Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Atten Percept Psychophys. Author manuscript; available in PMC 2016 August 01.

Published in final edited form as:

Atten Percept Psychophys. 2015 August ; 77(6): 1881–1895. doi:10.3758/s13414-015-0903-8.

Automatic guidance of attention during real-world visual search

Katharina N. Seidl-Rathkopf[a,b,1], **Nicholas B. Turk-Browne**[a,b], and **Sabine Kastner**[a,b]

[a]Department of Psychology, Princeton University, Princeton, NJ 08544

[b]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544

Abstract

Looking for objects in cluttered natural environments is a frequent task in everyday life. This process can be difficult, as the features, locations, and times of appearance of relevant objects are often not known in advance. A mechanism by which attention is automatically biased toward information that is potentially relevant may thus be helpful. Here we tested for such a mechanism across five experiments by engaging participants in real-world visual search and then assessing attentional capture for information that was related to the search set but was otherwise irrelevant. Isolated objects captured attention while preparing to search for objects from the same category embedded in a scene, as revealed by lower detection performance (Experiment 1A). This capture effect was driven by a central processing bottleneck rather than the withdrawal of spatial attention (Experiment 1B), occurred automatically even in a secondary task (Experiment 2A), and reflected enhancement of matching information rather than suppression of non-matching information (Experiment 2B). Finally, attentional capture extended to objects that were semantically associated with the target category (Experiment 3). We conclude that attention is efficiently drawn towards a wide range of information that may be relevant for an upcoming real-world visual search. This mechanism may be adaptive, allowing us to find information useful for our behavioral goals in the face of uncertainty.

Introduction

Searching for things in our environment is a common task in every day life. Searches can be directed toward different kinds of information, varying from individual objects (e.g., locating your shopping cart in a crowded grocery store) to entire object categories (e.g., finding fresh fruit in the produce section). The selection of relevant information in visual search is thought to be accomplished by matching incoming visual information to an internally generated attentional set (Bundesen, 1990; Duncan & Humphreys, 1989). Visual search appears to be most efficient when the exact appearance of a target is known in advance (Schmidt & Zelinsky, 2009; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004), enabling observers to implement a detailed attentional set. In naturalistic settings, however, visual search is made difficult by a number of uncertainties that are inherent to our typical visual environment.

[1]To whom correspondence should be addressed: Katharina Seidl-Rathkopf Princeton Neuroscience Institute Princeton University Princeton, NJ, 08544 phone: (609) 578-8213 kseidl@princeton.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

First, the appearance of any object in a scene is virtually unconstrained, as it depends on factors such as the perspective from which it is viewed, its distance from the observer, and the degree to which it is occluded by other objects. Second, visual search performance suffers when targets share features with surrounding distracters (Duncan & Humphreys, 1989). This challenge is exacerbated in the real world, where the properties of both targets and non-targets are not always stable across time. For instance, which fruits and vegetables are available depends upon the season. Third, the locations and points in time at which targets appear are often not known in advance.

The first two challenges suggest that searching for objects in the real-world requires an abstract attentional set that is not bound to low-level features and can accommodate large variation in target and distracter appearance. The third challenge suggests that it would be adaptive to have mechanisms that *automatically* bias attention toward objects related to the search target so that they do not go unnoticed when they appear at unforeseen locations or times. The current research aimed to establish the existence of, and investigate the properties of automatic capture by task-relevant information during real-world visual search that requires an abstract attentional set. For this purpose, we assessed the degree to which isolated and novel exemplars from an object category capture attention while participants prepare to rapidly detect the presence of objects from that category in subsequently presented natural scene photographs.

Past research using artificial search displays with relatively simple stimuli has demonstrated that attention is indeed reflexively captured by items that contain target-defining features — a phenomenon known as “contingent attentional capture” (Folk, Remington, & Johnston, 1992). For example, cues that suddenly appear in a display only disrupt the detection of a target at a different location if the target also appears suddenly, as opposed to being revealed by a color change; likewise, color cues only distract when searching for a color target and not an onset target (Folk et al., 1992). These findings show that non-target stimuli can capture attention when they match the current attentional set. Contingent attentional capture was first thought to be dependent exclusively upon the withdrawal of spatial attention from the task-relevant location (Folk, Leber, & Egeth, 2002). While monitoring a central rapid serial visual presentation (RSVP) stream of heterogeneously colored letters for a pre-specified target color, peripheral distracters surrounding the central stream impair detection of an upcoming target if they appear in the target color, but not in a neutral or non-target color. These results suggest that distracters that match the current attentional set divert spatial attention from the relevant target location. If the target follows in short succession, there is insufficient time to re-allocate attention to the relevant location, resulting in more failures to detect the target. In subsequent studies, it was found that capture of non-spatial attention can also be contingent on the attentional set (Folk, Leber, & Egeth, 2008). That is, a distracter sharing target-defining features disrupts detection of a target stimulus even if distracter and target are presented in the same spatial location. These results bear resemblance to the attentional blink (AB), in which the ability to report a second target in an RSVP task is impaired when the first target has been detected 200-600 ms beforehand (Chun & Potter, 1995).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Even though most research investigating contingent attentional capture has focused on relatively simple and predictable target-defining features such as onsets, color, and motion (Folk et al., 2002; Folk & Remington, 1998; Folk et al., 1992; Folk, Remington, & Wright, 1994), there is some evidence that contingent attentional capture may occur for more abstract visual information such as categorical information about objects. In one recent study (Wyble, Folk, & Potter, 2013) participants monitored an RSVP stream for a conceptually defined target such as “amusement park ride” and reported its identity at the end of the trial (e.g., “roller coaster”). Analogous to a standard contingent capture paradigm (Folk et al., 2002) two distracter images were presented in the periphery before the target. When one of these images belonged to the target concept (e.g., “ferris wheel”), target detection performance was impaired. As in the early studies of contingent attentional capture (Folk et al., 2002) this capture effect was found to depend on the withdrawal of spatial attention from the relevant target location. In this study the only factor disqualifying peripheral distracters from acting as targets was the location in which they were presented. Consequently, it is difficult to discern from these results whether capture was fully automatic or whether participants may have actually confused the peripherally presented distracter for the target and voluntarily directed attention towards it.

Another recent study (Reeder & Peelen, 2013) demonstrated contingent capture of spatial attention for isolated objects while participants were preparing to detect the presence of an object category in photographs of natural scenes. On a subset of trials the scenes were replaced by objects from a target and a distracter category, followed by a dot probe. Attentional capture was evident in that reaction times (RTs) on the probe detection task were shortest when the probe was presented in the same spatial location as the target category object. Since the probe detection task was presented only after the distracter displays in this experiment, there was no disincentive to attend to the target category exemplar. In consequence, participants may have voluntarily attended to the target category object rendering inferences regarding the automaticity of the observed capture effect difficult.

Finally, a related line of work is concerned with the effects an item held in working memory (WM) has on the deployment of attention during visual search (Soto, Heinke, Humphreys, & Blanco, 2005; Soto, Hodsoll, Rotshtein, & Humphreys, 2008). WM-driven attentional guidance is typically investigated in dual task paradigms in which participants retain a stimulus in WM while searching for an independent target in a search array. On some trials the search array contains a distracter that matches the WM item, while on others all distracters are novel items. Experiments using simple objects have found that RTs are prolonged on match trials relative to neutral trials, suggesting that the WM representation automatically guides attention towards the matching distracter (Soto et al., 2008). Studies using more complex objects have yielded only mixed results regarding the degree to which attentional capture is dependent on the visual similarity between the WM item and the irrelevant distracters in the search display. One study (Bahrami Balani, Soto, & Humphreys, 2010) found guidance towards distracters that were exemplars of the same category as the WM item. There was also some evidence that distracters that were semantically associated with the WM item captured attention. Other studies, however, found that capture by distracters that matched the WM item category was abolished, when visual similarity between the different exemplars was reduced (Calleja & Rich, 2013). In these experiments,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

both the WM item and the target stimulus tend to rely on similar representations, i.e. they are both objects, in some cases even objects from the same superordinate category (Bahrami Balani et al., 2010). Thus, it can be difficult to disentangle the effects of WM-guided capture and capture by the target-related attentional set. Capture by the WM item, for instance, may be the result of blurring of the two concurrently held representations.

The current study extends the work reviewed above in four important ways. First, for the first time we show that contingent attentional capture by abstract visual information is not fully accounted for by the withdrawal of spatial attention. As is the case for simple features (Folk et al., 2002), visual information that matches an abstract attentional set can capture non-spatial attention. Second, we investigated the extent to which contingent attentional capture can be explained by either enhanced processing of task-relevant information or the suppressed processing of task-irrelevant information. Third, contingent capture was previously examined within the search task itself and for objects that had some task-relevant features (by definition). Thus, it is hard to know whether such capture is purely reflexive or partially volitional. Studies of WM-driven attentional guidance begin to address this issue by measuring attentional capture on a secondary task. However, these studies have yielded mixed results regarding the extent to which attentional capture occurs for abstract visual information. Moreover, in some of these studies it can be difficult to disentangle the effects of the WM item and the search target on attentional capture, especially if they may rely on overlying attentional sets (Bahrami Balani et al., 2010). Critically, in our task the search targets for the primary and secondary tasks were chosen such that the overlap in attentional sets were minimal (people, cars, and trees, vs. triangle). And finally, we demonstrate that attention is not only captured by objects of a target category but also by visual information that is semantically associated with the target category. This is interesting for two reasons. First, it makes a strong case that attentional capture can occur for abstract visual information. Second, it adds to a growing body of research demonstrating that information stored in long-term memory influences attentional guidance (Hutchinson & Turk-Browne, 2012).

Experiment 1A

Our first goal was to establish contingent attentional capture when the overlap between targets and distracters goes beyond simple visual features. For this purpose, we investigated whether task-irrelevant object category exemplars capture attention during preparation for real-world visual search.

Methods

Participants—Twelve members of the Princeton University community (age 22-32, 10 female, normal or corrected-to-normal vision) provided informed consent to a protocol approved by the Institutional Review Board (IRB) of Princeton University. Procedures in all experiments adhered to the Declaration of Helsinki.

Stimuli & Apparatus—Participants were seated 60 cm from a 21” CRT monitor with a refresh rate of 75 Hz. Stimuli were presented using Matlab (The MathWorks, Natick, MA) and the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) against a white background. A

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

black fixation cross subtending 0.5 × 0.5° was presented at the center of the screen throughout the experiment. The category detection task included 1280 grayscale photographs from an online database (Russell, Torralba, Murphy, & Freeman, 2008), which depicted cityscapes and landscapes (Figure 1A). These scenes contained: only people, only cars, people and cars, or neither people nor cars. For every participant, 180 scenes from each of these four types were randomly selected. The scenes were presented centrally and subtended 9 × 7° of visual angle. Perceptual masks consisted of grayscale random noise masks with superimposed naturalistic structure (Walther, Caddigan, Fei-Fei, & Beck, 2009). They were presented centrally and were equal in size to the scenes. Letter cues ("p" for people and "c" for cars) were presented in blue and subtended 0.7 × 0.7°.

Distracter images consisted of grayscale photographs of 20 exemplars from three categories: people, cars, and trees (Figure 1B). The distracters subtended 5 × 5° and were presented at an eccentricity of 7.5° to either the left or right of the fixation cross.

Procedure & Design—An example trial is depicted in Figure 2A. Each trial began with a letter cue, indicating whether the task was to detect people ("p") or cars ("c") in the upcoming scene. The cue was followed by a delay period, ranging from 1200-2090 ms, before presentation of the scene. The duration of the scene was adjusted separately for each participant using a staircasing procedure that was conducted prior to the main experiment (80% accuracy threshold). Across participants, scene presentation durations ranged from 53-80 ms (66 ms on average). Together, the scene and subsequent mask presentation always totaled 400 ms in duration. Participants were instructed to indicate as quickly and accurately as possible whether or not the cued category was present in the scene by pressing one response key for "absent" and another for "present". Trials concluded with an intertrial interval that ranged from 1-1.75 s.

On 75% of trials, a distracter was presented during the delay period. The interstimulus interval (ISI) between the distracter and subsequent scene was 150, 300, 450, 600, or 750 ms. The presentation duration of the distracter was matched to the duration of the scene based on previous contingent attentional capture paradigms (Folk et al., 2002). Distracters were isolated exemplars of people, cars, or trees (Figure 1B). They could thus match (Figure 2B): (1) the currently cued object category (people or cars; cued condition), (2) a previously but not currently cued object category (people or cars; non-cued condition), or (3) a never task-relevant object category (trees; neutral condition).

All independent variables (search task, scene type, distracter type, distracter-scene ISI, and distracter location) were counterbalanced across 320 trials, which were divided into four blocks of 80 trials each. Time-permitting, participants completed either 2 or 3 full sets of counterbalanced trials (i.e., 8 blocks (N= 9) or 12 blocks (N=3)). This procedure ensured that, for each participant, every condition had the same number of trials. The different number of total trials across participants should have no systematic impact on the results, as this was a within-subjects design. Indeed, the pattern of results did not qualitatively change when the same number of trials was included in the analysis for all participants. Trials of all conditions were presented in random order.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Results & Discussion

Accuracy—Figure 3 depicts accuracy on the category detection task as a function of distracter type and distracter-scene ISI. These accuracy scores were submitted to a 4 (distracter type: cued, non-cued, neutral, no distracter) by 5 (ISI: 150, 300, 450, 600, 750) repeated-measures analysis of variance (ANOVA).

The analysis revealed a main effect of ISI [$F(4, 44) = 5.06$, $p = 0.002$, $\eta_p^2 = 0.32$], which was driven by lower accuracy at the shortest ISI compared to all other ISIs (ps < 0.05). Critically, there was also an interaction between ISI and distracter type [$F(12, 132) = 2.37$, $p = 0.009$, $\eta_p^2 = 0.18$]. To understand this interaction, we performed a one-way repeated measures ANOVA at each ISI with distracter type as the factor. At the shortest ISI, there was a main effect of distracter type [$F(3,33) = 5.74$, $p = 0.003$, $\eta_p^2 = 0.34$]. Planned pairwise comparisons revealed that the main effect was due to lower accuracy for the cued distracter compared to all other distracter types (ps < 0.05). In contrast, category detection performance did not depend upon distracter type at any other ISI [300: $F(3, 33) = 1.13$, $p = 0.352$, $\eta_p^2 = 0.09$; 450: $F(3, 33) = 0.61$, $p = 0.614$, $\eta_p^2 = 0.05$; 600: $F(3,33) = 2.03$, $p = 0.128$, $\eta_p^2 = 0.16$; 750: $F(3,33) = 0.88$, $p = 0.459$, $\eta_p^2 = 0.07$].

Signal detection theory measures—To unpack the accuracy results, we also analyzed performance with signal detection measures *d'* and *β*, which characterize sensitivity and criterion, respectively (Figure 4). With respect to *d'*, higher values indicate an increased ability to distinguish signal from noise. The 4 by 5 repeated-measures ANOVA revealed a main effect of ISI [$F(4, 44) = 3.98$, $p = 0.008$, $\eta_p^2 = 0.27$], which was driven by lower sensitivity at the shortest ISI compared to ISIs 450, 600 and 750 (ps< 0.05). Furthermore, there was an interaction between ISI and distracter type [$F(12, 132) = 2.08$, $p = 0.022$, $\eta_p^2 = 0.16$]. Follow-up one-way repeated-measures ANOVAs with the factor distracter type at each individual ISI revealed a significant main effect only at 150 ms [$F(3, 33) = 4.10$, $p = 0.014$, $\eta_p^2 = 0.27$]. This effect was accounted for by lower sensitivity following cued distracters compared to incongruent ($p = 0.012$), neutral ($p = 0.064$), or no distracters ($p = 0.017$). Thus, attentional capture by the cued object category reduced participants' ability to detect target objects.

With respect to *β*, higher values indicate a more conservative response criterion, that is, a reduced false alarm rate at the cost of a lower hit rate. Once more, we observed a main effect if ISI [$F(4, 44) = 4.40$, $p = 0.004$, $\eta_p^2 = 0.29$] that was driven by higher β values at the shortest compared to all other ISIs (ps < 0.08). Furthermore, there was a significant interaction between ISI and distracter type [$F(12, 132) = 2.16$, $p = 0.017$, $\eta_p^2 = 0.16$]. This interaction was driven by a significant main effect of distracter type [$F(3, 33) = 7.09$, $p = 0.001$, $\eta_p^2 = 0.39$] at the shortest ISI but not any of the remaining ISIs. Planned pairwise comparisons revealed that this main effect was driven by *β*s in the no distracter condition compared to neutral, non-cued, and cued distracter types (ps < 0.05). Thus, the appearance of a distracter led to a general shift of becoming more conservative about reporting a target object. This was numerically strongest in the cued distracter condition, suggesting that the lower sensitivity in this condition was caused by a reduction in the hit rate rather than an increase in the false alarm rate.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Reaction Times—Finally, we analyzed RTs in order to test for a speed-accuracy trade-off. RTs for the first two participants were unavailable due to technical issues. A 4 (distracter type: cued, non-cued, neutral, no distracter) by 5 (ISI: 150, 300, 450, 600, 750) repeated-measures ANOVA revealed a main effect of ISI [$F(4, 36) = 31.65$, $p < 0.001$, $\eta_p^2 = 0.78$]. This effect was driven by longer RTs for shorter ISIs (when accuracy was lower) compared to longer ISIs. Furthermore, there was no difference in mean RT across the different distracter types at the 150 ms ISI [$F(3, 27) = 1.844$, $p = 0.163$, $\eta_p^2 = 0.17$]. This pattern of results confirms that the decrement in accuracy for cued distracters at the shortest ISI cannot be explained by a speed-accuracy tradeoff.

In sum, the results from this experiment indicate that distracters that matched the cued object category transiently impaired the participants' ability to detect cued objects within scenes, providing evidence for selective attentional capture by distracters that match an abstract category-based attentional set.

Experiment 1B

Early research has suggested that attentional capture effects arise because the distracters that match the current attentional set lead to a disengagement of spatial attention from the location at which the target will appear (Folk et al., 2002). Later work demonstrated that attentional capture by distracters sharing simple target-defining features can also be driven by a central rather than a spatial bottleneck. In this experiment, we aimed to assess whether the attentional capture by complex visual information observed in Experiment 1A can be fully accounted for by the withdrawal of spatial attention.

Methods

Participants—Twenty-eight Princeton University undergraduates (age 18-22, 18 female, normal or corrected-to-normal vision) provided informed consent to a protocol approved by the Princeton University IRB. Four participants were excluded for not following task instructions (N=1) or for poor overall performance (accuracy < 60%, N = 2; RT > 2 *SD* from the group mean, N = 1). For purposes of counterbalancing, excluded participants were replaced so that the sample included 24 valid datasets. The number of participants in this experiment was increased from Experiment 1A due to the inclusion of an additional condition.

Stimuli & Apparatus—The apparatus was identical to Experiment 1A. To ensure that any differences between conditions were not driven by incidental differences in processing of the three object categories, the assignment of task-relevance to object categories was fully counterbalanced across three groups of participants in this experiment (as opposed to always using the same category, trees, as the neutral condition in the previous experiment). For each group of participants, the 1280 scenes were re-sorted according to the presence of (1) the cued category, (2) the non-cued category, (3) both the cued and non-cued category or (4) neither of the task-relevant categories. Any of the four scene types could randomly contain objects from the neutral category. For each subject, 120 scenes of each scene type were randomly selected from the overall scene database and presented twice over the course of the experiment. Distracter stimuli were identical to those used in Experiment 1A. The

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

critical manipulation was that they were presented either centrally, or 7.5° to the left or right of fixation.

Procedure & Design—The basic task was identical to Experiment 1A with the following exceptions: First, based on the results of Experiment 1A, we focused on only two ISIs: 150 and 650 ms. Second, distracters appeared with equal probability in the periphery or at fixation. Presentation durations were again determined using a staircasing procedure and across participants ranged from 40-90 ms (66 ms on average).

Results & Discussion

Accuracy—Figure 5 depicts accuracy on the category detection task as a function of distracter type, distracter-scene ISI, and distracter location. These accuracy scores were submitted to a 4 (distracter type: cued, non-cued, neutral, no distracter) by 2 (ISI: 150, 600 ms) by 2 (distracter location: central, peripheral) repeated-measures ANOVA. This analysis revealed main effects of distracter location [$F(1, 23) = 5.31$, $p = 0.031$, $\eta_p^2 = 0.19$], distracter type [$F(3, 69) = 5.19$, $p = 0.003$, $\eta_p^2 = 0.18$], and ISI [$F(1, 23) = 26.52$, $p < 0.001$, $\eta_p^2 = 0.54$].

Critically, there was an interaction between ISI and distracter type [$F(3, 69) = 3.81$, $p = 0.014$, $\eta_p^2 = 0.14$]. We conducted simple-effects analyses across distracter type at each ISI using one-way repeated measures ANOVAs. There was a main effect at the short ISI [$F(2.13, 49) = 7.56$, $p = 0.001$, $\eta_p^2 = 0.25$, Greenhouse-Geisser corrected for non-sphericity], which was driven by lower accuracy for cued distracters compared to all other distracter types (ps < 0.01). In contrast, accuracy at the long ISI did not depend upon distracter type [$F(3, 69) = 0.58$, $p = 0.633$, $\eta_p^2 = 0.02$]. The interaction between ISI and distracter type did not further interact with location [$F(3, 69) = 0.38$, $p = 0.767$, $\eta_p^2 = 0.02$], consistent with an interpretation that attentional capture did not depend on the withdrawal of spatial attention.

Signal detection theory measures—To complement the accuracy analysis, *d'* was also analyzed as a function of distracter type, distracter-scene ISI, and distracter location. Most noteworthy, there was again a significant interaction between distracter type and ISI [$F(3, 69) = 4.24$, $p = 0.008$, $\eta_p^2 = 0.16$], which was driven by a main effect of distracter type at the short ISI [$F(3, 69) = 7.10$, $p < 0.001$, $\eta_p^2 = 0.24$]. Planned post-hoc comparisons revealed that *d'* was lower in the cued distracter condition compared to all other distracter conditions (ps < 0.01), suggesting a selective decrease in sensitivity. As observed in the accuracy data, there was no further interaction between distracter type, ISI, and distracter location [$F(3, 69) = 0.65$, $p = 0.583$, $\eta_p^2 = 0.03$].

The pattern of results for β resembled that observed in Experiment 1A. A significant main effect of distracter type [$F(3, 69) = 6.96$, $p < 0.001$, $\eta_p^2 = 0.23$], followed by planned pairwise comparisons, showed that participants adopted a more conservative response criterion following cued distracters compared to all other distracter types (ps < 0.05). A marginally significant interaction between distracter type and ISI [$F(3,69) = 2.31$, $p = 0.084$,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

η_p^2 = 0.09] suggests that this effect was more pronounced at the 150 ms compared to the 650 ms ISI.

Reaction Times—Mean RTs were entered into a 4 (distracter type: cued, non-cued, neutral, no distracter) by 2 (ISI: 150, 600 ms) by 2 (distracter location: central, peripheral) repeated-measures ANOVA to test for a speed-accuracy tradeoff. This analysis yielded a significant main effect of ISI [$F(1, 23) = 66.02$, $p < 0.001$, $\eta_p^2 = 0.74$], which was driven by longer RTs at the 150 ms compared to the 600 ms ISI. There was a significant main effect of distracter type at the 150 ms ISI regardless of distracter location [central: $F(3, 69) = 4.77$, $p = 0.004$, $\eta_p^2 = 0.17$; peripheral: $F(3, 69) = 8.10$, $p < 0.001$, $\eta_p^2 = 0.26$]. In both cases, this effect was accounted for by longer RTs in the cued distracter and no-distracter conditions, compared to other distracter types. This pattern of results is inconsistent with a speed-accuracy tradeoff.

To summarize, Experiment 1B replicated the results of Experiment 1A in showing that object exemplars matching the current attentional set selectively capture attention at short ISIs. This was true regardless of whether the distracter overlapped spatially with the target location. Given that centrally presented distracters impaired performance on the category detection task in our study, explanations of the AB might apply to our findings (in place of, or in addition to, a spatial explanation). In particular, the AB is thought to arise from a central bottleneck, in which encoding of the second target fails when a capacity-limited second stage of processing is busy consolidating the first target (Chun & Potter, 1995; Dehaene, Sergent, & Changeux, 2003; Dux & Marois, 2009; Jolicoeur & Dell'Acqua, 1999).

In classical AB studies, the decrement in target performance lasts for stimulus onset asynchronies (SOAs) of 200 – 500 ms (Martens & Wyble, 2010). Here, we chose the ISIs based on Experiment 1A, in which the only significant effect was observed at the shortest ISI (corresponding on average to an SOA of 216 ms). Consequently, we cannot determine whether the capture effect observed here is comparable in duration to that reported in other studies for the AB. It is not clear whether we would expect it to be, as the timecourse of the AB may vary between simple RSVP-based search tasks and more real-world search tasks as used here. For example, the length of the AB has been shown to depend on target category (Einhauser, Koch, & Makeig, 2007).

Experiment 2A

In Experiments 1A and 1B attentional capture was measured indirectly as a decrement in performance on the real-world visual search task. The question arises whether the observed decrease in detection accuracy arises from truly involuntary attentional capture or rather reflects voluntary orienting towards the cued distracter. Therefore, in Experiment 2A we sought to directly measure attentional capture using a secondary task for which the cued object category was completely irrelevant.

Methods

Participants—Fourteen Princeton University students (age: 19-29, 7 female, normal or corrected-to-normal vision) provided informed consent to a protocol approved by the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Princeton University IRB. Two participants were excluded for performance on the category detection task (accuracy < 60%). For purposes of counterbalancing, excluded participants were replaced so that the sample included 12 valid datasets.

Stimuli & Apparatus—The apparatus was identical to Experiments 1A and 1B. The stimuli used in the category detection task were the same as those used in Experiment 1B. For each participant, 96 exemplars of each of the four scene types were randomly selected from the overall image database.

This experiment also included a new "triangle" task, in which two shapes were presented 2.5° to the left and right of the central fixation cross. The target in this task consisted of a triangle subtending 1.2 × 1.2° and outlined in light gray. It was presented in either an upright or inverted orientation. The distracter consisted of a square of the same size and color. Isolated exemplars from the people, car, and tree categories were presented within the triangle and square (Figure 6A). Sixteen different exemplars were used for each of the three categories.

Procedure & Design—The category detection task was identical to the one used in Experiment 1B. Scene presentation duration was fixed at 70 ms for all participants based on the previous experiments.

We combined this task with the secondary triangle task, for which the cued category was irrelevant, in order to assess the degree to which attention is captured automatically. This task was inserted during the preparation period between the cue and scene of the category detection task. On each trial, a triangle and square were presented on either side of the fixation cross 1500 ms after the cue. Participants were instructed to indicate the orientation of the triangle via button press. The square was never task-relevant and thus served as a distracter. The triangle and square display remained present on the screen until response.

Presented within both the triangle and the square was a member of either the cued, non-cued, or eutral category. An exemplar from one category was always paired with a member from another category, resulting in three different category pairs (cued – neutral, non-cued – neutral, cued – non-cued). For each pair, which category was presented within the triangle and the square varied across trials, resulting in a total of six display types (Figure 6B). Critically, the category exemplars were completely irrelevant to the triangle task. Participants completed 8 runs of 48 trials each. Search task and scene type in the category task as well as display type, triangle orientation, and triangle location in the triangle task were all counterbalanced across the 384 trials of the experiment.

Prior to the start of the experiment, participants completed a practice block consisting of 10 trials of the triangle task, 10 trials of the category detection task, and 25 trials of the combined category detection and triangle task.

Results & Discussion

Accuracy on the category detection task was on average 77.63% [vs. chance: $t(11) = 16.26$, $p < 0.001$, $d = 4.70$], indicating that participants were able to hold the cue in mind over the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

duration of a trial and effectively search for the cued object category, despite the intervening secondary task.

RTs on the triangle task were analyzed in order to assess whether attention was automatically drawn to the object belonging to the category that was relevant for the later search task. If so, we reasoned that participants should be faster to determine the orientation of the target triangle when it contains an exemplar from the cued category, as compared to when the cued category exemplar appears in the square. For trials on which objects from the cued and neutral categories were presented (Figure 7A), RTs were faster when the cued category occupied the triangle vs. the square [$t(11) = -4.28$, $p = 0.001$, $d = -1.24$], suggesting that the cued category captured attention. In contrast, for displays containing objects from the non-cued and neutral categories (Figure 7B), RTs did not differ based on whether the non-cued category occupied the triangle or square [$t(11) = -1.10$, $p = 0.294$, $d = -0.32$], suggesting that, despite having been relevant on previous trials, the non-cued category did not capture attention compared to a never-relevant category. Finally, for displays containing objects from the cued and non-cued categories (Figure 7C), RTs were again faster when the cued category occupied the triangle vs. the square [$t(11) = -2.41$, $p = 0.035$, $d = -0.70$].

Thus, we found evidence that exemplars from the cued category captured attention on a secondary task, relative to both the neutral and non-cued categories. Critically, the objects in this task were completely irrelevant for the triangle judgment, suggesting that attentional capture based on abstract categories can be automatic.

Experiment 2B

Previous research has shown that exchanging an established attentional set for a new one may require suppression of the no longer relevant attentional set (Seidl, Peelen, & Kastner, 2012). However, in Experiment 2A, there was no sign of suppression for the non-cued category relative to the neutral category. One reason might be that the constant switching of the cued and non-cued categories across trials prevented the buildup of inhibition for the non-cued category. In Experiment 2B, we evaluated this possibility by holding the cued category constant for longer periods of time.

Methods

Participants—Fourteen Princeton University undergraduates (age: 18-22, 11 female, normal or corrected-to-normal vision) provided informed consent to a protocol approved by the Princeton University IRB. Data from two participants were excluded because they did not follow task instructions. For purposes of counterbalancing, excluded participants were replaced such that the sample included 12 valid datasets.

Stimuli & Apparatus—The apparatus and stimuli were identical to Experiment 2A.

Procedure & Design—The experimental design was identical to Experiment 2A with the exception that the cued category no longer switched on a trial-by-trial basis but was held constant over an entire block (48 trials) of the experiment. Blocks of the two task-relevant

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

categories alternated back and forth. In order to encourage participants to start preparing for the category detection task at the beginning of every single trial, the triangle task was omitted on five randomly chosen trials per block. Further, to ensure that even on the first trial of the experiment participants had established an attentional set for the non-cued category, the practice phase now contained 48 trials of the pure category detection task. These 48 trials were divided into 4 blocks of 12 trials, during which the cued category was held constant.

Results

Accuracy on the category detection task was 77.91% [vs. chance: $t(11) = 14.82$, $p < 0.001$, $d = 4.29$]. RTs on trials containing objects from the cued and neutral categories differed depending on whether the cued category occupied the triangle vs. the square [$t(11) = -2.30$, $p = 0.042$, $d = -0.66$; Figure 8A], replicating the key result of Experiment 2A. Moreover, as in Experiment 2A, the same effect was observed for trials with objects from cued and non-cued categories [$t(11) = -3.77$, $p = 0.003$, $d = -1.09$; Figure 8C].

The critical test of suppression lies in the trials that contained objects from the non-cued and neutral categories (Figure 8B). If the non-cued attentional set was suppressed, RTs should be faster when the neutral category occupied the triangle vs. square. However, as in Experiment 2A, RTs did not differ depending on the placement of the categories [$t(11) = -0.52$, $p = 0.616$, $d = -0.15$]. Thus, even when there was a greater opportunity and incentive to discard the no longer relevant attentional set for the non-cued category, no such suppression was observed. This suggests that in this experiment attentional capture for target over distracter categories was driven by enhancement for matching objects. At this point we can only speculate as to why we did not find behavioral evidence of distracter suppression. It is possible that the time period over which the target was held constant was still too short and the switches of the target category too frequent to make suppression an adaptive strategy in the current experiment. Future experiments that systematically vary the length of search for the same category will yield more insight to this question.

Experiment 3

The four preceding experiments demonstrated that, during real-world visual search, an endogenously generated attentional set guides attention toward irrelevant exemplars of the task-relevant category – even if there is little visual overlap between the irrelevant exemplars and the target (in this case the category embedded in the scenes). In order to make an even stronger case for capture by visually dissimilar yet potentially relevant stimuli we tested whether attentional capture for a category would not only occur for exemplars of that category, but also for objects that are merely semantically related.

Methods

Participants—Twenty-seven members of the Princeton University community (age: 19-26, 21 female, normal or corrected-to-normal vision) provided informed consent to a protocol approved by the Princeton University IRB. Data from three participants were discarded due to low performance on the category detection task (accuracy < 60% or RT > 2

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

SD, from group mean). For purposes of counterbalancing, excluded participants were replaced such that the sample included 24 valid datasets. The number of participants was increased in this experiment due to the expectation that effect sizes for semantic associates would be lower than those observed for category exemplars.

Stimuli & Apparatus—The apparatus was identical to Experiments 1 and 2. Stimuli in the category detection task were identical to Experiments 2A and 2B. In the triangle task, the target and distracter stimuli were also identical to Experiments 2A and 2B. However, the objects presented within the triangle and square were not direct exemplars of the categories, but rather semantically related objects (Figure 9).

Semantic associates were chosen based on an online survey in which a separate set of participants (N = 40, age: 19-64, 26 female) were asked to name five objects that they associated with people, cars, and trees, but were not parts of the objects themselves. We included in the stimulus set two exemplars of the eight most commonly named objects that: (1) did not overlap with any of the objects associated with the other two categories, (2) were not parts of the actual category (e.g., tires), and (3) were not members of the same superordinate category (e.g., trucks).

To confirm that the semantic associates were indeed visually dissimilar from the category exemplars employed in the previous experiments, we quantified the visual similarity between exemplars and associates using the Gabor-jet model — a model of V1 neurons, simulating the filtering of the visual field at multiple scales and orientations (Biederman & Kalocsai, 1997; Lades et al., 1993). The implementation of the Gabor-jet model was based on Xu & Biederman (2010). The average pairwise similarity values between exemplars and associates from the three object categories are reported in Table 1. For each of the three categories, similarity among exemplars was higher than both similarity among associates and similarity between associates and exemplars.

Procedure & Design—The procedure and design were identical to Experiment 2A, with the exception that semantic associates rather than exemplars of the cued, non-cued, and neutral categories were presented during the triangle task (Figure 9).

Results & Discussion

Accuracy on the category detection task was 80.96% [vs. chance: $t(23) = 19.60$, $p < 0.001$, $d = 4.00$]. RTs on trials containing semantic associates of the cued and neutral categories differed depending on whether the semantic associate of the cued category occupied the triangle vs. the square [$t(23) = -3.06$, $p = 0.006$, $d = -0.62$; Figure 10A], with faster RTs when it occupied the triangle. Thus, semantic associates of a task-relevant category capture attention over objects that are semantically associated with a never-relevant category. In contrast, there were no RT differences on trials containing semantic associates of the non-cued and neutral categories [$t(23) = -1.06$, $p = 0.302$, $d = -0.22$; Figure 10B]. Surprisingly, there was also no difference in RTs on trials containing the cued and non-cued category [$t(23) = 0.95$, $p = 0.351$, $d = 0.19$; Figure 10C].

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The lack of a cued vs. non-cued difference was unexpected, both theoretically and based on the results of Experiments 2A and 2B, and at this point, we can only speculate on why this might have occurred. One possibility is that participants actively attended to objects in the triangle task to identify how they were related to the cued category. Such a strategy might interfere with the automatic capture effect observed in previous experiments.

To summarize, Experiment 3 demonstrated that attention during real-world visual search is automatically guided not only towards exemplars, but also towards semantic associates of the currently relevant object category. The semantic associates had little visual resemblance with the category exemplars used in the previous experiments. These results therefore offer further support for the claim that capture can occur for abstract visual information. Although at least one previous study has demonstrated attentional capture for visually dissimilar objects (Wyble et al., 2013), the capturing distracters in that case were always members of the target category. In the present study, by contrast, attention was automatically guided towards semantic associates even though they were not included in the target definition. This finding suggests that semantic information stored in long-term memory can influence the allocation of attention, in agreement with previous studies documenting interference by semantically associated distracters during visual search (Meyer, Belke, Telling, & Humphreys, 2007; Moores, Laiti, & Chelazzi, 2003). Our results add to these existing findings in two important ways.

First, in previous work (Moores et al., 2003), related distracters were always presented as part of the search display, and it is thus unclear to what extent the observed effects reflect volitional vs. reflexive orienting. In contrast, we showed that related objects capture attention even when presented during a secondary task in which they were completely irrelevant, demonstrating that capture is automatic.

Second, the associates used in previous work were arguably closer in semantic space than the ones used here. For instance, according to a large database of free association norms (Nelson, McEvoy, & Schreiber, 1998), the probability of generating the object pairs used in Moores (2003), when cued with one member of the pair, was on average 0.28 (range: 0 – 0.8). In contrast the probability of generating the associations used in the current study was on average only 0.04 (range: 0 – 0.37). Thus, the present results suggest that automatic activation of associated category representations can occur in a widespread manner during real-world visual search.

General Discussion

The present study aimed to determine whether contingent attentional capture extends beyond simple well-defined visual features (Folk et al., 2002; Folk & Remington, 1998; Folk et al., 1992; Folk et al., 1994) to situations of real-world visual search in which the specific appearance of a target object within a scene is largely unknown in advance. In Experiments 1A and 1B we found that distracter objects belonging to a category relevant for the search task impaired detection performance. This capture effect was limited to trials on which scenes followed the distracters within less than 300 ms — a time window consistent with studies investigating contingent capture for simple visual features (Folk et al., 2002; Lamy,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Leber, & Egeth, 2004; Leblanc & Jolicoeur, 2005). Experiment 1B further showed that impaired detection performance occurred for both central and peripheral distracters, suggesting that it does not necessarily depend on the misdirection of spatial attention. In Experiments 2A and 2B we demonstrated that capture by exemplars from the search target category is automatic, as it occurred on a secondary task when the object categories were completely task-irrelevant. Using a neutral third category, we were further able to deduce that these capture effects reflect enhancement of the cued category rather than suppression of the non-cued category. Finally, Experiment 3 showed that automatic capture does not only occur for members of the cued object category but also for objects that are semantically related.

The present results suggest that mechanisms of attentional capture (Folk et al., 2002; Folk & Remington, 1998; Folk et al., 1992; Folk et al., 1994) apply when target-defining features are not precisely known and hence an abstract attentional set is required. This is consistent with a recent study that reported contingent attentional capture during search for conceptually defined targets (Wyble et al., 2013). We extend these findings by showing that capture for complex visual information occurs in an automatic fashion, making it a useful mechanism for finding objects in the real world where behaviorally relevant information often appears in unexpected locations or unexpected points in time. Although the majority of the reported experiments investigated capture of spatial attention, Experiment 1A showed that capture can also occur for non-spatial attention. An interesting avenue for future research will be to compare capture of spatial and non-spatial attention during real-world visual search. These two forms of “capture” may rely on distinct mechanisms, with processing delays in spatial attention resulting from a need to reorient and in non-spatial attention from a central bottleneck, as in the AB (Chun & Potter, 1995). Regardless of the mechanisms, we assume that in both cases the origin of the effect is the match between an abstract attentional set and the distracter item. At the neural level this attentional set may be implemented through the pre-activation of category-selective neurons in object-selective cortex while preparing for real-world visual search (Peelen & Kastner, 2011).

We have interpreted the results of all experiments presented here as reflecting attentional capture that occurs when a distracter matches the abstract attentional set required for real-world visual search. However, there are some differences between the paradigm used in Experiment 1 and that used in Experiments 2 and 3 that are worth noting. In particular, in Experiments 1A and 1B, the delay period between the cue and scene onset was variable and distracters could appear at potentially relevant timepoints. Owing to the unpredictable timing of events, it is likely that participants immediately implemented the relevant attentional set following the cue stimulus. In contrast, in Experiments 2 and 3, the timing of events was fixed and for the most part the triangle task occurred on every trial (this was not true in Experiment 2B, which included a small percentage of catch trials, on which no triangle task was presented). Thus, it is possible that rather than immediately implementing the relevant attentional set, participants merely held the category cue in WM until completion of the triangle task, and only afterwards activated the attentional set for the search task. Previous work has shown that holding a verbal cue in WM can result in attentional capture by visual information (Soto & Humphreys, 2007), implying the existence of an automatic link between verbal WM and visual representations. Thus, one might argue

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

that the capture effects observed here are driven by WM representations rather than search-related attentional sets.

To understand whether this distinction is meaningful, a more complete account of the interactions between working memory and attention is needed. Early studies failed to show an interaction between working memory load and search performance, suggesting that target-related attentional sets do not rely on the same representations as items held in WM (Logan, 1978, 1979; Woodman, Vogel, & Luck, 2001). However, later work found interference between object WM and visual search when the need for maintaining an active target-related attentional set was increased due to frequent changes of the search target (Woodman, Luck, & Schall, 2007). Although it is unclear whether WM representations of the cues and search-related attentional sets rely on common resources, they do differ in terms of timing and processing steps with respect to how they could support the current task: the attentional set can be immediately adopted, whereas WM for the cue would allow the attentional set to be adopted strategically at a later point. Such differences may provide a way forward for future research on this issue.

Our results further show that objects that are merely semantically associated with the target category also automatically attract attention. This finding makes a strong case that attentional capture occurs for abstract information, since the capture stimuli were visually dissimilar from the target category exemplars. Attending to objects that are semantically associated with a target during visual search may be useful in that associated objects can be predictive of the location of the search target itself (Biederman, 1972). Recent work has shown that categories that are close in semantic space are also represented in close spatial proximity in the brain, creating a continuous representation of a semantic object space across cortex (Huth, Nishimoto, Vu, & Gallant, 2012). Thus, attentional capture for semantic associates may be driven by a spread of activation to close by object representations during preparation for visual search.

While the current result converges with other studies showing attentional guidance by semantically associated objects (Moores et al., 2003), other studies have found this kind of guidance to be dependent on visual similarity. In particular, Calleja and Rich (2013) found WM-guided capture for distracters that belonged to the same category as an item stored in WM, but only as long as the WM item and the same-category distracter were visually similar. We posit that whether capture extends to semantic associates depends on the specificity of the attentional set that is encouraged by task demands. In Calleja and Rich (2013), for instance, the WM item was always the image of a specific object. Even though participants performed a match-to-category task, the presentation of a specific image as the WM item may have interfered with the implementation of an abstract attentional set.

More broadly, the finding of attentional capture by objects that are associated in semantic long-term memory adds to a growing literature showing that a wide range of memory processes can guide the allocation of attention (Hutchinson & Turk-Browne, 2012). Different forms of memory may differ in the extent to which the control they exert over attention is volitional or automatic. Thus far, automatic capture has been demonstrated for incidentally learned statistical regularities (Zhao, Al-Aidroos, & Turk-Browne, 2013) and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

for items that are stored in working memory (Soto et al., 2005). Our results additionally show that attentional guidance by semantic memory can be automatic. Further work will be needed to determine whether the influence of episodic memories on the allocation of attention can occur in a similarly reflexive manner.

Finally, the paradigms developed here may prove useful for addressing other questions about the mechanisms involved in real-world visual search. For instance, one recent study used a paradigm similar to Experiments 2 and 3 to examine what people look for when they are instructed to detect object categories in a scene, that is, the contents of the attentional set used during real-world visual search (Reeder & Peelen, 2013). The study revealed attentional capture for silhouettes of objects, irrespective of their orientation and location in space, as well as for silhouettes of object parts. These results suggest that the attentional set is comprised of location- and view-invariant contours of representative object parts. Although it was not our goal to map the contents of attentional sets during real-world search, the results from Experiment 3 suggest that the attentional set for a category is not restricted to the diagnostic visual features of a prototype, but also includes semantic features shared with dissimilar looking objects and categories.

Conclusions

Across five experiments, we found reliable evidence that during visual search for categories, attention is automatically and efficiently drawn toward target category exemplars and semantic associates. The reflexive and selective nature of this orienting may constitute an adaptive mechanism for real-world visual search, allowing for the detection of useful information even when it appears at unexpected locations or points in time.

Acknowledgments

This work was supported by:

NSF BCS 1025149 (SK)

NSF BCS 1328270 (SK)

NIH R01 MH64043 (SK)

NIH R01 EY021755 (NTB)

References

Bahrami Balani AB, Soto D, Humphreys GW. Working memory and target-related distractor effects on visual search. Memory & Cognition. 2010; 38(8):1058–1076. doi: Doi 10.3758/Mc.38.8.1058. [PubMed: 21156870]

Biederman I. Perceiving real-world scenes. Science. 1972; 177(4043):77–80. [PubMed: 5041781]

Biederman I, Kalocsai P. Neurocomputational bases of object and face recognition. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences. 1997; 352(1358):1203–1219.

Brainard DH. The Psychophysics Toolbox. Spat Vis. 1997; 10(4):433–436. [PubMed: 9176952]

Bundesen C. A theory of visual attention. Psychological Review. 1990; 97(4):523–547. [PubMed: 2247540]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Calleja MO, Rich AN. Guidance of attention by information held in working memory. Atten Percept Psychophys. 2013; 75(4):687–699. doi: 10.3758/s13414-013-0428-y. [PubMed: 23404521]

Chun MM, Potter MC. A two-stage model for multiple target detection in rapid serial visual presentation. J Exp Psychol Hum Percept Perform. 1995; 21(1):109–127. [PubMed: 7707027]

Dehaene S, Sergent C, Changeux JP. A neuronal network model linking subjective reports and objective physiological data during conscious perception. Proc Natl Acad Sci U S A. 2003; 100(14): 8520–8525. doi: 10.1073/pnas.1332574100. [PubMed: 12829797]

Duncan J, Humphreys GW. Visual-search and stimulus similarity. Psychological Review. 1989; 96(3): 433–458. doi: Doi 10.1037//0033-295x.96.3.433. [PubMed: 2756067]

Dux PE, Marois R. The attentional blink: A review of data and theory. Attention Perception & Psychophysics. 2009; 71(8):1683–1700. doi: Doi 10.3758/App.71.8.1683.

Einhauser W, Koch C, Makeig S. The duration of the attentional blink in natural scenes depends on stimulus category. Vision Research. 2007; 47(5):597–607. doi: 10.1016/j.visres.2006.12.007. [PubMed: 17275058]

Folk CL, Leber AB, Egeth HE. Made you blink! Contingent attentional capture produces a spatial blink. Percept Psychophys. 2002; 64(5):741–753. [PubMed: 12201333]

Folk CL, Leber AB, Egeth HE. Top-down control settings and the attentional blink: Evidence for nonspatial contingent capture. Visual Cognition. 2008; 16(5):616–642. doi: Doi 10.1080/13506280601134018.

Folk CL, Remington R. Selectivity in distraction by irrelevant featural singletons: evidence for two forms of attentional capture. J Exp Psychol Hum Percept Perform. 1998; 24(3):847–858. [PubMed: 9627420]

Folk CL, Remington RW, Johnston JC. Involuntary covert orienting is contingent on attentional control settings. J Exp Psychol Hum Percept Perform. 1992; 18(4):1030–1044. [PubMed: 1431742]

Folk CL, Remington RW, Wright JH. The structure of attentional control: contingent attentional capture by apparent motion, abrupt onset, and color. J Exp Psychol Hum Percept Perform. 1994; 20(2):317–329. [PubMed: 8189195]

Hutchinson JB, Turk-Browne NB. Memory-guided attention: control from multiple memory systems. Trends Cogn Sci. 2012; 16(12):576–579. doi: 10.1016/j.tics.2012.10.003. [PubMed: 23141429]

Huth AG, Nishimoto S, Vu AT, Gallant JL. A continuous semantic space describes the representation of thousands of object and action categories across the human brain. Neuron. 2012; 76(6):1210–1224. doi: 10.1016/j.neuron.2012.10.014. [PubMed: 23259955]

Jolicoeur P, Dell'Acqua R. Attentional and structural constraints on visual encoding. Psychological Research. 1999; 62(2-3):154–164. doi: Doi 10.1007/S004260050048.

Lades M, Vorbruggen JC, Buhmann J, Lange J, Vandermalsburg C, Wurtz RP, Konen W. Distortion invariant object recognition in the dynamic link architecture. Ieee Transactions on Computers. 1993; 42(3):300–311. doi: Doi 10.1109/12.210173.

Lamy D, Leber A, Egeth HE. Effects of task relevance and stimulus-driven salience in feature-search mode. J Exp Psychol Hum Percept Perform. 2004; 30(6):1019–1031. doi: Doi 10.1037/0096-1523.30.6.1019. [PubMed: 15584812]

Leblanc E, Jolicoeur P. The time course of the contingent spatial blink. Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale. 2005; 59(2):124–131. doi: Doi 10.1037/H0087467. [PubMed: 16035345]

Logan GD. Attention in character-classification tasks - evidence for automaticity of component stages. Journal of Experimental Psychology-General. 1978; 107(1):32–63. doi: Doi 10.1037//0096-3445.107.1.32.

Logan GD. Use of a concurrent memory load to measure attention and automaticity. Journal of Experimental Psychology-Human Perception and Performance. 1979; 5(2):189–207.

Martens S, Wyble B. The attentional blink: past, present, and future of a blind spot in perceptual awareness. Neurosci Biobehav Rev. 2010; 34(6):947–957. doi: 10.1016/j.neubiorev.2009.12.005. [PubMed: 20025902]

Meyer AS, Belke E, Telling AL, Humphreys GW. Early activation of object names in visual search. Psychon Bull Rev. 2007; 14(4):710–716. doi: Doi 10.3758/Bf03196826. [PubMed: 17972738]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Moores E, Laiti L, Chelazzi L. Associative knowledge controls deployment of visual selective attention. Nat Neurosci. 2003; 6(2):182–189. doi: 10.1038/nn996. [PubMed: 12514738]

Nelson, D.; McEvoy, C.; Schreiber, T. The University of South Florida word association, rhyme, and word fragment norms. 1998. http://www.usf.edu/FreeAssociation/

Peelen MV, Kastner S. A neural basis for real-world visual search in human occipitotemporal cortex. Proc Natl Acad Sci U S A. 2011; 108(29):12125–12130. doi: 10.1073/pnas.1101042108. [PubMed: 21730192]

Pelli DG. The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis. 1997; 10(4):437–442. [PubMed: 9176953]

Reeder RR, Peelen MV. The contents of the search template for category-level search in natural scenes. J Vis. 2013; 13(3):13. doi: 10.1167/13.3.13. [PubMed: 23750015]

Russell SB, Torralba RM, Murphy KP, Freeman WT. LabelMe: a database and web-based tool for image annotation. International Journal of Computational Vision. 2008; 77:157–173.

Schmidt J, Zelinsky GJ. Search guidance is proportional to the categorical specificity of a target cue. Quarterly Journal of Experimental Psychology. 2009; 62(10):1904–1914. doi: Pii 911471427 Doi 10.1080/17470210902853530.

Seidl KN, Peelen MV, Kastner S. Neural evidence for distracter suppression during visual search in real-world scenes. J Neurosci. 2012; 32(34):11812–11819. doi: 10.1523/JNEUROSCI.1693-12.2012. [PubMed: 22915122]

Soto D, Heinke D, Humphreys GW, Blanco MJ. Early, involuntary top-down guidance of attention from working memory. J Exp Psychol Hum Percept Perform. 2005; 31(2):248–261. doi: 10.1037/0096-1523.31.2.248. [PubMed: 15826228]

Soto D, Hodsoll J, Rotshtein P, Humphreys GW. Automatic guidance of attention from working memory. Trends Cogn Sci. 2008; 12(9):342–348. doi: 10.1016/j.tics.2008.05.007. [PubMed: 18693131]

Soto D, Humphreys GW. Automatic guidance of visual attention from verbal working memory. J Exp Psychol Hum Percept Perform. 2007; 33(3):730–737. doi: 10.1037/0096-1523.33.3.730. [PubMed: 17563233]

Walther DB, Caddigan E, Fei-Fei L, Beck DM. Natural scene categories revealed in distributed patterns of activity in the human brain. J Neurosci. 2009; 29(34):10573–10581. doi: 10.1523/JNEUROSCI.0559-09.2009. [PubMed: 19710310]

Wolfe JM, Horowitz TS, Kenner N, Hyle M, Vasan N. How fast can you change your mind? The speed of top-down guidance in visual search. Vision Research. 2004; 44(12):1411–1426. doi: Doi 10.1016/J.Visres.2003.11.024. [PubMed: 15066400]

Woodman GF, Luck SJ, Schall JD. The role of working memory representations in the control of attention. Cereb Cortex. 2007; 17(Suppl 1):i118–124. doi: 10.1093/cercor/bhm065. [PubMed: 17725994]

Woodman GF, Vogel EK, Luck SJ. Visual search remains efficient when visual working memory is full. Psychol Sci. 2001; 12(3):219–224. doi: Doi 10.1111/1467-9280.00339. [PubMed: 11437304]

Wyble B, Folk C, Potter MC. Contingent attentional capture by conceptually relevant images. J Exp Psychol Hum Percept Perform. 2013; 39(3):861–871. doi: 10.1037/a0030517. [PubMed: 23163786]

Xu X, Biederman I. Loci of the release from fMRI adaptation for changes in facial expression, identity, and viewpoint. J Vis. 2010; 10(14) doi: 10.1167/10.14.36.

Zhao J, Al-Aidroos N, Turk-Browne NB. Attention is spontaneously biased toward regularities. Psychol Sci. 2013; 24(5):667–677. doi: 10.1177/0956797612460407. [PubMed: 23558552]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

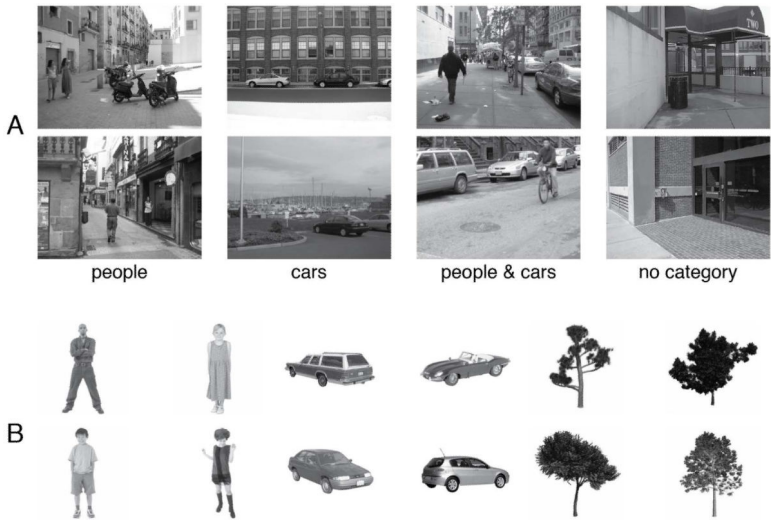

Figure 1.

Example stimuli. **A**, Scene stimuli. In Exp. 1A scenes contained either people, cars, people and cars or neither people nor cars. In Exps. 1B – 3, the same set of scene images was re-sorted for each group of particitpants to contain either the cued, non-cued, cued and non-cued or neither the cued nor the non-cued category. **B**, isolated exemplars from the people, cars and tree category served as distracters in Exps. 1A and 1B. In Exps. 2A and 2B the same stimuli were presented during the triangle task.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

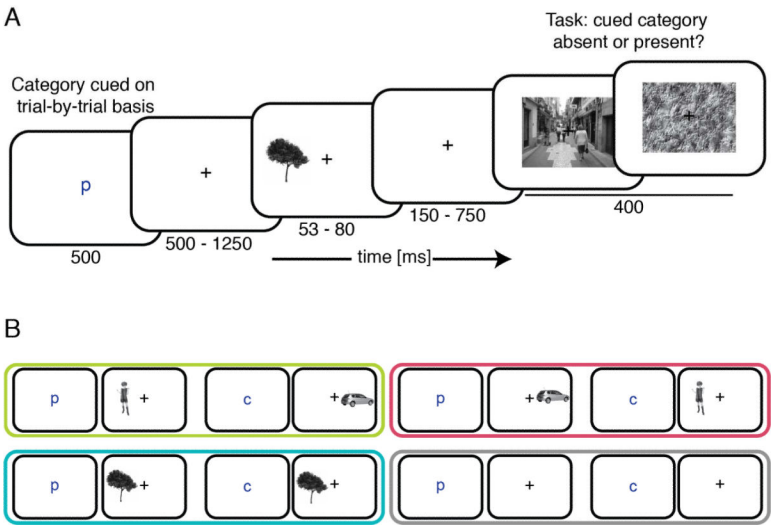

Figure 2. Experimental design for Exp 1a

A, trial structure. On each trial participants were presented with a letter cue, indicating whether people ("p") or cars ("c") were the relevant category for that trial. On 75% of all trials, an irrelevant distracter was presented to the left or right of the fixation cross with varying scene - distracter ISIs. Upon presentation of the scene, participants indicated via button press whether the cued object category was present in the scene or not. **B**, distracter conditions. Distracters could either match the currently relevant category (green, *cued*), match the previously but not currently relevant category (red, *non-cued*), or match a never task-relevant category (blue, *neutral*). On 25% of all trials, no distracter was presented (grey, *no distracter*).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

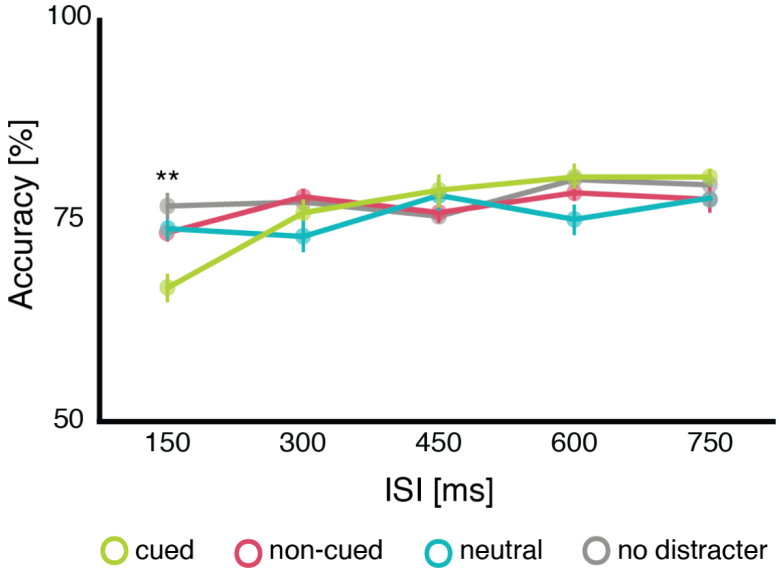

Figure 3.

Results for Experiment 1A. Mean category detection accuracy as a function of distracter type and scene – distracter ISI. Distracters that matched the cued object category selectively reduced category detection accuracy at the shortest ISI. Error bars denote ± standard errors of the mean (*SEM*), corrected for within-subject comparisons. ** $p < 0.01$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

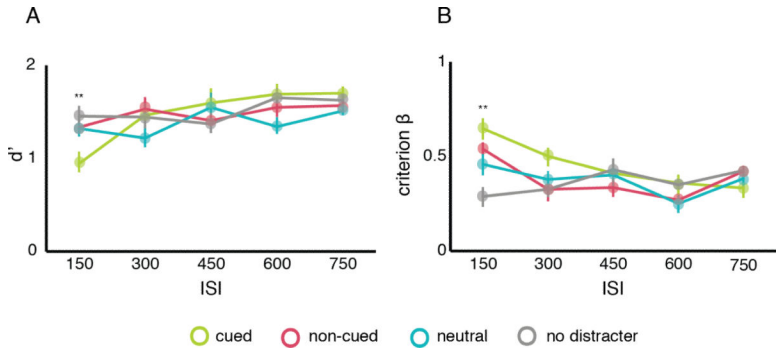

Figure 4.

Signal detection theory measures for Experiment 1A. **A.** *D'* as a function of ISI and distracter type. **B.**Response criterion β as a function of ISI and distracter type. Error bars denote ± *SEM*, corrected for within-subject comparisons. ** $p < 0.01$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

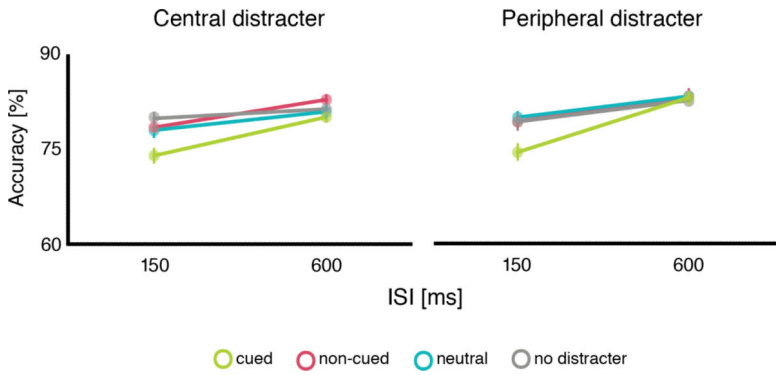

Figure 5.

Results for Experiment 1B. Mean category detection accuracy as a function of distracter type, distracter location and scene-distracter ISI. Category detection accuracy was selectively impaired for distracters that matched the cued object category at short ISIs, independent of the distracter location. Error bars denote ± *SEM*., corrected for within-subject comparisons.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

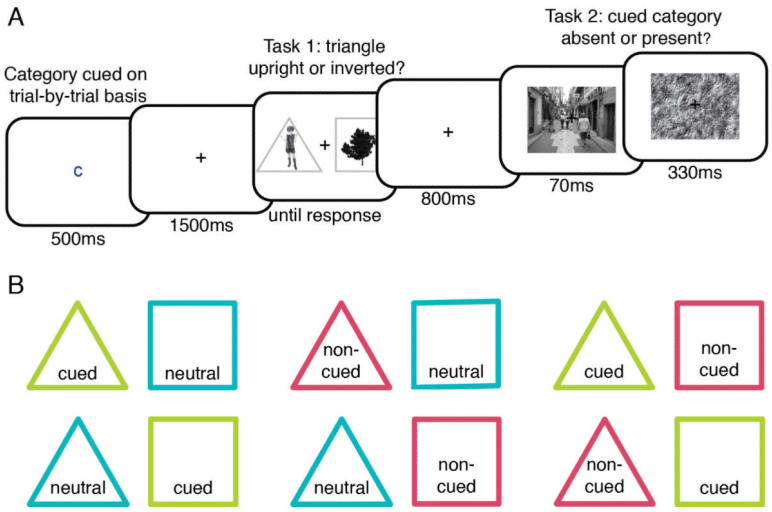

Figure 6.

Experimental design for Experiment 2A. **A**, trial structure. A cue at the beginning of each trial indicated which object category to detect in the upcoming natural scene photograph. The degree to which attention is automatically allocated to one object category over another was measured on the "triangle task" which was inserted in between the cue and the scene. **B**, triangle task display types. Members from the cued, non-cued and neutral object categories were presented within the triangle and square shapes. An object from one category was always paired with an object from a second category. Within each category pair, the category to shape assignment randomly switched across trials.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

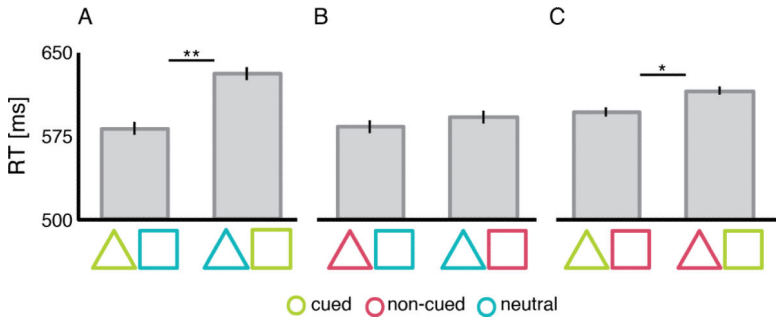

Figure 7.

Results for Experiment 2A. Mean RTs on the triangle task. For analysis, we compared RTs on displays that contained the same object categories but with the opposing category to shape assignment. **A**, displays containing the cued and the neutral categories revealed capture for the cued over the neutral category. **B**, RTs for displays containing the non-cued and neutral categories did not depend on the category to shape assignment. **C**, displays containing the cued and the non-cued categories revealed capture for the cued over the non-cued category. Error bars denote ± *SEM*, corrected for within-subject comparisons. * $p < 0.05$, ** $p < 0.01$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

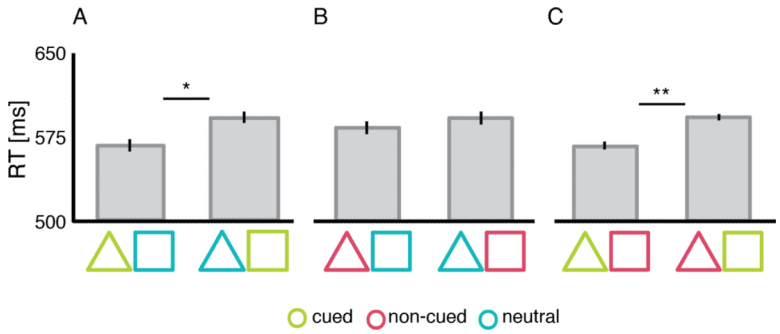

Figure 8.

Results for Experiment 2B. Mean RTs on the triangle task. **A**, displays containing the cued and the neutral categories revealed capture for the cued over the neutral category. **B**, RTs for displays containing the non-cued and neutral categories did not depend on the category to shape assignment. **C**, displays containing the cued and the non-cued categories revealed capture for the cued over the non-cued category. Error bars denote ± *SEM*., corrected for within-subject comparisons. * $p < 0.05$, ** $p < 0.01$

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Figure 9.

Stimuli in triangle task for Experiment 3. **A**, semantic associates of tree category. **B**, semantic associates of people category **C**, semantic associates of car category.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

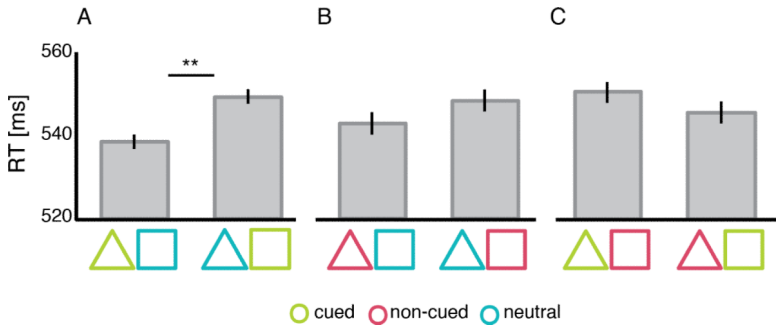

Figure 10.

Results for Experiment 3. Mean RTs on the triangle task. For analysis, we compared RTs on displays that contained the associates from the same object categories but with the opposing category to shape assignment. **A**, displays containing the semantic associates of the cued and the neutral categories revealed capture for the cued over the neutral category. **B**, RTs for displays containing the associates of the non-cued and neutral categories did not depend on the category to shape assignment. **C**, there was no evidence for capture by associates of the cued category over associates of the non-cued category. Error bars denote ± *SEM*, corrected for within-subject comparisons. * $p < 0.05$, ** $p < 0.01$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Average pairwise Gabor-jet similarity values for category exemplars (Experiments 1-2) and category associates (Experiment 3).

Category	Exemplar-Exemplar	Associate-Associate	Exemplar-Associate
Trees	0.55 (0.08)	0.34 (0.12)	0.40 (0.10)
People	0.57 (0.06)	0.23 (0.12)	0.29 (0.12)
Cars	0.56 (0.07)	0.30 (0.12)	0.26 (0.12)

Standard deviations are indicated in parentheses.