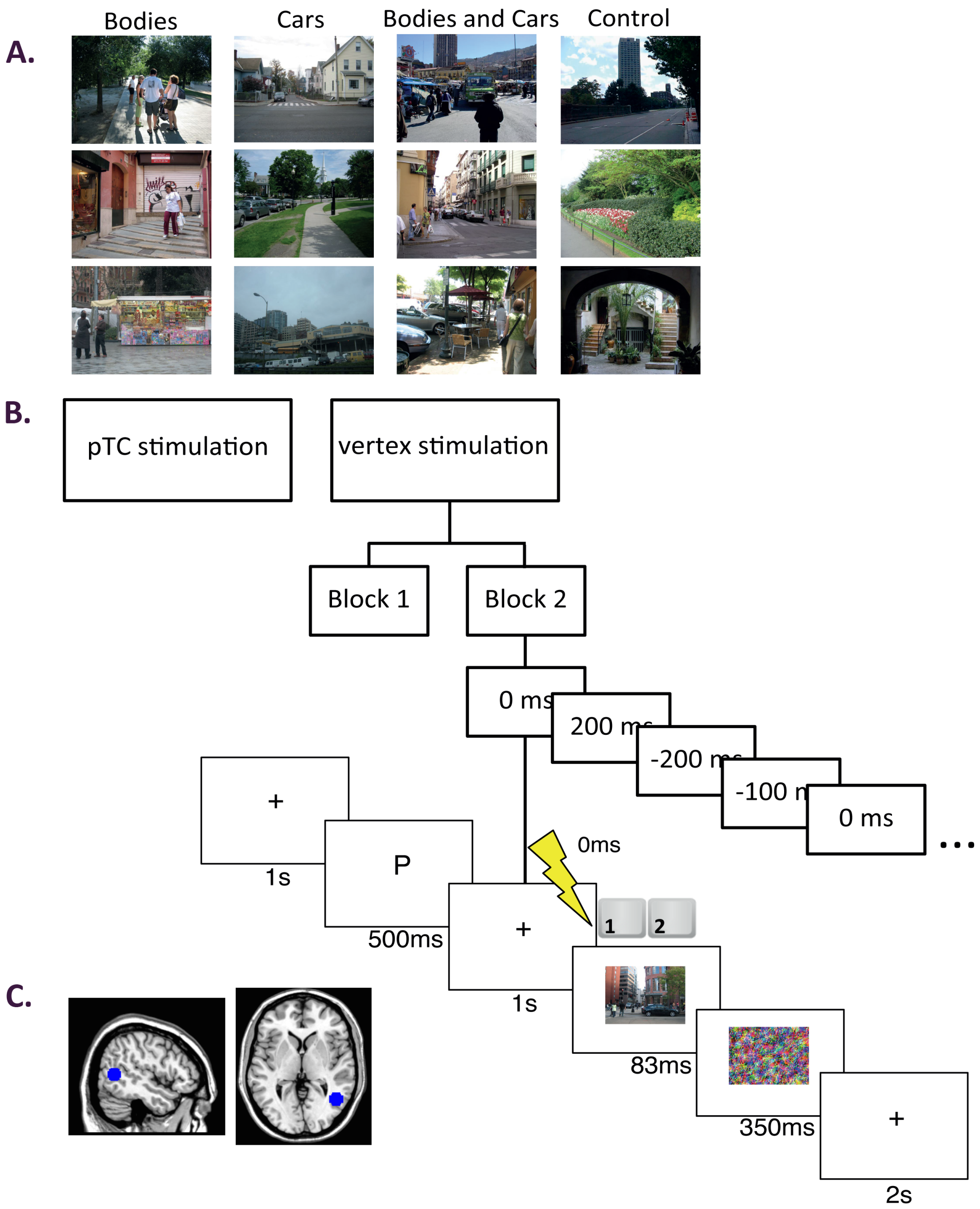


Figure 1. A) Example stimuli from the visual search task of Experiment 1. B) The experimental paradigm for Experiment 1. C) The Talairach coordinates for posterior temporal cortex.

4.4.3 Stimuli

Stimuli were presented in a dimly lit room on a 17-inch gamma-corrected DELL 1908FP-BLK monitor with a screen resolution of 1280x960 pixels and a 60Hz refresh rate. Stimuli were presented using A Simple Framework (Schwarzbach, 2011), based on the Psychophysics Toolbox for MATLAB.

640 color photographs of real-world scenes were obtained from the LabelMe online database (Russell, Torralba, Murphy, & Freeman, 2008). Photographs could contain people (160), cars (160), both people and cars (160), or other objects (160). See Figure 1a for a sample of the images used in the current experiment.

4.4.4 Procedure

Figure 1b provides a schematic of the experimental paradigm. All subjects completed a practice block followed by 8 experimental blocks of 80 trials each. A trial began with the presentation of a centrally presented fixation cross (0.81° visual angle) for 500 ms, followed by a centrally presented single letter cue (1.83° visual angle) for 500 ms, indicating the target category for that trial: “P” for “persona” or “M” for “macchina” (the Italian words for person and car, respectively). After the cue, another fixation cross appeared for 1 s, followed by a centrally presented photograph of a real-world scene (16.08 x 12.18° visual angle) for 83 ms. All four scene types were presented an equal number of times within a block (20 each), with presentations of the “P” and “M” cues distributed evenly across the scene types. This ensured that subjects would see the cued category on 50% of trials. Scenes were backward masked for 350 ms, after which a final fixation cross appeared for 2 s, for a total trial duration of approximately 4.4 seconds. Subjects were instructed to respond whether the cued object (person or car) appeared in the scene by pressing 1 for "yes" or 2 for "no" on the keyboard number pad as accurately as possible. Subjects never saw the same scene twice within an experimental session.

The order in which pTC and vertex were stimulated was counterbalanced across subjects. Subjects received stimulation to pTC or vertex alternating every two blocks (e.g., 2 blocks pTC stimulation, 2 blocks vertex stimulation, 2 blocks pTC, etc.). Single-pulse TMS was administered at four time points within a block (200 ms or 100 ms prior to scene onset, or 100 ms or 200 ms after scene onset), with one pulse per trial. Stimulation onset times were randomized and each subject received an equal number of pulses at the different onset times within a block.

Following each session of the experiment, subjects filled out a questionnaire that assessed their search strategy as either general (e.g., based on high-level object features such as shapes of diagnostic parts; Reeder & Peelen, 2013) or specific (e.g., based on low-level features such as the color of a particular exemplar). Subjects read a list of 10 statements pertaining to general or specific search strategies and rated their agreement with each statement on a 5-point Likert scale, with "1" indicating that the subject fully disagreed and "5" indicating that the subject fully agreed with the statement. Statements that assessed the generality of the subject's search strategy included *After the "P" cue I looked out for persons, but I didn't have a vivid mental image of a person*. Statements that assessed the specificity of the subject's search strategy included *After the "M" cue I vividly imagined a car, as if I could almost see it in front of me*. Statements that assessed the orientation-based strategy included *After the "M" cue I looked out for horizontal objects that were about the size of a car*. A full list of the statements can be found in the appendix.

4.4.5 Analyses

Search accuracy and reaction time (RT) on correct trials were recorded for all subjects. Mean accuracy and median RT values were normalized for each subject using the same procedure: the mean accuracy (or RT) score of the four experimental conditions (pTC pre-stimulus (average of -200 ms and 100 ms time points), pTC post-stimulus (average of +100 ms and +200 ms time points), vertex pre-stimulus (average of -200 ms and -100 ms time points), and vertex post-stimulus (average of +100 ms and +200 ms time points)) was subtracted from the mean accuracy (or median RT) of each condition separately. Each of these values was then divided by the standard deviation (SD) of the mean accuracy (or RT) of the four experimental conditions. This normalized each subject's scores so that they had a mean accuracy (or median RT) of 0 with a SD of 1 for all conditions. Once accuracy and RT scores were normalized, they were combined to create a mean performance score for each of the four experimental conditions (with normalized RT values multiplied by -1 to be consistent with normalized accuracy values).

Search strategy was assessed taking the average reported strength of agreement with the 4 general statements minus the average reported strength of agreement with the 6 specific statements on the search strategy questionnaire, consistent with the individual differences measure used in Peelen & Kastner (2011). A more positive score indicated a more general search strategy.

The results for individual differences in search strategy were reported using correlation analyses on baseline performance (vertex) and under stimulation to pTC compared to vertex (pTC-vertex). To analyze baseline performance for different strategies, accuracy on vertex stimulation trials was averaged across both time windows and then correlated with self-reported search strategy. To reduce the influence of outliers, we used Spearman's rho for our correlations.

4.5 Results

In this experiment, subjects were required to detect people and cars in briefly presented photographs of real-world scenes while undergoing TMS to pTC or vertex prior to scene onset (-200 ms, -100 ms) or after scene onset (+100 ms, +200 ms). First, there was a correlation between generality of search strategy (general-specific) and baseline search performance (mean performance score across vertex stimulation blocks), *r*(14)=0.658, *p*=0.006. This indicates that subjects with a more general search strategy performed category-level search in natural scenes better than those with a more specific search strategy, replicating the results of Peelen & Kastner (2011). Critically, we examined the correlation between generality of search strategy and the performance difference between pre-stimulus TMS to pTC and vertex. This would elucidate the extent to which pTC is involved in activating a high-level search strategy for category detection. This analysis revealed a negative correlation between generality of search strategy and the performance difference between pTC and vertex stimulation in the pre-stimulus time window, *r*(14)=-0.670, *p*=0.004, indeed reflecting a greater performance impairment under pTC stimulation compared to vertex for those subjects with a more general search strategy (Figure 2a). Looking at the correlations for pTC and vertex separately, we found a significant negative correlation between generality of search strategy and pre-stimulus TMS to pTC, *r*(14)=-0.504, *p*=0.046, and a significant positive correlation between generality of search strategy and pre-stimulus TMS to vertex, *r*(14)=0.692, *p*=0.003, indicating that more general strategists show both a greater performance impairment under pre-stimulus TMS to pTC and better baseline performance under vertex stimulation (Figure 2b). These separate results also illustrate that the pre-stimulus pTC-vertex effect is not simply driven by better baseline performance by more general strategists. To further account for this, we conducted a partial correlation between generality of search strategy and the performance difference between pTC and vertex stimulation in the pre-stimulus time window controlling for average pre-stimulus performance variability (average of the pTC and vertex pre-stimulus conditions), using the partial correlation measure in SPSS. This revealed a significant relationship, *r*(13)=-0.688, *p*=0.003, in support of our interpretation of the TMS effects. Finally, there was no significant correlation between generality of search strategy and pTC-vertex search performance in the post-stimulus time window, *r*(14)=-0.369, *p*>0.158, suggesting that there was no difference in the extent to which pTC was recruited by general and specific search strategists after scene onset. These results collectively suggest that stimulation to pTC prior to scene onset, compared to vertex, is more detrimental to performance in searchers who use a more general search strategy compared to those who use a more specific strategy for category-level search in natural scenes.

This experiment delivers encouraging results on the role of pTC in preparing for category-level search, but further investigations are needed to determine whether this region is specifically involved in category-level search preparation or feature-based search preparation in general. Furthermore, it is important to distinguish whether more classical visual regions (EVC) also contribute to category-level search preparation, or whether object-selective cortex plays a more dominant role as previously suggested by Peelen and Kastner (2011). We conducted Experiment 2 to investigate these questions.

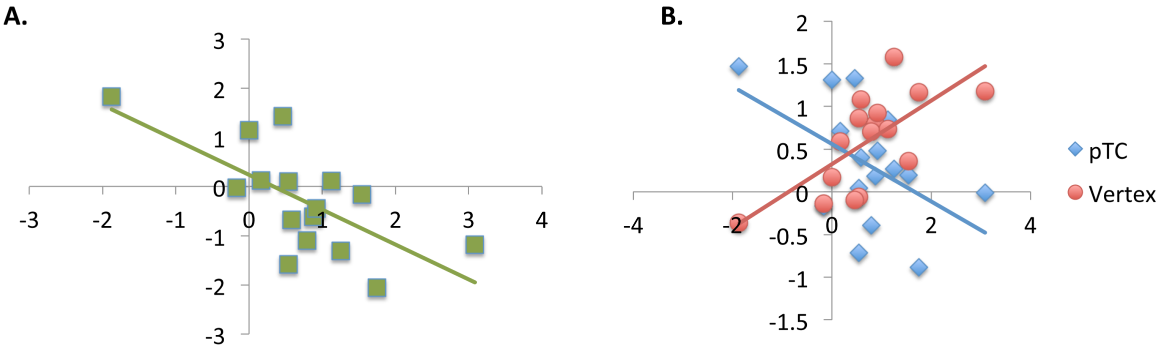


Figure 2. A) The correlation between generality of search strategy and the difference in search performance between TMS to pTC and vertex in the pre-stimulus time window. B) The separate correlations between generality of search strategy and search performance following pre-stimulus TMS to pTC (blue) and vertex (red).

4.6 Experiment 2

4.6.1 Introduction

Previous behavioral studies have suggested that different category levels of search require preparing for different target features (e.g., Bravo & Farid, 2012). For example, preparing for a specific target (e.g., my car) requires the activation of precise attributes of that target (e.g., a car's particular color and model; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004; Vickery, King, & Jiang, 2005) whereas preparing for a target category (e.g., people) requires the activation of attributes that can generalize across various exemplars (e.g., general shapes of body parts; Reeder & Peelen, 2013). Therefore, searchers may prepare for different visual features depending on the nature of the task. In the current experiment, as in Experiment 1, we investigated the causal role of pTC in preparation for category-level search. However, instead of evaluating the extent to which searchers recruit pTC based on self-reported search strategy (general or specific), here we required searchers to change their search strategies in two different (general and specific) search tasks.

Subjects were cued to detect the presence of cars and people in real-world scenes. In separate tasks, subjects either searched for different cars and people on every trial (category-level search task) or a specific car and person repeated across trials (individual-level search task; see Figure 3a). Following the cue, 10 Hz double-pulse TMS was delivered over pTC or EVC (in different blocks) 200 ms prior to scene onset. Based on previous fMRI results (Peelen & Kastner, 2011), we predicted an interaction between task and region, with TMS over pTC disproportionally impairing category-level search compared to TMS over EVC, with no such distinction during individual-level search.

4.7 Methods

4.7.1 Subjects

We recruited 40 healthy undergraduate and graduate students from the University of Trento (25 female, all right-handed) aged 19-34 years (mean=24.6 years). All subjects completed a screening questionnaire to ensure that they met the published safety criteria to undergo TMS experimentation (Rossi, Hallett, Rossini, Pascual-Leone, & The Safety of TMS Consensus Group, 2009). Subjects received stimulation to both target areas prior to the task to expose them to the sensation of TMS and to ensure that stimulation was comfortable for both stimulation sites. All subjects were comfortable with the TMS procedure and received monetary compensation for their participation. All subjects provided written, informed consent prior to taking part in the experiment. The experiment was approved by the human research ethics committee of the University of Trento.

4.7.2 TMS methods

A Magstim Rapid stimulator with a 50mm D70 Alpha coil was used for the TMS. The hand area of the motor cortex was first localized in the left hemisphere. We then determined resting motor threshold (MT) for each participant based on the minimum stimulation intensity needed to produce a visible right hand twitch in at least 5/10 pulses. Stimulation intensity for experimentation was set at 120% of each participant's MT.

For each participant, pTC was localized using anatomical brain scans acquired for previous (unrelated) fMRI experiments. Anatomical data were normalized to Talairach space to localize pTC based on the Talairach coordinates reported in Peelen & Kastner (2011; peak: *xyz* = 46, -58, 8); see Figure 1c for an illustration of the pTC coordinates mapped on a sample brain. Once pTC was located for each individual, brain images were transformed back to native space and used to position the coil during the experiment. pTC was located for each participant using Zebris neuronavigation software, and its position was marked on the scalp with a permanent marker. The coil was placed over pTC with the handle pointed toward the back of the head. The coil was turned slightly clockwise or counterclockwise to produce inferior-to-superior current flow to find the best stimulation angle that minimized muscle twitching and eye blinking.

The early visual cortex (EVC) region was located by placing the coil 2 cm above the inion pointed inferiorly and adjusting it slightly to the left and right until stimulation evoked the perception of phosphenes (flashes of light) in blindfolded subjects. Stimulation intensity was varied until subjects reported seeing phosphenes, and phosphene threshold (PT) was determined as the lowest stimulation intensity needed to produce phosphenes in approximately 3/5 pulses. For subjects who did not report seeing phosphenes, EVC was located by placing the coil 2cm above the inion pointed inferiorly. PT was not used to determine stimulation intensity during the experiment.

During the experiment, subjects sat in a straight-backed chair and rested their heads on a chinrest to minimalize body and head movement during experimentation. Once the stimulation site was located, the TMS coil was stabilized against the participant's head using a metal arm.

4.7.3 Stimuli

All stimuli were presented on a 22-inch Dell E228 WFP monitor with a screen resolution of 1680x1050 pixels and 60Hz refresh frequency. Stimuli were presented using A Simple Framework (Schwarzbach, 2011), based on the Psychophysics Toolbox for MATLAB.

Photographs of real-world scenes were obtained from the LabelMe online database (Russell, Torralba, Murphy, & Freeman, 2008) and converted to grayscale. Object stimuli for the category-level search task consisted of 48 different images of people and 48 different images of cars that were manually inserted into different scenes (Figure 3a); no people or cars were inserted in the remaining 96 scenes; 48 of these were used in the category-level search task and 48 were used in the individual-level search task. Two object stimulus sets were used for the individual-level search task, and each set of person and car images was inserted into the same 96 person or car scenes as in the category-level task. Importantly, in the individual-level task, the inserted person and car images were of the same person and car, repeated in 48 scenes each (Figure 3a). In both tasks, person and car images were placed in natural locations within a scene (i.e., a person could appear on a staircase and a car could appear in a driveway, but neither could appear floating in the sky). Targets were placed in various parts of the scene (far or near, to the left or right), and thus could appear large or small depending on the appropriateness of the scale. Category-level and individual-level search targets were matched by location and size (i.e., the person in scene 1 of the category-level task would be in the same location and of the same size as the person in scene 1 of the individual-level task; Figure 3a).

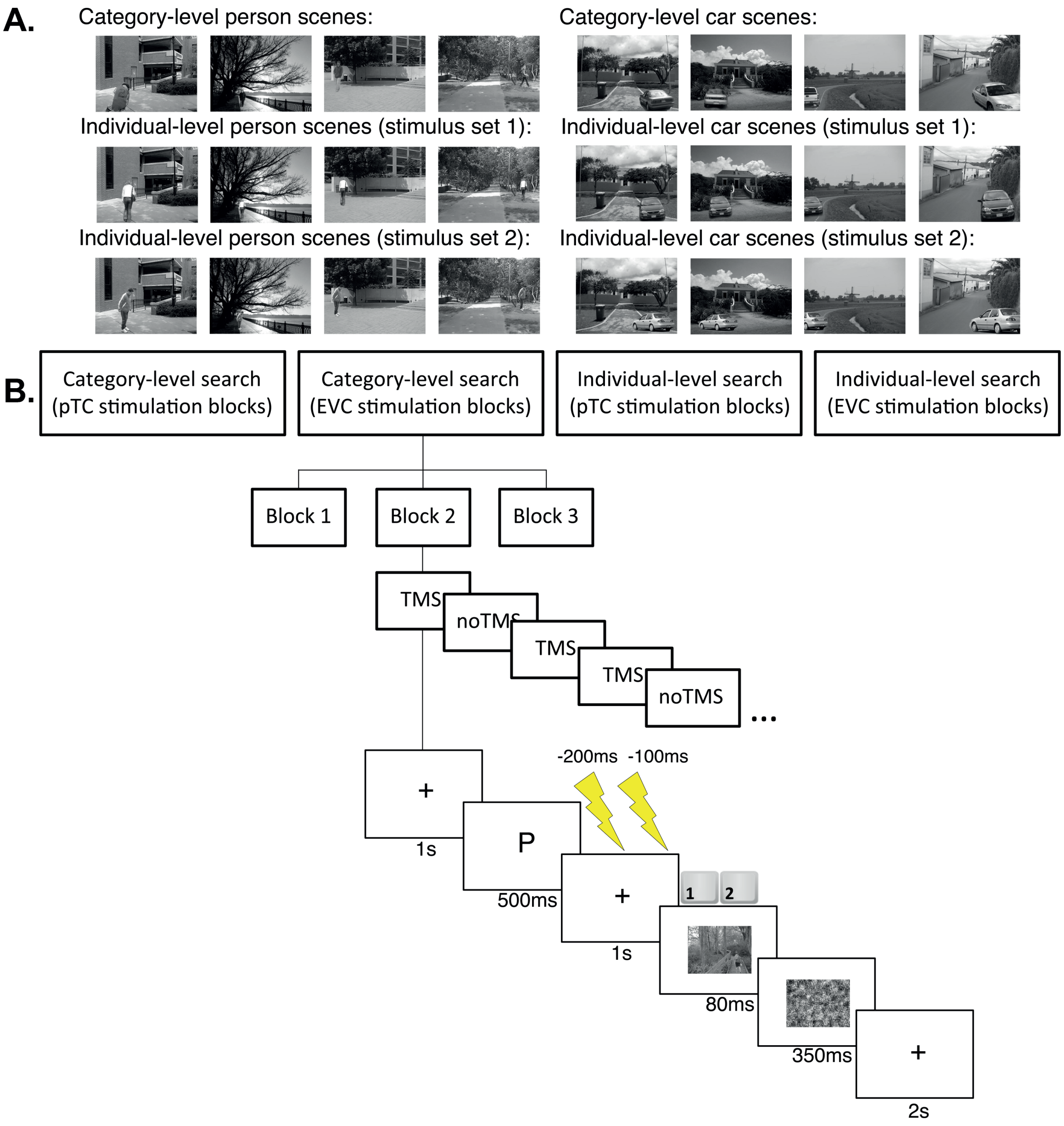

A.
Category-level person scenes:
Category-level car scenes:
Individual-level person scenes (stimulus set 1):
Individual-level car scenes (stimulus set 1):
Individual-level person scenes (stimulus set 2):
Individual-level car scenes (stimulus set 2):
B.
Category-level search (pTC stimulation blocks)
Category-level search (EVC stimulation blocks)
Individual-level search (pTC stimulation blocks)
Individual-level search (EVC stimulation blocks)
Block 1
Block 2
Block 3
TMS
noTMS
TMS
TMS
noTMS
...
+
1s
P
500ms
-200ms
-100ms
+
1s
1
2
80ms
350ms
+
2s

Figure 3. A) Examples of the stimuli used in the category-level and individual level search tasks of Experiment 2. B) The experimental paradigm of Experiment 2.

4.7.4 Procedure

See Figure 3b for a schematic of the experimental paradigm. All subjects completed one experimental session in which they were given a practice block followed by 12 experimental blocks of 48 trials each. A trial began with the presentation of a centrally presented fixation cross (0.81° visual angle) for 500 ms, followed by a centrally presented single letter cue (1.83° visual angle) for 500 ms, indicating the target category for that trial: "P" for "persona" or "M" for "macchina" (the Italian words for person and car, respectively). After the cue, another fixation cross appeared for 1 s, followed by a centrally presented photograph of a real-world scene (16.08 x 12.18° visual angle) for 83 ms in the category-level task or 50 ms in the individual-level task; these presentation times were determined after extensive behavioral piloting (with different subjects from those tested in the TMS experiment) with the aim to match the difficulty of the two tasks. There was an equal number (16) of scenes containing a person, a car, or neither person nor car within a block, and the two letter cues ("P", "M") appeared an equal number of times in each of the different scene types. Scenes were backward masked for 350 ms, after which a final fixation cross appeared for 2 s, for a total trial duration of approximately 4.4 seconds. Subjects were instructed to respond whether the cued object (person or car) appeared in the scene by pressing 1 for "yes" or 2 for "no" on the keyboard number pad as fast and accurately as possible. In the category-level task, subjects never saw the same person or car twice within a block, while in the individual-level task the same person and car were repeated throughout a block.

Subjects performed 6 consecutive blocks of the category-level task and 6 consecutive blocks of the individual-level task within a session, with 3 consecutive blocks of pTC stimulation and 3 consecutive blocks of EVC stimulation for each task. The stimulus images used in the general task (48 car scenes, 48 person scenes, and 48 with neither cars nor people) and the images used in the specific task were repeated for each of the two stimulation sites. The order of the stimulated regions (EVC-pTC or pTC-EVC) was the same in both tasks for a single participant, and this order was counterbalanced across subjects. Task order was also counterbalanced across subjects for a total of 10 subjects in each region/task order condition (i.e., if 1=pTC stimulation in the category-level task, 2=EVC stimulation in the category-level task, 3=pTC stimulation in the individual-level task, and 4=EVC stimulation in the individual-level task, then there were 10 subjects each in the following task orders: 1-2-3-4, 2-1-4-3, 3-4-1-2, and 4-3-2-1.

10 Hz double-pulse TMS was administered 200 prior to scene onset on half of trials. No TMS was administered on the other half of trials as a baseline condition for stimulation. TMS and no-TMS trials were interspersed randomly without replacement throughout a block. Each of the two cue conditions (person, car) and each of the three scene types (person, car, neither) received an equal number of TMS and no-TMS trials within a block.

4.7.5 Analyses

Search accuracy and reaction time (RT) on correct trials were recorded for all subjects. Mean accuracy and median RT values were normalized for each subject to remove response variation that was not related to the experimental conditions using a similar procedure to that of Peelen & Kastner (2011): the mean accuracy (or

RT) score of the four experimental conditions (category-level task, pTC stimulation; category-level task, EVC stimulation; individual-level task, pTC stimulation; individual-level task, EVC stimulation) was subtracted from the mean accuracy (or median RT) of each condition separately. This value was then divided by the standard deviation (SD) of the mean accuracy (or RT) of the four experimental conditions. This normalized each subject's scores so that they had a mean accuracy (or median RT) of 0 across the four experimental conditions, with a SD of 1. This procedure removed stimulation artifacts from the data and allowed us to analyze effects that were due to the different experimental conditions. Once accuracy and RT scores were normalized, they were combined into a single mean performance score (with normalized RT values multiplied by -1 to be consistent with normalized accuracy values).

4.8 Results

Normalized RT/accuracy scores for TMS trials were submitted to a 2x2 repeated-measures ANOVA with Task (category-level task, individual-level task) and Region (pTC, EVC) as factors. Despite the large sample size (N=40), there was no main effect of Task, $F(1,39) = 0.266$, $p = 0.609$, ηp2 < 0.007, indicating that the tasks were well matched for difficulty. Importantly, confirming our hypothesis, there was a highly significant Task x Region interaction (see Figure 4), $F(1,39) = 12.771$, $p = 0.001$, ηp2 = 0.247, reflecting worse detection performance in the category-level task during pTC stimulation blocks than during EVC stimulation blocks, $t(39) = -3.043$, $p =$ 0.004, $d = -0.704$. No such effect was observed for the individual-level task, $t(39) =$ -

1.298, p = 0.202, d = 0.284 Neither of the main effects or the Task x Region interaction were significant on noTMS trials, all Fs<2.3, all ps>0.14.

The results of Experiment 2 provide TMS evidence for a dissociation between preparatory neural mechanisms involved in category-level and individual-level search in real-world scenes. These findings directly support the results of a recent fMRI study (Peelen & Kastner, 2011), which showed that the category specificity of preparatory activity patterns in pTC, but not EVC, was positively correlated with RT on a naturalistic category-level search task that was similar to the category-level task used in the current study.

The present finding of impaired category-level search in pTC relative to EVC stimulation blocks highlights a significant difference in the contribution of these regions to category-level search. Activating high-level, category diagnostic features (recruiting pTC) is advantageous for category-level search (consistent with previous findings by Reeder & Peelen, 2013), whereas activating low-level features of category exemplars (recruiting EVC) may be detrimental to this process. These results provide evidence that pTC is causally involved in preparing for category-level search to a greater extent than EVC.

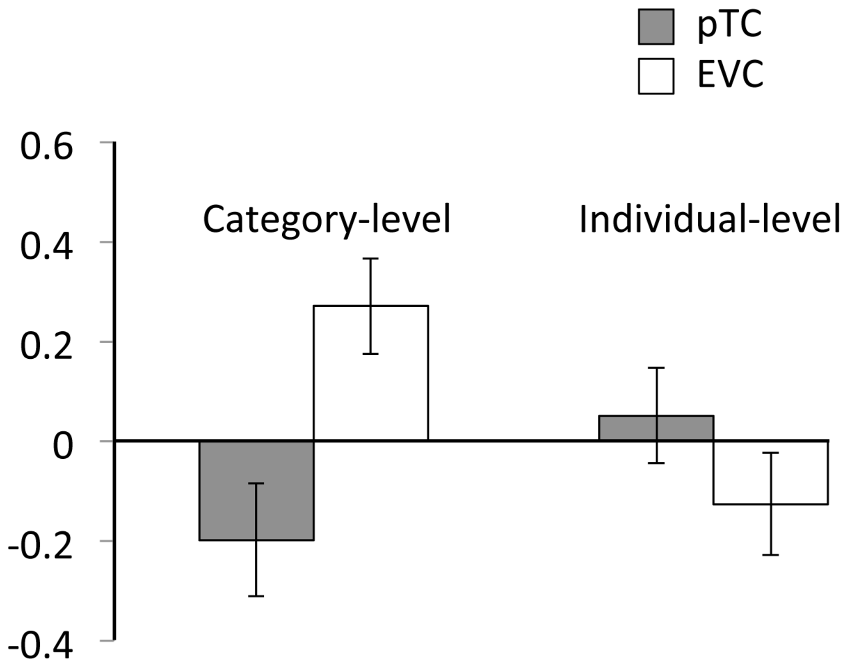


Figure 4. Normalized RT/accuracy scores in the category-level and individual-level search tasks following stimulation to pTC and EVC. Error bars represent the standard error of the mean.

4.9 General Discussion

Attention can be biased to predetermined targets that subsequently appear in a visual scene (Desimone & Duncan, 1995). Whether the neural mechanisms of this bias are part of a location-based feedback system following target onset (Olshausen et al., 1993; Riesenhuber & Poggio, 2000; Serre et al., 2007; VanRullen & Thorpe, 2002) or a feature-based feedforward system activated prior to target onset (Chelazzi et al., 1993; Chelazzi et al., 1998; Deco & Rolls, 2004; Kastner et al., 1999; Peelen & Kastner, 2011; Reeder & Peelen, 2013; Zhang & Luck, 2009) is widely contested. In this study, we report two experiments that provide evidence for a causal involvement of objective-selective visual cortex (right posterior temporal cortex, or pTC) in preparing for high-level features of real-world objects. In Experiment 1, we found that stimulation to pTC prior to scene onset impairs category-level detection performance to a greater extent in subjects who reported to prepare for general, high-level (e.g., non-specific, view-invariant) features of objects compared to subjects who reported to prepare for specific, low-level (e.g., color, orientation) features of objects. In Experiment 2, we found that pre-stimulus TMS to pTC relative to early visual cortex (EVC) impairs category-level search to a greater extent than individual-level search, suggesting that pTC is required to prepare for high-level, but not low-level, features of objects. These results collectively suggest a causal role for pTC in preparing for category detection in naturalistic scenes.

The correlation between search strategy and performance in Experiment 1 suggests that category-level search relies on a high-level cognitive strategy for effective performance. The neural dissociation between category-level and individual-level search tasks of Experiment 2 suggests that different levels of search rely on different cognitive strategies. Activating a low-level search strategy is detrimental to performance in a category-level search task, in which objects of the same category may appear very differently from one exemplar to the next. A more general search strategy that could account for these variations between exemplars is optimal in category-level search situations. Previous behavioral and computational research suggest that the optimal search strategy for category detection relies on activating a preparatory set of category-diagnostic, view-invariant shape features of intermediate complexity that can generalize across most members of the target category (Delorme, Richard, & Fabre-Thorpe, 2010; Evans & Treisman, 2005; Reeder & Peelen, 2013; Rousselet, Mace, & Fabre-Thorpe, 2003; Ullman, Vidal-Naquet, & Sali, 2002). Nevertheless, a specific search strategy is not always disadvantageous, and is even necessary when searching for specific targets, e.g., during search for your car in a parking lot or a friend in a crowd. Our results indicate that different mechanisms underlie search at different levels of specificity, consistent with behavioral studies showing that the target features that are activated in preparation for search depend on task demands and target-distracter similarity (Boot, Becic, & Kramer, 2009; Bravo & Farid, 2009; 2012; Collin & McMullen, 2005; Schmidt & Zelinsky, 2009; Vickery et al., 2005; Yang & Zelinsky, 2009).

One question that emerges from our study is whether the current results reflect a disruption of preparatory attention for visual features of targets or expected target locations. The latter possibility was raised by Beck and Kastner (2009), who suggested that an increase in "baseline activity" (i.e., preparatory activity) has only been replicated in studies that required subjects to attend to particular locations, rather than features. Although this question was not directly addressed in the current study, we provide evidence that our tasks required global search across the display for target features. Because our stimuli were naturalistic and changed on every trial (except in the individual-level search task of Experiment 2), targets could appear in hundreds of possible locations. Furthermore, stimulus presentations were rapid (always less than 100 ms) and backward-masked, so attending to one location in a scene would result in missing targets that appeared outside of the focus of attention. In fact, we have previously found evidence that target features presented in task-irrelevant locations can nevertheless capture attention during search, suggesting that preparatory attention for naturalistic categories may be represented globally across the visual field (Experiment 5, Reeder & Peelen, 2013). It is therefore unlikely that subjects used a location-based search strategy in the current study.

The results of the current study also contribute to the literature on pre-stimulus TMS effects. In the early days of TMS, Becker and Homburg (1991) found that visual task performance could be disrupted following stimulation to early visual cortex (EVC) 40 ms prior to stimulus onset, resulting in TMS-induced forward masking (classically, forward masking occurs when the processing of one target suppresses the processing of a second target presented in rapid succession). Several more recent studies have found that applying TMS to EVC anywhere between 80 ms and 25 ms prior to display onset leads to a significant dip in performance when subjects are required to discriminate briefly presented (16.7-33.4 ms) stimuli, even when controlling for eye blink artifacts (de Graaf, Jacobs, Cornelsen, & Sack, 2011; Jacobs, de Graaf, Goebel, & Sack, 2012; Jacobs, de Graaf, & Sack, 2014; Jacobs, Goebel, & Sack, 2012).

Although the pre-stimulus TMS effects in all of these studies were attributed to visual suppression, performance impairments could alternatively reflect a disruption of top-down preparatory activity (Laycock, Crewther, Fitzgerald, & Crewther, 2007). Stimulating EVC in time windows around 50 ms prior to stimulus onset may hinder the influx of feedback from higher order regions, which are likely activated more than 100 ms earlier than EVC (as suggested by the current study). This would thus interrupt the flow of top-down information along the visual hierarchy rather than suppress subsequent visual processing. To test this hypothesis, Vetter, Grosbras, and Muckli (2013) required subjects to determine the presence of a briefly presented target when it followed the apparent motion of a stimulus (congruent) or did not follow this trajectory (incongruent). Subjects showed an improvement in performance when the target was congruent with the apparent motion stimulus, but this congruency advantage was abolished under double-pulse TMS to extrastriate area V5 at 53 ms and 13 ms prior to target onset. This is evidence for a diminished effect of *expectation* due to pre-stimulus TMS to extrastriate cortex, and cannot be attributed to visual suppression (which would have resulted in diminished performance in both congruent and incongruent conditions).

The current study provides further evidence that pre-stimulus TMS can be used to disrupt preparatory activity. If the effects were simply due to forward masking, we likely would have found performance impairments under pTC and EVC stimulation in both the category-level and individual-level search tasks of Experiment 2. Forward masking also cannot explain the strong correlations between search strategy and the performance impairment due to pTC stimulation in Experiment 1. These results instead point toward a specific performance impairment under pre-stimulus TMS to pTC during category-level search, rather than a general TMS-induced masking effect.

4.9.1 Conclusions

The results of the current study suggest that efficient real-world category detection requires searchers to activate a preparatory set of high-level object features in posterior temporal cortex. Our results cannot be attributed to a feedback mechanism limited by the scope of spatial selective attention, or a forward masking effect of pre-stimulus TMS. The current study is the first to provide evidence for a causal role of posterior temporal cortex in preparing for object detection in naturalistic scenes.

Chapter five: Effects of perceptual expertise on category-level detection in natural scenes

Reshanne R. Reeder[1], Timo Stein[1], and Marius V. Peelen[1]

[1]Center for Mind/Brain Sciences, University of Trento, 38068 Rovereto, Italy

5.1 Abstract

There is much debate about how categorization, individuation, and detection relate to one another during object recognition. Whether these tasks rely on the same representations or different representations may be determined by whether training on one of these tasks (i.e., individuation) can facilitate performance in another (i.e., detection). Individuation training results in perceptual expertise for a particular category (e.g., cat enthusiasts are experts at discriminating different car models), but thus far, there has been no research into the effects of perceptual expertise on category-level detection in naturalistic scenes, for which the ability to discriminate individual objects based on subtle visual differences is seemingly irrelevant. In the current study, a large group of car experts (N=34) were recruited to search for cars and people in hundreds of photographs of naturalistic scenes. Results revealed that car experts were more accurate than novices in detecting cars relative to people, indicating that perceptual expertise boosts category-level detection ability without any detection training. This suggests that different aspects of object recognition (i.e., individuation, detection) likely draw from the same representations.

5.2 Introduction

Categorization, individuation, and detection are all part of object recognition, but there is much debate about how these processes relate to one another (Delorme, Rousselet, Macé, & Fabre-Thorpe, 2004; Grill-Spector & Kanwisher, 2005; Large, Kiss, & McMullen, 2004; Mack, Gauthier, Sadr, & Palmeri, 2008; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; Thorpe, Fize, & Marlot, 1996). Categorization and category detection are similarly efficient, whereas individuation is typically a much slower process (Grill-Spector & Kanwisher, 2005, suggesting that individuation perhaps relies on different cognitive resources. Intuitively, this makes sense: labeling an object as a car rather than a dog requires the ability to attend to low-frequency features (Collin & McMullen, 2005) that are both generalizable to other objects of the same category and optimally discriminable from objects of other categories. For example, object shape features such as the rim/wheel of a car are highly diagnostic and are rapidly classified as belonging to the category "cars" (Harel, Ullman, Harari, & Bentin, 2011; Ullman, 2006; Ullman, Vidal-Naquet, & Sali, 2002). Similarly, a previous behavioral study found that detecting cars and people in real-world scenes requires the ability to prepare for shapes of object parts that may appear in various views, disregarding high-frequency information such as textures and color patterns (e.g., racing stripes on the side a car) or prototypical configurations of parts (e.g., a car as seen from the side; Reeder & Peelen, 2013). Conversely, high-frequency information (Collin & McMullen, 2005) and particular configurations of parts (e.g., eyes above a nose above a mouth in faces) are critical for individuation, which requires a slower, more detailed level of processing relative to categorization and detection (Grill-Spector & Kanwisher, 2005; Op de Beeck & Baker, 2010). Nevertheless, studies of expertise have found that extensive individuation training can benefit performance on these tasks, suggesting that they may actually rely on similar representations. Discrimination, or perceptual, experts can individuate objects of expertise with the same efficiency as categorization (Curby & Gauthier 2009; Gauthier, Williams, Tarr, & Tanaka, 1998; Robbins & McKone, 2007; Tanaka & Taylor, 1991). In fact, objects of expertise are also categorized faster than objects from other categories (e.g., cars vs. faces or airplanes; Harel et al., 2011). The fact that individuation training can have such an impact on categorization suggests that these tasks are actually quite closely related and perhaps rely on the same representations (e.g., activity in the same neural regions; see General Discussion.

Just as with categorization, the expert ability to individuate objects seems irrelevant for category-level detection. However, there is some evidence that experts have a category-level detection advantage for objects of expertise compared to objects of other categories (Hershler & Hochstein, 2009). It is possible that discrimination training enhances detection for a particular object category, which would suggest that individuation is also linked to detection: in this case, it would seem that individuation, categorization, and detection all draw from the same representations for object recognition.

In a previous study by Hershler and Hochstein (2009), bird experts and car experts were recruited to actively search through displays of 9, 16, 25, or 36 photographs of various object categories for a single photograph of a bird, car, or face in separate blocks of trials. Experts were instructed to determine whether a target was present or absent as fast as possible, and a decision terminated the trial. The authors found faster and more accurate detection of objects from the expert category compared to the non-expert category in both bird experts and car experts, which is evidence for a detection advantage for objects of expertise even at the category level. However, there are some points that need to be addressed before more substantial conclusions can be made about the relationship between individuation and detection. In real-world search environments, one must search for category members that may appear in different viewpoints or sizes depending on the perspective of the observer, and this study did not take real-world perspective into account in its design. Target objects in this task were always separated from other objects in the display in different photographs, and there was relatively little contextual information that would be present in a more realistic search environment. Furthermore, there was a small number of experts in this experiment (5 car experts and 6 bird experts), with each group acting as a novice control group for the other. It is possible that low statistical power produced false positive results, and furthermore, selection bias may have played a role in shaping the observed effects, particularly because of the blocked design; that is, car experts may have simply found the car detection task more interesting than the bird detection task (and vice versa for bird experts), leading to significant performance differences between the "interesting" and "uninteresting" target categories for the two groups. To follow up on this study, we investigated whether the category-level detection advantage in perceptual experts holds up in more naturalistic settings using a larger set of experts who could act as their own controls.

The current study was conducted to explore the effects of perceptual expertise on category-level detection in naturalistic scenes. A large group (N=34) of self-proclaimed car experts were cued to detect people or cars in diverse real-world photographs in which targets could appear in various positions, sizes (perceived distance), and levels of occlusion. We then correlated each subject's car and person detection performance with their performance on a car discrimination task. We hypothesized that discrimination is a more detailed part of the detection process rather than a separate process, and that car experts who scored higher on the car discrimination task should show more efficient detection for cars compared to experts who scored lower on the car discrimination task. This should manifest as a correlation between car discrimination performance and car detection performance, with no such correlation between car discrimination and person detection.

5.3 Methods

5.3.1 Subjects

34 self-proclaimed car experts (2 women) participated in the current study for payment. Subjects were between the ages of 19 and 63 (mean age = 25.98) and had normal or corrected-to-normal vision. 4 subjects were left-handed. Subjects were recruited from the Rovereto and Trento communities by responding to fliers that called for car experts to participate in the experiment. 30 subjects had completed at

least some university education, and the other 4 had completed a high school education. All subjects were given monetary compensation for their time.

5.3.2 Stimuli

All stimuli were presented on a 19-inch Dell 1905 FP monitor with a screen resolution of 1280 x 1024 pixels and 60Hz refresh frequency (Dell Inc., Round Rock, TX). Subjects sat 57cm from the screen. Stimuli were presented using A Simple Framework (Schwarzbach, 2011), a toolbox based on the Psychophysics Toolbox for MATLAB (The MathWorks, Inc., Natick, MA).

Expertise assessments

Stimuli in the expertise assessments were 480 centrally presented black and white photographs of cars (160), birds (160), and faces (160) shown in isolation on a white background. Images of modern cars (no more than ~5 years out of production) commonly seen on European streets were retrieved from free-access online image searches. A car expert created pairs of cars that were difficult to distinguish as belonging to the same or different make or model based on perceptual similarities alone. We additionally ensured that names written on the side of the car or on the license plate were not visible in the selected images. 40 pairs of cars were the same make and model, 20 were the same make but different models, and 20 were different makes and models, for a total of 80 pairs of cars. Cars of even the same make and model could appear in different positions and colors, and could be from different years and series. Bird images were the same as those used by Isabel Gauthier and colleagues in previous experiments (e.g., Gauthier, Skudlarski, Gore, & Anderson, 2000), with 40 pairs belonging to the same species and 40 pairs belonging to different species. Face images were obtained from a database created by the Olivetti Research Laboratory in Cambridge, UK (e.g., Samaria & Harter, 1994). Faces were cropped around the forehead, cheeks, and chin to fit within a standard oval shape, omitting the hair and ears. Faces were Caucasian or mixed-ethnicity, reflecting the major ethnicities of the subjects of the current study. All faces were presented without eyewear or jewelry to prevent subjects from identifying faces based on non-facial features. There were 40 "same" pairs in which two different images of the same person were presented, and 40 "different" pairs that were two different people matched by gender and hair color. See Figure 1a for some examples of stimuli in the expertise assessments.

Images of cars, birds, and faces were scaled to fit inside 300x300, 256x256, and 250x250 pixel-sized boxes, subtending 8.8°, 7.5°, and 7.4° in height and width, respectively. The sizes of the bird and face images were not changed from their original sources, whereas the size of the car images was adjusted for optimal visibility based on the resolution of the photographs (as determined by a car expert).

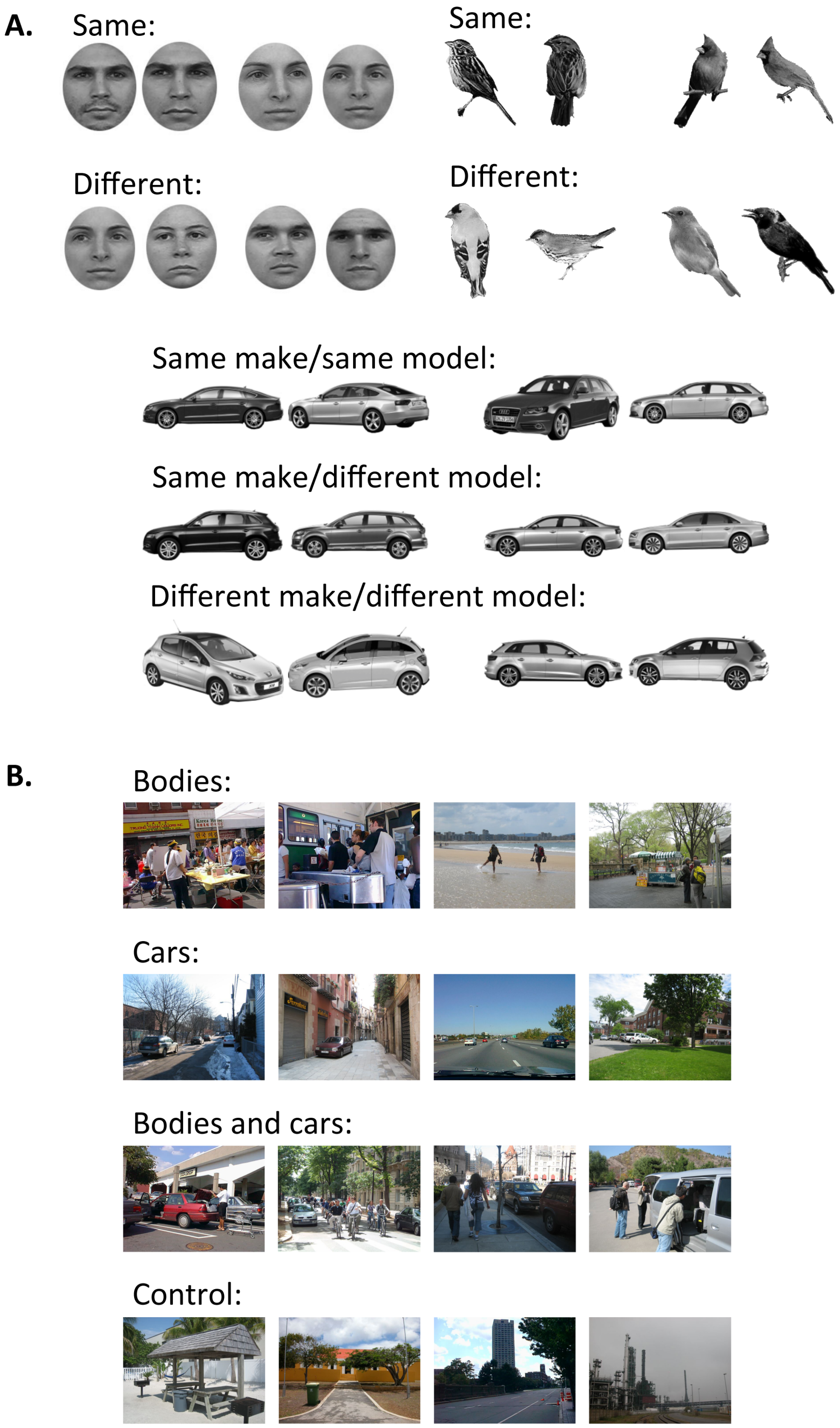

A.
Same:
Same:
Different:
Different:
Same make/same model:
Same make/different model:
Different make/different model:
B.
Bodies:
Cars:
Bodies and cars:
Control:

Figure 1. A) Examples of the stimulus pairs used in the discrimination tasks of the expertise assessments. B) Examples of the photographs used in the visual search task.

Experimental stimuli

A fixation cross and uppercased letter cues appeared centered on the screen in 70-point "strong" Times New Roman font. The fixation cross had dimensions of 31 x 31 pixels subtending 0.92° in height and width, and letters had dimensions of 70 x 70 pixels subtending 2.1° in height and width.

Natural scene stimuli

Stimuli presented in the search task were 960 color photographs (see Figure 1b) of real-world scenes obtained from the LabelMe online database (Russell, Torralba, Murphy, & Freeman, 2008) and were divided into scenes containing cars (240), people (240), both cars and people (240), or neither cars nor people (240). Two scenes appeared on every trial and no scene was repeated within the experiment. Scenes containing people exclusively were always paired with scenes containing cars exclusively, while scenes containing both cars and people were always paired with scenes containing neither cars nor people.

Scenes were scaled to 548 x 411 pixel resolution, subtending a visual angle of 16.08 x 12.18°. Scenes were presented 7.41° from the center of the screen to the center of the image, above and below fixation.

5.3.3 Experimental procedure

Expertise assessments

Prior to the main experiment, all subjects underwent a series of four discrimination tasks: one each for upright cars, upright birds, upright and inverted cars, and upright and inverted faces, completed in separate blocks (see Figure 2a). In the upright car discrimination task, subjects were required to determine whether two cars presented in succession were the same or different model (e.g, Honda Civic). On each trial, a car appeared in the center of the screen for one second, followed by a fixation for 500 ms, then another car that would remain on screen until the subject made a button press. Subjects responded by pressing the "1" key on the number pad if they believed the two cars were the same model and the "2" key if they believed the two cars were different models. The same procedure was used for the upright bird discrimination task, except two pictures of birds appeared instead of cars, and subjects were required to respond whether they believed the two birds were the same ("1" key) or different ("2" key) species. The upright car and bird discrimination tasks were 80 trials each and were performed in succession. The upright and inverted car discrimination task was twice as long as the upright task (160 trials), with each car presented upright and inverted once. Upright and inverted trials were intermixed, but image pairs were always either both upright or both inverted. The upright and inverted face discrimination task followed the same procedure (160 trials), except subjects were required to respond whether two faces were the same person ("1" key) or two different people ("2" key). Each subject performed the discrimination tasks in the same order, starting with upright and inverted cars, followed by upright and inverted faces, then upright cars, and finally upright birds.

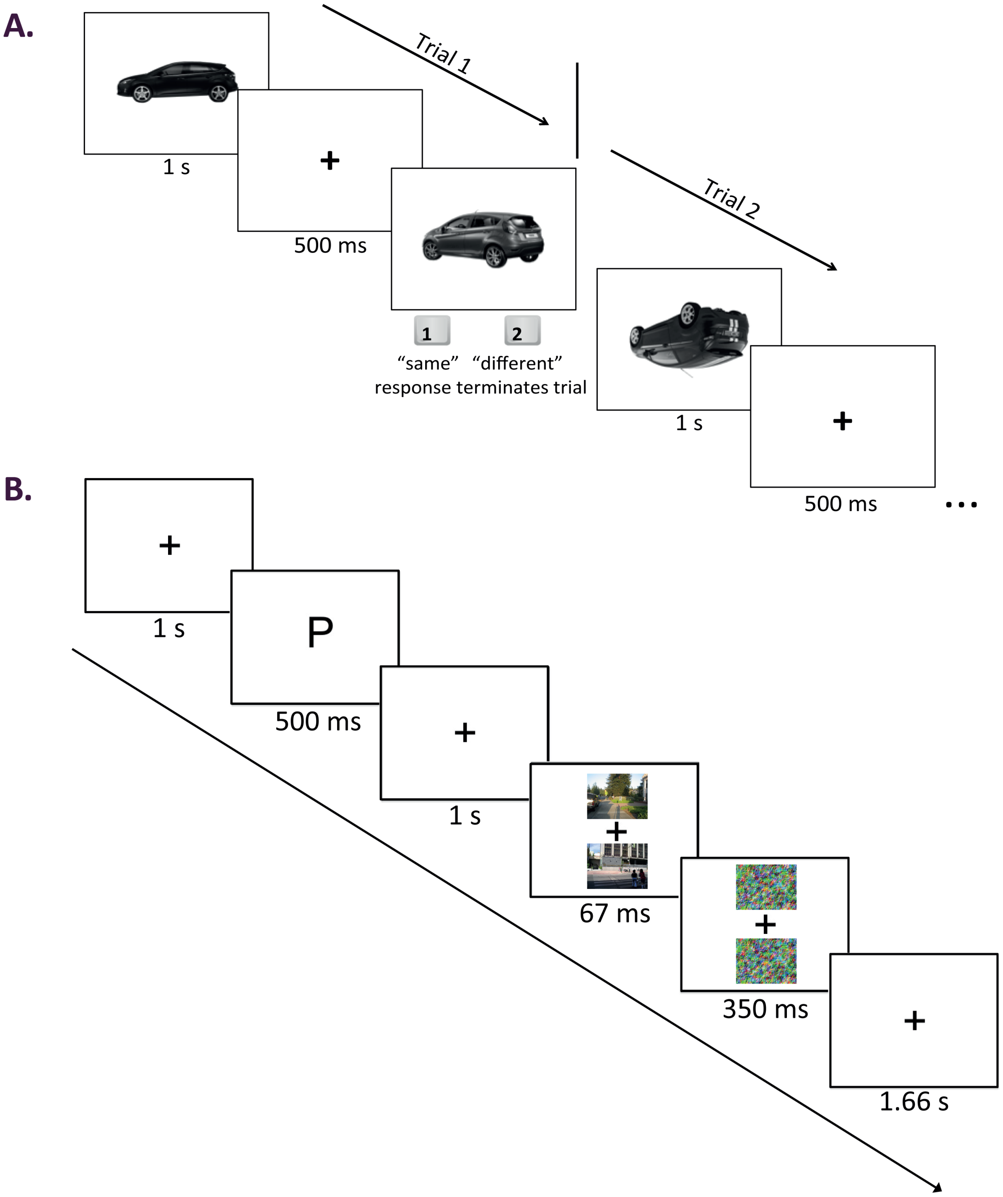


Figure 2. A) The experimental paradigm of the discrimination tasks of the expertise assessments. Subjects completed 1 block each for upright cars, upright birds, upright and inverted faces, and upright and inverted cars (shown here). B) The experimental paradigm of the visual search task.

Visual search

All subjects (N=34) took part in one practice block followed by 8-10 blocks of 64 trials each. The search task started with the presentation of a fixation cross for 500 ms, followed by a single letter for 500 ms: “P” for “persona: or “M” for “macchina” (the Italian words for person and car, respectively). After the letter, another fixation cross appeared for 1 s. Subjects would then see two photographs of natural scenes for 67 ms, followed by a 350-ms mask. Subjects were required to respond whether the cued object category appeared above or below fixation using the “up” and “down” arrow keys, respectively. The two scenes that appeared could either be one containing cars and the other containing people, or one containing both cars and people and the other containing no cars or people. This structure allowed us to present people and cars on every trial. Each of the four scene types appeared in each possible location an equal number of times above and below fixation. See Figure 2b for a schematic of the visual search paradigm.

Within each experimental block, 16 attentional capture trials were interspersed with the 64 search trials (see Experiment 5, Reeder & Peelen, 2013). These trials were recorded to add to a larger dataset of capture trials from different experiments, and results of these trials are not reported here as they were not the focus of the current paper.

5.3.4 Analysis

We analyzed d prime (*d’*) for the responses on the expertise assessments, reported in Table 1. In the main experiment, we used correlational analyses to compare search task performance with average upright car *d'* and car-bird *d'* values.

Table 1. Mean and median *d'* for the different discrimination tasks of the expertise assessments, with standard error of the mean (SEM). Inversion effects were calculated as upright (car or person) *d'* – inverted (car or person) *d'*.

	Upright Cars	Upright Faces	Car Inversion Effect	Face Inversion Effect	Upright Cars - Birds
Mean	1.50 ± 0.21	2.01 ± 0.11	1.07 ± 0.14	0.76 ± 0.10	0.33 ± 0.19
Median	1.40	1.93	0.91	0.75	0.25

Table 2. Accuracy and RT scores in the search task with standard error of the mean (SEM). Results are further broken down for person and car search.

	Total	Person Search	Car Search
Accuracy	73.3% ± 1.4	71.5% ± 1.5	75.3% ± 1.7
RT	665 ms ± 19	678 ms ± 19	654 ms ± 20

5.4 Results

5.4.1 Expertise assessments

Table 1 reports *d'* values for upright cars (from the upright vs. inverted car discrimination task), upright faces (from the upright vs. inverted faces discrimination task), the inversion effect (upright-inverted *d'*) for cars, the inversion effect for faces, and the *d'* difference between upright cars (from the upright car discrimination task) and birds (from the upright birds discrimination task) on the expertise assessments. We used average upright car *d'* (mean upright car *d'* from the upright vs. inverted car discrimination task and the upright car discrimination task) and car-bird *d'* as our measures of car expertise in the current paper, consistent with previous studies that have used these values as criteria of expertise (Curby & Gauthier, 2009; Gauthier et al, 2000).

5.4.2 Correlational analyses

Previous studies have determined that an upright car *d'* greater than 2 and a car-bird *d'* greater than 1 denotes car expertise. Our main measures of expertise in the correlational analyses were average upright car *d'* and car-bird *d'*. To reduce the influence of outliers, we used Spearman's rho for our correlations.

Visual search

Figure 3 depicts the correlations between car *d'* and visual search accuracy (3a), and car-bird *d'* and visual search accuracy (3b) for car and person search. Correlational analyses on the visual search task revealed a significant positive correlation between car *d'* and car search accuracy, $r(32)=0.576$, $p<0.001$, as well as a positive correlation between car-bird *d'* and car search accuracy, $r(32)=0.399$, $p=0.019$. There was no such correlation between car *d'* and person search accuracy, $r(32)=0.245$, $p=0.163$, nor car-bird *d'* and person search accuracy, $r(32)=0.118$, $p=0.506$. The expert accuracy advantage for car search compared to person search was further confirmed by correlations between car *d'* and the difference in detection accuracy during car versus person search (car-person accuracy). More positive difference scores indicated higher detection accuracy during car search than person search. There was a significant positive correlation between car *d'* and car-person accuracy, *r*(32)=0.435, *p*=0.010, and also with car-bird *d'* and car-person accuracy, *r*(32)=0.401, *p*=0.019. There were no significant correlations between car *d'* or car-bird *d'* and RT for either car or person search (all *ps*>0.057).

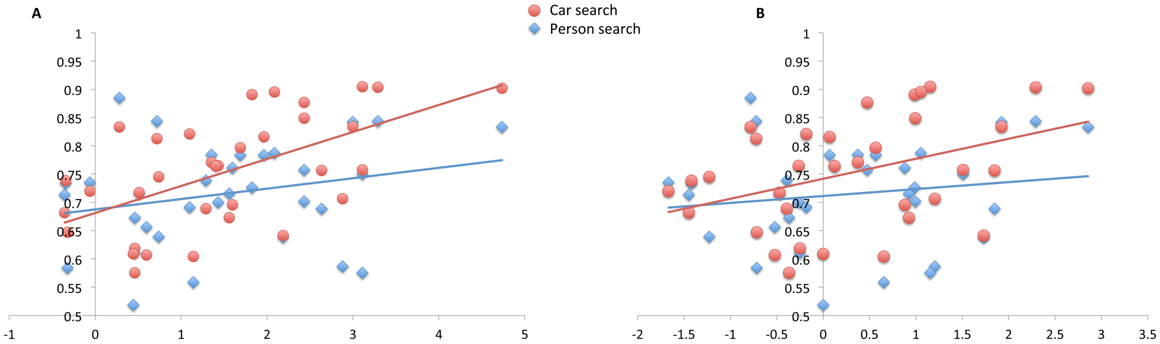


Figure 3. A) The correlation between average upright car discrimination *d'* (x axis) and search accuracy (y axis) during search for people (blue) and cars (red). B) The correlation between car-bird *d'* (x axis) and search accuracy (y axis) during search for people (blue) and cars (red).

5.5 Discussion

The goal of the current study was to determine the generalizability of perceptual expertise to visual search for object categories in naturalistic scenes. Specifically, we were interested in examining the effects of car discrimination expertise on category detection ability in real-world contexts. We used a visual search paradigm to evaluate the accuracy of car and person detection in natural scenes. Results revealed a performance advantage for car experts during car detection but not person detection. The expert ability to individuate cars correlates with a category-level detection advantage, in line with a previous study by Hershler and Hochstein (2009). Thus, individuation training seems to improve category-level detection, suggesting these two tasks may rely on the same preparatory representations.

If our claim is correct that individuation training facilitates detection, then detection advantages should be observed for other categories of perceptual expertise. As previously stated, human faces are a natural category of expertise: they are individuated as fast as they are categorized (Tanaka, 2001) and are processed configurally (Maurer, Le Grand, & Mondloch, 2002), showing marked discrimination impairments when inverted (Freire, Lee, & Symons, 2000; Rossion & Gauthier, 2002). Faces furthermore show a distinct neural potential approximately 170 ms (N170) after their presentation in a visual display (see Bentin, Allison, Puce, Perez, & McCarthy, 1996 for a review). This is also the case for other objects of expertise (Gauthier, Curran, Curby, & Collins, 2003; Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002; Tanaka & Curran, 2001). Importantly, similar to other objects of expertise, faces show category-level detection advantages, supporting the current results. Natural images of human faces are detected more efficiently among distractors than other objects, including animal faces (Hershler, Golan, Bentin, & Hochstein, 2010; Hershler & Hochstein, 2005).

Our results are in line with several studies on the neural basis of object recognition. Although identifying objects at different category levels requires attention to different features (i.e., categorization requires part-based attention, whereas individuation requires attention to configurations of parts), the cortical regions recruited for these tasks are widely overlapping (Mason & Macrae, 2004;

Tarr & Cheng, 2003). We suggest that these different tasks rely on the same neural representations, which are enhanced (e.g., gain higher "resolution": Scolari, Vogel, & Awh, 2008) by perceptual expertise (Harel, Gilaie-Dotan, Malach, & Bentin, 2010). These findings collectively support the hypothesis that expertise boosts the neural representation of objects, which facilitates object recognition at every level.

Although the current study supports the hypothesis that individuation training implicitly improves detection performance, the results could alternatively be attributed to explicit search training. It is possible that because car experts are highly interested in cars (and therefore look at many different car models), they may also enjoy searching for different models of cars in their surroundings. In this case, perceptual experts may train themselves in both car individuation and detection. The results of several previous studies suggest that intensive search training leads to superior detection ability compared to untrained individuals (Koller, Hardmeier, Michel, & Schwaninger, 2008; Manning, Ethell, Donovan, & Crawford, 2005; McCarley, Kramer, Wickens, Vidoni, & Boot, 2004; Nodine, Kundel, Lauver, & Toto, 1996). It would be interesting for future studies to compare the search strategies of expert searchers and perceptual experts to investigate how they compare in detection performance based on their different training and experience.

5.5.1 Conclusions

The current study tested the extent to which perceptual expertise generalizes to category-level search performance in naturalistic scenes. Results showed that car discrimination expertise is correlated with superior car search performance relative to person search performance. This indicates that perceptual expertise facilitates search even without specific search training. Our results support the hypothesis that different levels of the object recognition process rely on the same representations, which are enhanced by perceptual expertise.

Chapter six: General discussion

The preparatory template activated in preparation to detect object categories in real-world scenes is not well understood. We are highly effective in using top-down knowledge of familiar categories and contexts during search, as evidenced by our rapid and almost attention-free ability to detect targets in cluttered, naturalistic scenes. But what is this top-down knowledge? When searchers are cued to detect people in a photograph, what features can they prepare for? They cannot prepare for a particular person, specific clothing or colors of clothing, a certain posture, position, or size. They do not know whether there will be multiple people in the photograph, whether people will be near or far in perspective, or occluded by trees or other objects. The main objective of this thesis was to shed light on our ability to prepare for targets under such conditions by studying the contents, characteristics, neural underpinnings, and individual differences involved in the preparatory template for object categories embedded in real-world contexts.

The second and third chapters of this thesis were two papers in which we showed that an efficient search template for naturalistic presentations of object categories is comprised of a set of view-invariant shapes of category-diagnostic object parts represented globally across the visual field, activated prior to search and deployed automatically to template-matching features in the environment. The fourth chapter of this thesis revealed that efficient category-level search requires the preparatory activation of high-level, object-selective cortex. In Experiment 1 of the fourth chapter, we found evidence that searchers use a range of strategies that recruit different neural areas in preparation for search, consistent with previous work (Peelen & Kastner, 2011). Searchers who rely more on early visual cortex (EVC) during preparation for category-level search show worse detection performance compared to searchers who rely more on high-level, object-selective cortex (such as right posterior temporal cortex, or pTC). Considering the results of the paper reported in the fifth chapter of this thesis, it is likely that perceptual experts activate more enhanced representations of their category of expertise in object-selective cortex compared to novices, which further optimizes search performance.

In sum, the studies reported in Chapters 2 and 3 provide evidence for the contents and characteristics of the search template for real-world categories, whereas Chapter 4 elucidates the critical neural regions involved in such a search template. Experiment 1 of Chapter 4 and Chapter 5 provide evidence for a range of search strategies from poor, to efficient, to optimal.

6.1 The search template for category-level search in naturalistic scenes

6.1.1 Contents and characteristics

It is interesting to note that prior to the studies presented in this thesis, there was no experiment-based evidence that the search template for categories in real-world scenes is composed of view-invariant, spatially global shapes of diagnostic object parts. In a review of the visual search literature, Treisman suggested there is a special role for real-world search templates but only provided a guess as to what kinds of features are needed to perform this task (Treisman, 2006). Treisman proposed that humans are predisposed to detect familiar object categories in natural environments, for which we have immense experience. The simplest features that allow us to differentiate one category from another are diagnostic parts of objects such as the arms and legs of people (which simultaneously apply to most members of the category "people" and distinguish people from all other categories). These naturalistic features may be represented as efficiently as "simple" features (such as a target color) when the task requires it. This hypothesis is supported by evidence that category-level detection in real-world scenes is nearly attention free (e.g., Li et al., 2002; Peelen et al., 2009). Experimental evidence for a template composed of diagnostic part features was found in Experiment 4 (capture by parts of objects vs. whole objects) of Chapter 2 of this thesis.

Several previous studies have found that both humans (Silvanto, Schwarzkopf, Gilaie-Dotan, & Rees, 2010) and monkeys (Booth & Rolls, 1998; Li & DiCarlo, 2008) activate view-invariant representations of objects, in that targets are processed efficiently even if there are changes in size (Fiser & Biederman, 2001; Vickery et al., 2005), orientation (Biederman & Bar, 1999; Vickery et al., 2005), or occlusion (Behrmann, Zemel, & Mozer, 1998; Moore, Yantis, & Vaughan, 1998) from one presentation to the next. Because the visual system is so good at processing objects in many different views, it is likely that it represents parts of objects rather than whole objects in particular configurations, also consistent with Treisman's account (Treisman, 2006) and computational accounts (Ullman et al., 2002). This additionally supports the hypothesis that searchers activate view-invariant templates for objects that can appear in various views across diverse natural photographs. Indeed, Experiments 2 (capture by upright vs. inverted objects) and 3 (capture by upright vs. rotated objects) of Chapter 2 of the current thesis suggest that we activate view-invariant search templates for category-level search in naturalistic contexts.

One behavioral study found evidence that we preferentially represent low-frequency information over high-frequency information for category-level processing (Collin & McMullen, 2005). This corresponds to Experiment 1 of the study presented in Chapter 2 of this thesis, which suggests that shape features are more informative for category-level detection than surface features. This is further supported by computational studies that have found that categorization performance receives the greatest benefit when it is based on the detection of shapes of "intermediate" object fragments, suggesting that shape features provide the most diagnostic information about a category (Ullman et al., 2002).

Probably the least-studied aspect of the search template for real-world category detection is its spatial attention properties. The results of previous studies indicate that naturalistic category detection consumes surprisingly little attention resources and occurs prior to feature binding to locations (Evans & Treisman, 2005; Li et al., 2002; Peelen et al., 2009), both of which are attributes of efficient, parallel processing. Therefore, it is likely that target features of real-world categories are detected in parallel across the visual field (Rousselet, Fabre-Thorpe, & Thorpe, 2002). The results of Experiment 5 (capture by objects in search-irrelevant locations) in Chapter 2, and Experiments 1-2b in Chapter 3 of the current thesis provide additional support for a spatially global search template.

In sum, Chapter 2 of this thesis brings together piecemeal evidence from many of the studies mentioned here to provide cohesive evidence for a search template composed of a set of view-invariant shapes of diagnostic object parts represented globally across the visual field. In our second study (Chapter 3), we tested the reliability and validity of the paradigm we developed for the experiments reported in Chapter 2 and found results that strengthened our initial findings, particularly that the search template is composed of visual features of object categories that are represented spatially globally. The first two experimental chapters of this thesis are therefore essential in driving future research on the contents and characteristics of the search template.

6.1.2 Neural underpinnings

Few studies have explored the neural regions involved in the search template for real-world object categories (Çukur, Nishimoto, Huth, & Gallant, 2013; Peelen et al., 2009; Peelen & Kastner, 2011, Soon et al., 2012). In this thesis, I described two experiments in Chapter 4 that provide evidence for a causal involvement of high-level, objective-selective cortex (right posterior temporal cortex, or pTC) during the preparatory phase of category-level search in naturalistic scenes.

Previous behavioral studies have suggested that different category levels of search rely on different templates (e.g., Bravo & Farid, 2012). For example, a template for a specific target (e.g., my sister) can be composed of specific attributes of that target (e.g., blond hair, big teeth) whereas a template for a target category (e.g., people) must be composed of attributes that can generalize across various exemplars (e.g., general shapes of body parts). We found that templates that rely on different types of features (i.e., low-level or high-level features) also recruit different neural regions (Experiment 2 of Chapter 4), consistent with previous studies that have found correlations between the strength of category-selective BOLD activity in object-selective cortex and performance on category detection tasks (Peelen & Kastner, 2011; Soon et al., 2012), and the strength of BOLD activity in EVC and performance on "simple" feature detection tasks (Ress et al., 2000). These studies collectively suggest that we recruit different neural regions (and thus, cognitive strategies) to search at different category levels.

Another question is whether these visual regions are activated prior to search in preparation for targets, or whether they are activated only once attention falls on a search target in the environment. A previous neuroimaging study provides evidence for the former (Peelen & Kastner, 2011); searchers were cued to detect people and cars in natural scenes, but some trials only presented a fixation cross following the cue, on which no scene ever appeared. The authors found category-selective BOLD signal even on fixation-only trials, suggesting purely top-down, category-selective activity in high-level visual neural areas. It is therefore likely that neural regions involved in the search template are active prior to stimulus onset, but this has not yet been causally determined save for one study reported in this thesis (Chapter 4). In Experiment 1 of this study, we found evidence that a disruption of category-selective neural activity specifically prior to stimulus onset impairs goal-driven search performance in those searchers who use high-level search strategies for category detection. In Experiment 2 of this study, we found that pre-stimulus TMS to object-selective cortex selectively impairs category-level search performance but not individual-level search performance. These experiments collectively suggest that efficient real-world search requires the preparatory activation of high-level object features in object-selective visual cortex.

6.1.3 Individual differences

One of the most interesting findings of the current thesis was that searchers show widespread individual differences in the neural regions (thus, the contents and strategies) involved in preparing for category detection, as reported in Experiment 1 of Chapter 4. Although these findings are consistent with two previous studies of naturalistic search situations (Peelen & Kastner, 2011; Soon et al., 2012), why some people use strategies that are detrimental to performance is not well understood. Poor search strategies are associated with lower working memory capacity (Kane et al., 2001) and worse attention orienting abilities (Giesbrecht et al., 2006), but may additionally be due to some searchers being unaware of “bad habit” shortcut strategies, since previous studies suggest that it is possible to train people to be better searchers (Koller et al., 2008; McCarley et al., 2004).

Looking at the different ways that people search could provide valuable information in determining what strategies lead to detrimental, efficient, or optimal search performance. Chapter 5 of the current thesis reported a paper in which expertise for discriminating members of a particular object category (cars) correlated strongly with detection accuracy for that category compared to other categories. Perceptual experts likely represent exemplars from their category of expertise in a similar way to novices, but are able to use the same attention resources more efficiently, which leads to optimal search performance. It is possible that perceptual experts activate a variety of highly salient visual features of the category of expertise that are boosted above the representations activated by novices. Future research would need to explore this possibility further.

It is important to explore the template features and neural regions that are recruited by poor, efficient, and optimal searchers to determine the crucial differences between them. Not only would this allow us to gain a better understanding of various search strategies and the neural regions associated with them, but it would also open the door to creating better models of human attention for both healthy and clinical populations, improved machine (or machine-assisted) search, and better search training for professions that require it (e.g., radiology, airport security).

6.2 Future directions

The first and second chapters of this thesis dealt with the contents and characteristics of the search template for categories embedded in naturalistic scenes. It would be an interesting future avenue to investigate how these aspects of the search template change depending on the task; for example, how does the template change during search in more naturalistic situations such as moving scenes, real environments, or scenes containing multimodal information (e.g., audio-visual), or alternatively, during search in repeated or more artificial displays? The template we explored in our recent studies was shaped for quite a specific stimulus set (color photographs presented around a fixation cross on a computer screen), so we cannot rule out that the contents and characteristics we observed thus far are specific to the tasks we presented.

It is also important to explore further into the critical timing of search template formation and the interactions between different neural regions involved in preparatory activity for search. Chapter 4 used TMS to investigate some aspects of timing, but electroencephalography (EEG) or magnetoencephalography (MEG) would allow us to explore neural network activity across time, providing a bigger picture than TMS, which can only provide information about one or two small neural regions and specific time windows involved in the search template.

Finally, Experiment 1 of Chapters 4 and Chapter 5 of this thesis explored individual differences in search abilities, which opens up many new avenues for future research on these topics. For example, it would be interesting to investigate why some people are better natural searchers than others, and how training (e.g., discrimination training, search training) can affect search behavior and neural activity. Related to this is how visual search strategies change over a lifespan of natural visual experience. How do different levels of visual experience play a role in the contents and characteristics of, and neural regions recruited for, templates that are activated for search in scenes (e.g., what features and neural regions are recruited by young children compared to adults when preparing for search)?

6.3 Conclusions

This thesis explored the contents, characteristics, neural regions, and individual differences associated with the search template for familiar object categories embedded in real-world photographs. I provide evidence that such a search template is composed of a collection of view-invariant shapes of diagnostic object parts represented globally across the visual field, activated prior to search and deployed automatically to matching features in the environment. The category-level search template recruits right posterior temporal cortex (pTC) to a greater extent than early visual cortex (EVC), whereas the template activated for detection of specific targets does not show such a dissociation between these two regions, suggesting a unique role of pTC in high-level preparatory attention. However, search ability differs across individuals and so does the reliance on pTC, in that better searchers recruit pTC to a greater extent than EVC prior to search. Finally, people who are experts at discriminating exemplars from a particular object category are more efficient at detecting objects from their category of expertise compared to other categories, even without search training. These results collectively support the hypothesis that poor and efficient searchers activate different features and neural regions in preparation for category-level search, whereas the difference between an efficient and optimal searcher is in the efficiency with which one is able to recruit attention resources. The papers reported here expand on a little-explored area of research, at the meeting point between the study of preparatory templates, naturalistic search, and individual differences.

References

Ariga, A., & Yokosawa, K. (2008). Contingent attentional capture occurs by activated target congruence. *Percept Psychophys, 70*(4), 680-687. doi: 10.3758/PP.70.4.680

Bansal, A. K., Madhavan, R., Agam, Y., Golby, A., Madsen, J. R., & Kreiman, G. (2014). Neural Dynamics Underlying Target Detection in the Human Brain. *J Neurosci, 34*(8), 3042-3055. doi: 10.1523/Jneurosci.3781-13.2014

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Schmidt, A. M., . . . Halgren, E. (2006). Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A, 103*(2), 449-454. doi: 10.1073/pnas.0507062103

Barceló, F., Suwazono, S., & Knight, R. T. (2000). Prefrontal modulation of visual processing in humans. *Nat Neurosci, 3*(4), 399-403. doi: 10.1038/73975

Beck, D. M., & Kastner, S. (2009). Top-down and bottom-up mechanisms in biasing competition in the human brain. *Vision Res, 49*(10), 1154-1165. doi: 10.1016/J.Visres.2008.07.012

Beckers, G., & Hömberg, V. (1991). Impairment of Visual-Perception and Visual Short-Term-Memory Scanning by Transcranial Magnetic Stimulation of Occipital Cortex. *Exp Brain Res, 87*(2), 421-432. doi: 10.1007/BF00231859

Behrmann, M., Zemel, R. S., & Mozer, M. C. (1998). Object-based attention and occlusion: evidence from normal participants and a computational model. *J Exp Psychol Hum Percept Perform, 24*(4), 1011-1036. doi: 10.1037/0096-1523.24.4.1011

Biederman, I., & Bar, M. (1999). One-shot viewpoint invariance in matching novel objects. *Vision Res, 39*(17), 2885-2899. doi: 10.1016/S0042-6989(98)00309-5

Booth, M. C., & Rolls, E. T. (1998). View-invariant representations of familiar objects by neurons in the inferior temporal visual cortex. *Cereb Cortex, 8*(6), 510-523. doi: 10.1093/cercor/8.6.510

Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *J Vis, 9*(1), 34 31-39. doi: 10.1167/9.1.34

Bravo, M. J., & Farid, H. (2012). Task demands determine the specificity of the search template. *Atten Percept Psychophys, 74*(1), 124-131. doi: 10.3758/s13414-011-0224-5

Bukach, C. M., Phillips, W. S., & Gauthier, I. (2010). Limits of generalization between categories and implications for theories of category specificity. *Atten Percept Psychophys, 72*(7), 1865-1874. doi: 10.3758/APP.72.7.1865

Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Atten Percept Psychophys, 72*(5), 1283-1297. doi: 10.3758/APP.72.5.1283

Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychon Bull Rev, 15*(4), 795-801. doi: 10.3758/PBR.15.4.795

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*(6427), 345-347. doi: 10.1038/363345a0

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *J Neurophysiol, 80*(6), 2918-2940.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (2001). Responses of neurons in macaque area V4 during memory-guided visual search. *Cereb Cortex, 11*(8), 761-

772. doi: 10.1093/cercor/11.8.761

Collin, C. A., & McMullen, P. A. (2005). Subordinate-level categorization relies on high spatial frequencies to a greater degree than basic-level categorization. *Percept Psychophys, 67*(2), 354-364. doi: 10.3758/BF03206498

Corbetta, M., Miezin, F. M., Dobmeyer, S., Shulman, G. L., & Petersen, S. E. (1990). Attentional modulation of neural processing of shape, color, and velocity in humans. *Science, 248*(4962), 1556-1559. doi: 10.1126/science.2360050

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nat Neurosci, 16*(6), 763-770. doi: 10.1038/nn.3381

Curby, K. M., & Gauthier, I. (2009). The temporal advantage for individuating objects of expertise: perceptual expertise is an early riser. *J Vis, 9*(6), 7.1-13. doi: 10.1167/9.6.7

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *J Exp Psychol Hum Percept Perform, 35*(1), 94-107. doi: 10.1037/0096-1523.35.1.94

David, S. V., Hayden, B. Y., Mazer, J. A., & Gallant, J. L. (2008). Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron, 59*(3), 509-521.

Deco, G., & Rolls, E. T. (2004). A Neurodynamical cortical model of visual attention and invariant object recognition. *Vision Res, 44*(6), 621-642. doi: 10.1016/J.Visres.2003.09.037

de Graaf, T. A., Cornelsen, S., Jacobs, C., & Sack, A. T. (2011). TMS effects on subjective and objective measures of vision: Stimulation intensity and pre- versus post-stimulus masking. *Consciousness Cog, 20*(4), 1244-1255. doi: 10.1016/J.Concog.2011.04.012

de Graaf, T. A., Goebel, R., & Sack, A. T. (2012). Feedforward and quick recurrent processes in early visual cortex revealed by TMS? *Neuroimage, 61*(3), 651-659. doi: 10.1016/J.Neuroimage.2011.10.020

Delorme, A., Richard, G., & Fabre-Thorpe, M. (2010). Key visual features for rapid categorization of animals in natural scenes. *Front Psychol, 1*, 21. doi: 10.3389/fpsyg.2010.00021

Delorme, A., Rousselet, G. A., Macé, M. J., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Brain Res Cogn Brain Res, 19*(2), 103-113. doi: 10.1016/j.cogbrainres.2003.11.010

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci, 18*, 193-222. doi: 10.1146/annurev.ne.18.030195.001205

De Weerd, P., Peralta, M. R., Desimone, R., & Ungerleider, L. G. (1999). Loss of attentional stimulus selection after extrastriate cortical lesions in macaques. *Nat Neurosci, 2*(8), 753-758. doi: 10.1038/11234

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *J Exp Psychol Gen, 115*(2), 107-117. doi: 10.1037/0096-3445.115.2.107

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychol Sci, 11*(6), 467-473. doi: 10.1111/1467-9280.00290

Duncan, J. (1980). The Locus of Interference in the Perception of Simultaneous Stimuli. *Psychol Rev, 87*(3), 272-300. doi: 10.1037/0033-295x.87.3.272

Duncan, J. (1983). Category Effects in Visual-Search - a Failure to Replicate the Oh-Zero Phenomenon. *Percept Psychophys, 34*(3), 221-232. doi: 10.3758/Bf03202949

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychol Rev, 96*(3), 433-458. doi: 10.1037/0033-295X

Evans, K. K., & Treisman, A. (2005). Perception of objects in natural scenes: is it really attention free? *J Exp Psychol Hum Percept Perform, 31*(6), 1476-1492. doi: 10.1037/0096-1523.31.6.1476

Findlay, J. M., & Gilchrist, I. D. (2003). *Active Vision: the Psychology of Looking and Seeing*. Oxford University Press, Oxford. doi: 10.1002/acp.1019

Fiser, J., & Biederman, I. (2001). Invariance of long-term visual priming to scale, reflection, translation, and hemisphere. *Vision Res, 41*(2), 221-234. doi: 10.1016/S0042-6989(00)00234-0

Folk, C. L., Leber, A. B., & Egeth, H. E. (2002). Made you blink! Contingent attentional capture produces a spatial blink. *Percept Psychophys, 64*(5), 741-753. doi: 10.3758/BF03194741

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *J Exp Psychol Hum Percept Perform, 18*(4), 1030-1044. doi: 10.1037//0096-1523.18.4.1030

Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual salience in scene perception? *Perception, 36*(8), 1123-1138. doi: 10.1068/p5659

Freire, A., Lee, K., & Symons, L. A. (2000). The face-inversion effect as a deficit in the encoding of configural information: Direct evidence. *Perception*, *29*(2), 159. doi: 10.1068/p3012

Fuster, J. M. (1990). Inferotemporal units in selective visual attention and short-term memory. *J Neurophysiol, 64*(3), 681-697.

Fuster, J. M., & Jervey, J. P. (1982). Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *J Neurosci, 2*(3), 361-375.

Gauthier, I., & Curby, K. M. (2005). A perceptual traffic jam on highway N170: Interference between face and car expertise. *Curr Dir Psychol Sci, 14*(1), 30-33. doi: 10.1111/j.0963-7214.2005.00329.x

Gauthier, I., Curran, T., Curby, K. M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nat Neurosci, 6*(4), 428-432. doi: 10.1038/nn1029

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nat Neurosci*, *3*(2), 191-197. doi:10.1038/72140

Gauthier, I., & Tarr, M. J. (2002). Unraveling mechanisms for expert object recognition: bridging brain activity and behavior. *J Exp Psychol Hum Percept Perform, 28*(2), 431-446. doi: 10.1037/0096-1523.28.2.431

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training 'greeble' experts: a framework for studying expert object recognition processes. *Vision Res, 38*(15-16), 2401-2428. doi: 10.1016/S0042-6989(97)00442-2

Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D'Esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *J Cogn Neurosci, 17*(3), 507-517. doi: 10.1162/0898929053279522

Ghose, T., & Liu, Z. (2013). Generalization between canonical and non-canonical views in object recognition. *J Vis, 13*(1). doi: 10.1167/13.1.1

Giesbrecht, B., Weissman, D. H., Woldorff, M. G., & Mangun, G. R. (2006). Pre-target activity in visual cortex predicts behavioral performance on spatial and feature attention tasks. *Brain Res, 1080*(1), 63-72. doi: 10.1016/j.brainres.2005.09.068

Gleitman, H., & Jonides, J. (1978). The effect of set on categorization in visual search. *Percept Psychophys, 24*(4), 361-368. doi: 10.3758/BF03204254

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends Neurosci, 15*(1), 20-25. doi: 10.1016/0166-2236(92)90344-8

Grill-Spector, K., & Kanwisher, N. (2005). Visual recognition: as soon as you know it is there, you know what it is. *Psychol Sci, 16*(2), 152-160. doi: 10.1111/j.0956-7976.2005.00796.x

Hannus, A., van den Berg, R., Bekkering, H., Roerdink, J. B. T. M., & Cornelissen, F. W. (2006). Visual search near threshold: Some features are more equal than others. *J Vis, 6*(4), 523-540. doi: 10.1167/6.4.15

Harel, A., Gilaie-Dotan, S., Malach, R., & Bentin, S. (2010). Top-down engagement modulates the neural expressions of visual expertise. *Cereb Cortex*, *20*(10), 2304-2318. doi: 10.1093/cercor/bhp316

Harel, A., Ullman, S., Harari, D., & Bentin, S. (2011). Basic-level categorization of intermediate complexity fragments reveals top-down effects of expertise in visual perception. *J Vis, 11*(8), 18. doi: 10.1167/11.8.18

Hershler, O., Golan, T., Bentin, S., & Hochstein, S. (2010). The wide window of face detection. *J Vis*, *10*(10), 21. doi: 10.1167/10.10.21

Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Res*, *45*(13), 1707-1724. doi: 10.1016/j.visres.2004.12.021

Hershler, O., & Hochstein, S. (2009). The importance of being expert: top-down attentional control in visual search with photographs. *Atten Percept Psychophys, 71*(7), 1478-1486. doi: 10.3758/APP.71.7.1478

Herzmann, G., & Curran, T. (2011). Experts' memory: an ERP study of perceptual expertise effects on encoding and recognition. *Mem Cognit, 39*(3), 412-432. doi: 10.3758/s13421-010-0036-1

Houtkamp, R., & Roelfsema, P. R. (2009). Matching of visual input to only one item at any one time. *Psychol Res, 73*(3), 317-326. doi: 10.1007/s00426-008-0157-3

Huang, L., Mo, L., & Li, Y. (2012). Measuring the interrelations among multiple paradigms of visual attention: an individual differences approach. *J Exp Psychol Hum Percept Perform, 38*(2), 414-428. doi: 10.1037/a0026314

Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *J Vis, 9*(5), 25.21-18. doi: 10.1167/9.5.25

Jacobs, C., de Graaf, T. A., Goebel, R., & Sack, A. T. (2012). The temporal dynamics of early visual cortex involvement in behavioral priming. *PLoS One*, *7*(11): e48808. doi: 10.1371/journal.pone.0048808

Jacobs, C., de Graaf, T. A., & Sack, A. T. (2014). Two distinct neural mechanisms in early visual cortex determine subsequent visual processing. *Cortex, 59*, 1-11. doi: 10.1016/j.cortex.2014.06.017

Jacobs, C., Goebel, R., & Sack, A. T. (2012). Visual awareness suppression by pre-stimulus brain stimulation; a neural effect. *Neuroimage, 59*(1), 616-624. doi: 10.1016/J.Neuroimage.2011.07.090

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of working-memory capacity. *J Exp Psychol Gen, 130*(2), 169-183. doi: 10.1037//0096-3445.130.2.169

Kastner, S., De Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *J Neurophysiol, 86*(3), 1398-1411.

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron, 22*(4), 751-761. doi: 10.1016/S0896-6273(00)80734-5

Koller, S., Hardmeier, D., Michel, S., & Schwaninger, A. (2008). Investigating training, transfer and viewpoint effects resulting from recurrent CBT of X-Ray image interpretation. *Journal of Transportation Security, 1*(2), 81-106. doi: 10.1007/s12198-007-0006-4

Korjoukov, I., Jeurissen, D., Kloosterman, N. A., Verhoeven, J. E., Scholte, H. S., & Roelfsema, P. R. (2012). The Time Course of Perceptual Grouping in Natural Scenes. *Psychol Sci, 23*(12), 1482-1489. doi: 0956797612443832

Krueger, L. E. (1984). The Category Effect in Visual-Search Depends on Physical Rather Than Conceptual Differences. *Percept Psychophys, 35*(6), 558-564. doi: 10.3758/Bf03205953

Kuo, B. C., Stokes, M. G., & Nobre, A. C. (2012). Attention modulates maintenance of representations in visual short-term memory. *J Cogn Neurosci, 24*(1), 51-60. doi: 10.1162/jocn_a_00087

Large, M. E., Kiss, I., & McMullen, P. A. (2004). Electrophysiological correlates of object categorization: back to basics. *Brain Res Cogn Brain Res, 20*(3), 415-426. doi:

10.1016/j.cogbrainres.2004.03.013

Laycock, R., Crewther, D. P., Fitzgerald, P. B., & Crewther, S. G. (2007). Evidence for fast signals and later processing in human V1/V2 and V5/MT+ : A TMS study of motion perception. *J Neurophysiol, 98*(3), 1253-1262. doi: 10.1152/Jn.00416.2007

Lewis, M. B., & Ellis, H. D. (2003). How we detect a face: A survey of psychological evidence. *Int J Imag Syst Tech, 13*(1), 3-7. doi: 10.1002/Ima.10040

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proc Natl Acad Sci U S A, 99*(14), 9596-9601. doi: 10.1073/pnas.092277599

Li, N., & DiCarlo, J. J. (2008). Unsupervised natural experience rapidly alters invariant object representation in visual cortex. *Science, 321*(5895), 1502-1507. doi: 10.1126/science.1160028

Lien, M. C., Ruthruff, E., Goodin, Z., & Remington, R. W. (2008). Contingent attentional capture by top-down control settings: converging evidence from event-related potentials. *J Exp Psychol Hum Percept Perform, 34*(3), 509-530. doi: 10.1037/0096-1523.34.3.509

Loftus, G. R., & Masson, M. E. J. (1994). Using Confidence-Intervals in within-Subject Designs. *Psychon Bull Rev, 1*(4), 476-490. doi: 10.3758/Bf03210951

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol, 77*(1), 24-42.

Macé, M. J., Joubert, O. R., Nespoulous, J. L., & Fabre-Thorpe, M. (2009). The time-course of visual categorizations: you spot the animal faster than the bird. *PLoS One, 4*(6), e5927. doi: 10.1371/journal.pone.0005927

Mack, M.L., Sadr, J., Gauthier, I., & Palmeri, T.J. (2008). Object detection and basic-level categorization: Sometimes you know it is there before you know what it is. *Psychon Bull Rev, 15*(1), 28-25. doi: 10.3758/PBR.15.1.28

MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *J Abnorm Psychol, 95*(1), 15-20. doi: 10.1037/0021-843X.95.1.15

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: evidence from eye movements. *J Vis, 9*(11), 8.1-13. doi: 10.1167/9.11.8

Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *J Vis, 10*(2), 4.1-11. doi: 10.1167/10.2.4

Manning, D., Ethell, S., Donovan, T., & Crawford, T. (2006). How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography, 12*(2), 134-142. doi: 10.1016/j.radi.2005.02.003

Markman, A. B., & Wisniewski, E. J. (1997). Similar and different: The differentiation of basic-level categories. *J Exp Psychol Learn Mem Cogn*, *23*(1), 54. doi: 10.1037/0278-7393.23.1.54

Mason, M. F., & Macrae, C. N. (2004). Categorizing and individuating others: The neural substrates of person perception. *J Cogn Neurosci*, *16*(10), 1785-1795. doi: 10.1162/0898929042947801

Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends Cogn Sci*, *6*(6), 255-260. doi: 10.1016/S1364-6613(02)01903-4

McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., & Boot, W. R. (2004). Visual skills in airport-security screening. *Psychol Sci, 15*(5), 302-306. doi: 10.1111/j.0956-7976.2004.00673.x

Miyashita, Y., & Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature, 331*(6151), 68-70. doi: 10.1038/331068a0

Moore, C. M., Yantis, S., & Vaughan, B. (1998). Object-based visual selection: Evidence from perceptual completion. *Psychol Sci, 9*(2), 104-110. doi: 10.1111/1467-9280.00019

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*(4715), 782-784. doi: 10.1126/science.4023713

Motter, B. C. (1993). Focal Attention Produces Spatially Selective Processing in Visual Cortical Areas V1, V2, and V4 in the Presence of Competing Stimuli. *J Neurophysiol, 70*(3), 909-919.

Myles-Worsley, M., Johnston, W. A., & Simons, M. A. (1988). The influence of expertise on X-ray image processing. *J Exp Psychol Learn Mem Cogn, 14*(3), 553-557. doi: 10.1037/0278-7393.14.3.553

Nodine, C. F., Kundel, H. L., Lauver, S. C., & Toto, L. C. (1996). Nature of expertise in searching mammograms for breast masses. *Acad Radiol, 3*(12), 1000-1006. doi: 10.1016/S1076-6332(96)80032-8

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature, 401*(6753), 584-587. doi: 10.1038/44134

Olivers, C. N., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: visual working memory content affects visual attention. *J Exp Psychol Hum Percept Perform, 32*(5), 1243-1265. doi: 10.1037/0096-1523.32.5.1243

Olshausen, B. A., Anderson, C. H., & Vanessen, D. C. (1993). A Neurobiological Model of

Visual-Attention and Invariant Pattern-Recognition Based on Dynamic Routing of Information. *J Neurosci, 13*(11), 4700-4719.

Op de Beeck, H. P., & Baker, C. I. (2010). The neural basis of visual object learning. *Trends Cogn Sci*, *14*(1), 22-30. doi: 10.1016/j.tics.2009.11.002

Peelen, M. V., Fei-Fei, L., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature, 460*(7251), 94-97. doi: 10.1038/nature08103

Peelen, M. V., & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proc Natl Acad Sci U S A, 108*(29), 12125-12130. doi: 10.1073/pnas.1101042108

Peelen, M. V., & Kastner, S. (2014). Attention in the real world: toward understanding its neural basis. *Trends Cogn Sci*. doi: 10.1016/j.tics.2014.02.004

Pomplun, M. (2006). Saccadic selectivity in complex visual search displays. *Vision Res, 46*(12), 1886-1900. doi:10.1017/j.visres.2005.12.003

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Res, 1301*, 89-99. doi: 10.1016/j.brainres.2009.08.085

Reeder, R. R., & Peelen, M. V. (2013). The contents of the search template for category-level search in natural scenes. *J Vis, 13*(3), 13. doi: 10.1167/13.3.13

Ress, D., Backus, B. T., & Heeger, D. J. (2000). Activity in primary visual cortex predicts performance in a visual detection task. *Nat Neurosci, 3*(9), 940-945. doi: 10.1038/78856

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nat Neurosci, 3*(11), 1199-1204. doi: 10.1038/81479

Robbins, R., & McKone, E. (2007). No face-like processing for objects-of-expertise in three behavioural tasks. *Cognition, 103*(1), 34-79. doi: 10.1016/j.cognition.2006.02.008

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychol*, *8*(3), 382-439.

Rossi, S., Hallett, M., Rossini, P. M., Pascual-Leone, A., & Safety of, T. M. S. C. G. (2009). Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. *Clin Neurophysiol, 120*(12), 2008-2039. doi: 10.1016/j.clinph.2009.08.016

Rossion, B., & Curran, T. (2010). Visual expertise with pictures of cars correlates with RT magnitude of the car inversion effect. *Perception, 39*(2), 173-183.

Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behav Cogn Neurosci Rev*, *1*(1), 63-75. doi: 10.1177/1534582302001001004

Rossion, B., Gauthier, I., Goffaux, V., Tarr, M. J., & Crommelinck, M. (2002). Expertise training with novel objects leads to left-lateralized facelike electrophysiological responses. *Psychol Sci*, *13*(3), 250-257. doi: 10.1111/1467-9280.00446

Rousselet, G. A., Fabre-Thorpe, M., & Thorpe, S. J. (2002). Parallel processing in high-level categorization of natural images. *Nat Neurosci, 5*(7), 629-630. doi: 10.1038/nn866

Rousselet, G. A., Macé, M. J., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *J Vis, 3*(6), 440-455. doi: 10:1167/3.6.5

Russell, B. C., Torralba, A., Murphy, K. P., & Freeman, W. T. (2008). LabelMe: A database and web-based tool for image annotation. *Int J Comput Vision, 77*(1-3), 157-173. doi:

10.1007/S11263-007-0090-8

Samaria, F. S., & Harter, A. C. (1994). *Parameterisation of a stochastic model for human face identification.* Paper presented at the 1994 Second IEEE Workshop on Applications of Computer Vision.

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Q J Exp Psychol, 62*(10), 1904-1914. doi: 10.1080/17470210902853530

Schreuder, R., d'Arcais, G. B. F., & Glazenborg, G. (1984). Effects of perceptual and conceptual similarity in semantic priming. *Psychol Res, 45*(4), 339-354. doi: 10.1007/BF00309710

Schwarzbach, J. (2011). A simple framework (ASF) for behavioral and neuroimaging experiments based on the psychophysics toolbox for MATLAB. *Behav Res Methods, 43*(4), 1194-1201. doi: 10.3758/s13428-011-0106-8

Scolari, M., Vogel, E. K., & Awh, E. (2008). Perceptual expertise enhances the resolution but not the number of representations in working memory. *Psychon Bull Rev, 15*(1), 215-222. doi: 10.3758/PBR.15.1.215

Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron, 55*(2), 301-312. doi: S0896-6273(07)00445-X [pii]10.1016/j.neuron.2007.06.015

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proc Natl Acad Sci U S A, 104*(15), 6424-6429. doi: 10.1073/Pnas.0700622104

Silvanto, J., Schwarzkopf, D. S., Gilaie-Dotan, S., & Rees, G. (2010). Differing causal roles for lateral occipital cortex and occipital face area in invariant shape recognition. *Eur J Neurosci, 32*(1), 165-171. doi: 10.1111/j.1460-9568.2010.07278.x

Sobel, K. V., Gerrie, M. P., Poole, B. J., & Kane, M. J. (2007). Individual differences in working memory capacity and visual search: the roles of top-down and bottom-up processing. *Psychon Bull Rev, 14*(5), 840-845. doi: 10.3758/BF03194109

Soon, C. S., Namburi, P., & Chee, M. W. (2012). Preparatory patterns of neural activity predict visual category search speed. *Neuroimage, 66*, 215-222. doi: 10.1016/j.neuroimage.2012.10.036

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *J Exp Psychol Hum Percept Perform, 31*(2), 248-261. doi: 10.1037/0096-1523.31.2.248

Soto, D., Humphreys, G. W., & Heinke, D. (2006). Working memory can guide pop-out search. *Vision Res, 46*(6-7), 1010-1018. doi: 10.1016/j.visres.2005.09.008

Stein, T., Sterzer, P., & Peelen, M. V. (2012). Privileged detection of conspecifics: evidence from inversion effects during continuous flash suppression. *Cognition, 125*(1), 64-79. doi: 10.1016/j.cognition.2012.06.005

Tanaka, J. W. (2001). The entry point of face recognition: evidence for face expertise. *J Exp Psychol Gen*, *130*(3), 534. doi: 10.1037/0096-3445.130.3.534

Tanaka, J. W., & Curran, T. (2001). A neural basis for expert object recognition. *Psychol Sci, 12*(1), 43-47.

Tanaka, J. W., Curran, T., & Sheinberg, D. L. (2005). The training and transfer of real-world perceptual expertise. *Psychol Sci, 16*(2), 145-151. doi: 10.1111/j.0956-7976.2005.00795.x

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychol, 23*(3), 457-482. doi: 10.1016/0010-0285(91)90016-H

Tarr, M. J., & Cheng, Y. D. (2003). Learning to see faces and objects. *Trends Cogn Sci*, *7*(1), 23-30. doi: 10.1016/S1364-6613(02)00010-4

Theeuwes, J., Reimann, B., & Mortier, K. (2006). Visual search for featural singletons: No top-down modulation, only bottom-up priming. *Vis Cogn, 14*(4-8), 466-489. doi:10.1080/13506280500195110

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520-522. doi: 10.1038/381520a0

Treisman, A. (2006). How the deployment of attention determines what we see. *Vis Cogn, 14*(4-8), 411-443. doi: 10.1080/13506280500195250

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychol, 12*(1), 97-136. doi: 10.1016/0010-0285(80)90005-5

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nat Neurosci, 5*(7), 682-687. doi: 10.1038/nn870

Underwood, G., Foulsham, T., van Loon, E., & Underwood, J. (2005). Visual attention, visual saliency and eye movements during the inspection of natural scenes. In J. Mira and J. R. Alvarez (Eds.), *Artificial intelligence and knowledge engineering applications: a bioinspired approach* (pp. 459-468). Berlin: Springer. doi: 10.1007/11499305_47

VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. *Perception, 30*(6), 655-668. doi: 10.1068/p3029 van Zoest, W., & Donk, M. (2006). Saccadic target selection as a function of time. *Spatial Vision*, *19*(1), 61- 67. doi: 10.1163/156856806775009205

Vetter, P., Grosbras, M.-H., & Muckli, L. (2013). TMS Over V5 Disrupts Motion Prediction. *Cereb Cortex*. doi: 10.1093/cercor/bht297

Vickery, T. J., King, L. W., & Jiang, Y. (2005). Setting up the target template in visual search. *J Vis, 5*(1), 81-92. doi: 10:1167/5.1.8

Vuilleumier, P., Henson, R. N., Driver, J., & Dolan, R. J. (2002). Multiple levels of visual object constancy revealed by event-related fMRI of repetition priming. *Nat Neurosci, 5*(5), 491-499. doi: 10.1038/nn839

Wolfe, J. M. (2007). Guided Search 4.0: Current Progress with a model of visual search. In W. Gray (Ed.), *Integrated Models of Cognitive Systems* (pp. 99-119). New York: Oxford.

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Atten Percept Psychophys, 73*(6), 1650-1671. doi: 10.3758/s13414-011-0153-3

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *J Exp Psychol Hum Percept Perform, 15*(3), 419-433. doi: 10.1037/0096-1523.15.3.419

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Res, 44*(12), 1411-1426. doi: 10.1016/j.visres.2003.11.024

Wyble, B., Folk, C., & Potter, M. C. (2012). Contingent Attentional Capture by Conceptually Relevant Images. *J Exp Psychol Hum Percept Perform*. doi: 2012-30622-001 [pii]

10.1037/a0030517

Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Res, 49*(16), 2095-2103. doi: 10.1016/j.visres.2009.05.017

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *J Exp Psychol Hum Percept Perform, 16*(1), 121. doi: 10.1037/0096-1523.16.1.121

Zhang, W., & Luck, S. J. (2008). Feature-based attention modulates feedforward visual processing. *Nat Neurosci*, *12*(1), 24-25. doi:10.1038/nn.2223