Templates for rejection can specify semantic properties of nontargets in natural scenes

Jennifer L. Daffron — Department of Psychology, University of Cambridge, Cambridge, UK

Greg Davis — Department of Psychology, University of Cambridge, Cambridge, UK

In contrast to standard search templates that specify a target object's expected features, templates for rejection (TFR) may specify features of nontargets, biasing attention away from irrelevant objects. Little is known about TFR, and virtually nothing is known about their role in guiding search across natural scenes. In such scenes, targets and nontargets may not be easily distinguished on the basis of their visual features; it has been claimed that standard search templates may therefore specify target objects' semantic features to guide attention. Here, we ask whether TFR can do so. Noting a limitation of previous procedures used to study standard search templates, we trialed an alternative method to examine semantic templates for nontarget exclusion in natural scene search. We found that when nontargets belonged unpredictably to either of two physically distinct categories, search was less efficient than when targets belonged to one known category. This two-category cost, attributed to inefficient application of search templates, was absent for two physically dissimilar but semantically related categories. Adding a training phase to highlight semantic distinctiveness of two object categories reinstated the two-category cost, precluding stimulus-based accounts of the effect. These patterns were not observed for one-image displays or when observers searched for object categories rather than ignoring them, demonstrating their specificity to TFR, the inadequacy of search-and-destroy models to account for them, and likely basis in attentional guidance. TFR can specify semantic information to guide attention away from nontargets.

Introduction

The human visual system presumably is optimized for processing the complex combinations of objects, colors, surfaces, and textures that typify natural scenes (Burge, Fowlkes, & Banks, 2010; Sigman, Cecchi, Gilbert, & Magnasco, 2001; Tkacik, Prentice, Victor, & Balasubramanian, 2010). Efficient search of such complex stimuli requires flexible, top-down representations that specify which stimuli are relevant and which are not (Arita, Carlisle, & Woodman, 2012; Bundesen, 1990; Cohen, Konkle, Rhee, Nakayama, & Alvarez, 2014; Koshino, 2001; Telling, Kumar, Meyer, & Humphreys, 2010; Wolfe, 1994). These top-down representations, or search templates, must often specify a target object's features, influencing search in at least two ways. First, when a searched-for target object's position is unknown, a template may specify the target's properties, biasing attention shifts toward objects that share one or more of the target's properties (Duncan & Humphreys, 1989; Malcolm & Henderson, 2009; Reeder & van Peelen, 2013; Wolfe, Cave, & Franzel, 1989; Wyble, Folk, & Potter, 2013). Second, a template may establish response criteria, releasing a response only when the match between a stimulus and the template reaches threshold.

Search templates may also operate in an alternative mode. When the target's properties are not known, search may proceed by rejecting stimuli that are known to be irrelevant nontargets. This mode of search has received much less attention than search for known categories of targets. However, surely it must play a key role in behavioral inhibition—inhibiting attention to, and responses to, irrelevant stimuli that are perceptually salient or that elicit affective responses (substances of addiction, phobic stimuli). It is our primary focus here. As opposed to target templates that specify target features, the types of templates active in designating a stimulus as not relevant, or as a nontarget, have been termed *nontarget templates* (Müller, Humphreys, & Donnelly, 1994) or templates for rejection (TFR; Woodman & Luck, 2007). TFR may operate in concert with target templates during many searches but likely influence behavior, particularly when target features

Citation: Daffron, J. L., & Davis, G. (2015). Templates for rejection can specify semantic properties of nontargets in natural scenes. *Journal of Vision, 15*(15):16, 1–20, doi:10.1167/15.15.16.

doi: 10.1167/15.15.16 Received May 8, 2015; published November 25, 2015 ISSN 1534-7362 © 2015 ARVO

are unknown (e.g., Müller et al., 1994; Woodman & Luck, 2007).

What information do search templates specify?

Models of attention, such as attentional engagement theory, Bundesen's theory of visual attention (TVA), Wolfe's guided search, and contingent attentional capture (Bundesen, Habekost, & Kyllingsbaek, 2005; Dent, Allen, Braithwaite, & Humphreys, 2012; Seidl-Rathkopf, Turk-Browne, & Kastner, 2015; Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011; Woodman & Luck, 2007; Wyble et al., 2013), specify how search templates (both target templates and TFR) may guide search. Much of the work that has shaped these views of attention has used simple, abstract, two-dimensional arrays of objects (Dent et al., 2012; Seidl-Rathkopf et al., 2015). Perhaps as a reflection of this choice, many previous studies have stressed the importance and influence of visual features specified by templates (Moores, Laiti, & Chelazzi, 2003; Vickery, King, & Jiang, 2005) rather than semantic features.

The effects of templates on performance typically are measured by cuing observers to particular target features before a given search display, informing them about one or more of the target's properties or, in the case of TFR, the nontargets' properties. For target templates, when cues accurately specify the target's visual properties such as its precise retinal image (Reeder & van Peelen, 2013; Spotorno, Malcolm, & Tatler, 2015; Wolfe et al., 2011; Yang & Zelinsky, 2009), other views of the same object (Reeder & van Peelen, 2013), or a verbal description (e.g., "red car"), observers' search typically becomes more efficient (Reeder & van Peelen, 2013; Wolfe et al., 2011). That is, the more precise the information about the target provided by the cue, the more effective the cue is in visual guidance to the target (Moores et al., 2003; Reeder & van Peelen, 2013; Schmidt & Zelinsky, 2009; Seidl-Rathkopf et al., 2015; Vickery et al., 2005). Presumably, these cues permit observers to establish target templates that match incoming target information more precisely, increasing search efficiency. Rejection of nontargets in such tasks seems to operate along similar lines (Woodman & Luck, 2007).

These findings are consistent with a common view that search templates' efficacy is determined by their specification of target- or nontarget-specific visual properties. Possible exceptions to this pattern involve overlearned, limited categories of stimuli, such as faces or alphanumeric characters (Cohen, 2009; García-Orza, Perea, Abu Mallouh, & Carreiras, 2012; Godwin, Hout, & Menneer, 2014; Lupyan, 2008; Schwarz & Eiselt, 2012), for which there is limited evidence of attentional guidance on the basis of category-level or semantic properties. However, in search of naturalistic objects against natural scene backgrounds, the same emphasis on visual features might not be expected.

Templates for natural scene search

In contrast to the simple, abstract shapes used in conventional visual search tasks, natural scene search tasks often involve increasingly complex targets hidden among diverse nontargets that may have many features in common (Bravo & Farid, 2009; Castelhano, Pollatsek, & Cave, 2008; Seidl-Rathkopf et al., 2015). Accordingly, the types of rapid, bottom-up parsing processes that yield efficient pop-out search in laboratory-based tasks would not be expected to function as optimally in natural scene search due to nontarget variance and absence of target-unique features (Dent et al., 2012; Duncan, 1983; Duncan & Humphreys, 1989; Friedman-Hill & Wolfe, 1995; Kim & Cave, 1999; Menneer, Cave, & Donnelly, 2009). Furthermore, real-world tasks often require search for broad categories (e.g., food or shelter), the members of which may be so diverse in terms of their retinal image features and the target–nontarget feature overlap so great, that specifying retinal-image information would not support effective search. Under some circumstances, therefore, the efficacy of search templates may be limited primarily not on the basis of the visual features they specify but rather the semantic properties.

Despite their physical complexity, the processing of natural scenes' primary objects and semantic gist is very efficient (Delorme, Rousselet, Macé, & Fabre-Thorpe, 2004; Kaiser, Stein, & Peelen, 2014; Mack & Eckstein, 2011; Potter, 1976; Rousselet, Macé, & Fabre-Thorpe, 2003; Spotorno et al., 2015; Walker, Stafford, & Davis, 2008). While these processes may exploit differences in spatial-frequency spectra across orientations, such low-level processes do not always seem to capture human visual search abilities (Wichmann & Gegenfurtner, 2010), raising the possibility that top-down templates may sometimes operate more efficiently by specifying semantic or object category representations directly rather than specifying visual features that are typical of particular object categories (Wu, Wick, & Pomplun, 2014).

Perhaps the most direct evidence to date consistent with semantic templates in real-world search has been provided by priming and cueing studies of search (Belke, Humphreys, Watson, Meyer, & Telling, 2008; Çukur, Nishimoto, Huth, & Gallant, 2013; Hwang, Wang, & Pomplun, 2011; Moores et al., 2003). These studies reported incidental attention to nontarget objects that were semantically related to verbally specified targets or primes. Such findings suggest that the biasing effect of templates can extend to semantically related nontarget stimuli and that semantic—rather than visual—characteristics may limit which objects a template guides attention to.

Interpretation of these results is complicated, however, by a limitation of this paradigm. Nontargets that are semantically related to the prespecified target category are overrepresented among the nontargets in the experiments. Observers will likely detect this and begin to voluntarily bias their attention toward such nontargets. Indeed, one recent study (Reeder, van Zoest, & van Peelen, 2015) found that under such conditions, attention spreads in an obligatory fashion only to target-category objects and not to semantically related nontargets. This suggests that attentional bias toward semantically related items, when it does occur, is voluntary and an exploratory extension of the template, not a characteristic of the search template itself.

Using two-category costs to reveal template structure

To circumvent this limitation of previous procedures, we adopted a different approach to study TFR in search across photorealistic objects against natural backgrounds. The logic of our procedure was based on recent findings that search is less efficient when two or more distinct categories of targets may be present (a dual-target cost; Cunningham & Wolfe, 2014; Kristjánsson, Jóhannesson, & Thornton, 2014; Menneer et al., 2009). Presumably, this inefficiency reflects difficulties either (a) in applying two visual feature templates simultaneously or (b) in applying a single template effectively to two types of objects simultaneously. In either case, the costs of needing to apply a template or templates to two types of targets simultaneously seem to offer a method for elucidating their structure.

While the above studies examined standard target templates specifying visual features, the same logic can be applied to TFR and to the roles of semantic properties in TFR. Extending this approach required a further assumption: that if two types of target or nontarget objects had overlapping features (e.g., common color) that distinguished them usefully from other objects, the observer could use a template that specified this overlap between the two categories and search could be as efficient as when searching for one category only. We expected this assumption to be uncontroversial. Given this principle, we examined costs of ignoring two categories of visually dissimilar objects versus one category, manipulating the amount of semantic overlap between the two categories. As there was no obvious overlap in visual features between the two categories observers were asked to ignore, we expected the following effects of semantic overlap in our experiments.

First, we expected that if two semantically unrelated categories of objects with little overlap of physical features must be rejected simultaneously, then search should be less efficient than when only one such category must be rejected. This basic finding would support our fundamental assumption that for TFR, rejecting either of two categories of objects from search is less efficient than rejecting just one, on the basis of the lack of overlap between the two categories in terms of visual features or in terms of semantic features.

Second, if the two categories of objects to be ignored, though physically very different, were to overlap in terms of semantic features, two opposing predictions were possible. If, as is commonly assumed, the primary limitations on applying templates (including TFR) reflect the efficiency with which templates specify visual features, the same pattern of costs should emerge as for two unrelated categories. In contrast, if, as we supposed, TFR in natural search are sometimes limited by their ability to specify semantic information, a single TFR should be able to specify the semantic overlap between the two categories of nontargets and reject both categories as efficiently as one category alone. That is, there should now be no performance cost; rejecting two categories of nontargets should be as efficient as rejecting one.

Our experiments here used a pared-down search task similar to that used by other recent work (e.g., Reeder & van Peelen, 2013; Reeder, van Zoest, & van Peelen 2015). Each trial presented two scenes, one of which would always contain a category of objects designated as nontargets. The nature of the target was undefined except in that it was not one of the specified nontargets. This was intended to prohibit the creation of specific search templates for target properties and to encourage search on the basis of rejecting nontarget scenes. Further, by ensuring that the information given to observers about nontargets was category level and that the examples were diverse in their appearances, we hoped to encourage observers to use broad, semantic-level templates (that typify many real-world search tasks) rather than templates based on information available in the visual features.

The task was to move a cursor over the target image using a mouse and to press the mouse button as quickly and accurately as possible. We sought to compare the efficiency with which observers could reject two categories of objects simultaneously versus one category. As only one nontarget scene was presented on any trial (comprising only one category of objects), the stimuli were the same in these two conditions. Only the observer's foreknowledge of which nontarget objects would be presented differed between the two conditions. For two separate groups of observers, we examined search in these conditions for two unrelated categories of objects (Experiment 1A; clocks and keys) and two semantically related objects (Experiment 1B; locks and keys).

Experiments 1A and 1B

Each of the first two experiments (Experiments 1A and 1B) presented two images simultaneously on each trial. These images were presented in paired combinations of locations above, below, left of, and right of a central fixation cross. Speeded response times to move a cursor over one of the images (the target) and to press the mouse button were recorded. The nature of the target image was unspecified other than that observers knew it would not contain one of two categories of objects. Each nontarget image contained one of the specified categories of objects and could be rejected as a possible target on that basis. The basic performance measure in each experiment compared performance in two types of blocks of trials: (a) one-category (OC) blocks, in which observers knew beforehand that all the nontargets in the block would comprise one particular category of objects, and (b) two-categories (TC) blocks, in which nontargets unpredictably contained either of two possible categories of objects. The cost of inhibiting two categories of objects on each trial was calculated by subtracting mean response times for OC blocks from those for TC blocks (mean TC – mean OC). If the TFR used by observers were semantic or category level in nature, we expected the cost to be significant when the two categories were semantically unrelated, as in Experiment 1A, while we expected the cost to be relatively reduced or absent when the two categories of objects were semantically related, as in Experiment 1B.

Observers

A total of 40 observers (Experiment 1A: eight males, 12 females, aged 18–30 years; Experiment 1B: 12 males, eight females, aged 22–33 years) from the University of Cambridge and the local area gave written informed consent and were paid for participating.

Materials

Observers were asked to complete the Fagerstrom Test for Nicotine Dependence, the Alcohol Use Identification Test, and the UPPS-P Impulsivity Questionnaire. These measures were taken because the current experiments are part of a larger project on behavioral response inhibition in addiction, results of which are not analyzed or reported here. The experiment was run on an Apple MacBook Pro laptop (Apple, Cupertino, CA) with a 15-in. liquid crystal display screen at full brightness and an attached Apple mouse. The experiment was programmed in Psyscope XB57 (Cohen et al., 1993). The observers for Experiment 1A (semantically unrelated nontarget categories) viewed 288 randomly selected neutral images, 144 images of keys, and 144 images of clocks. The observers for Experiment 1B (semantically related nontarget categories) viewed 288 randomly selected neutral images, 144 images of keys, and 144 images of locks. The viewing distance was around 50 cm. Images were presented in boxes (150 × 150 pixels; 35 × 35 mm) centered at 32.5 mm above, left of, below, and right of fixation.

Procedure

The experiment began with an instruction screen detailing the structure of the block of trials to follow. Observers learned that they would see three categories of images: keys, clocks (Experiment 1A) or locks (Experiment 1B), and a random neutral category, which would contain images from the neutral category that did not include keys or clocks/locks. They were instructed to click on the image from the neutral category (not the key or clock/lock) as quickly and accurately as possible.

The experimenter checked that the observer fully understood the instructions before initiating the trials. On each trial, a fixation cross (Arial font size 12; 5 mm) appeared at the center of the screen until the observer used the mouse to click directly on it. This triggered presentation of the fixation cross for 500 ms, followed by the search display comprising two images: one of random neutral subject matter (e.g., book, chair, computer) and the other containing an image of keys or clocks (Experiment 1A) or an image of keys or locks (Experiment 1B). The observer was instructed to click on the neutral picture and to ignore any image containing keys or clocks/locks. The types of stimuli that would appear with the neutral pictures were indicated to the observer before the start of each block. A 300-ms intertrial interval passed, and the next trial began with the fixation cross to click in the center of the screen. Figure 1 shows the sequence of events in a typical trial.

All location combinations of nontarget and neutral stimuli were equally represented across the experiment. Each experiment (Experiments 1A and 1B) was organized into four blocks. Two blocks had two sets of 36 TC trials presenting unpredictably either a key or clock nontarget image on each trial; the other two

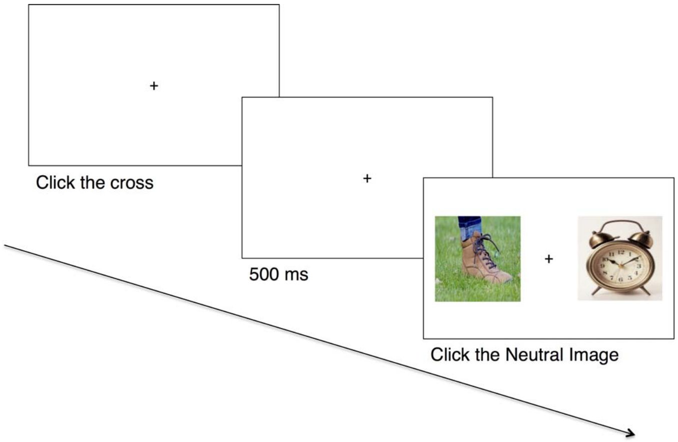

Figure 1. Sequence of displays in a typical trial from Experiments 1A and 1B. Each trial ended and RTs were recorded when the neutral category image was clicked.

blocks were OC blocks containing two segregated sets of 36 trials in which each set contained only one type of nontarget stimulus paired with the neutral target (36 key paired with neutral followed by 36 clock paired with neutral). The intermixed nontarget TC blocks and segregated nontarget OC blocks were presented in a counterbalanced ABBA sequence, where "A" denotes intermixed blocks for half the observers and segregated blocks for the other half. Additionally, within the segregated blocks, the order of the nontarget stimuli to be paired with the neutral target was counterbalanced across observers.

Results and discussion

Accuracy was high ($M = 98.51\%$, $SD = 1.96$), and a stem-and-leaf diagram identified no observers with extreme reaction times (RTs). Figure 2 plots mean RTs across observers separately for Experiments 1A and 1B (left and right pairs of columns, respectively) and for OC versus TC blocks (light gray and dark gray, respectively).

Visual inspection of the plot suggested that, as we had predicted, the TC condition showed equivalent performance to the OC condition in Experiment 1B (semantically related nontarget categories) but worse performance in Experiment 1A (semantically unrelated nontarget categories). These impressions were confirmed in a mixed two-way analysis of variance (ANOVA), with a between-observers factor of experiment (Experiment 1A: semantically unrelated categories; Experiment 1B: semantically related categories) and a within-observer factor of categories (categories to be inhibited on each trial: OC vs. TC). This yielded no main effect of experiment, $F(1, 38) = 0.014$, $p = 0.906$, a main effect of categories, $F(1, 38) = 6.402$, $p = 0.016$, and an interaction between categories and experiment, $F(1, 38) = 7.322$, $p = 0.010$. This interaction indicated that the costs of rejecting two categories in the TC condition (relative to the OC condition) were greater in Experiment 1A than in Experiment 1B.

To investigate the source of this interaction, we performed paired-samples t tests on OC versus TC

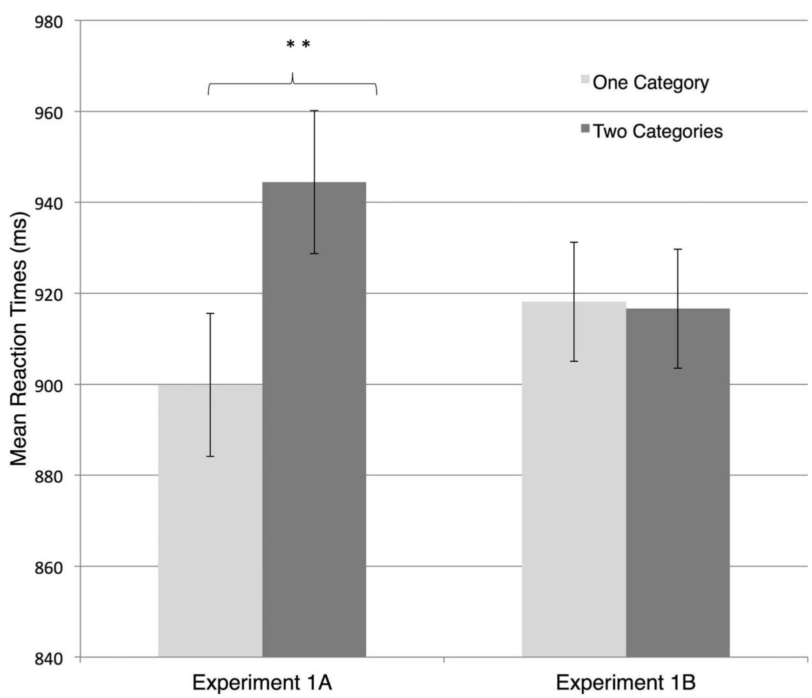

Figure 2. Mean RTs for OC and TC blocks in Experiments 1A (unrelated nontarget stimuli) and 1B (semantically related nontarget stimuli). Error bars = 1 standard error of the mean for paired differences (SEMpaireddiffs -Franz & Loftus, 2012).

means separately for Experiments 1A and 1B. As the semantic TFR view had predicted, there was a significant difference between OC ($M = 900$ ms, $SD = 143$) and TC ($M = 944$ ms, $SD = 150$) RTs for Experiment 1A, $t(19) = 4.075$, $p < 0.001$, when the two categories of nontargets were not semantically related. This difference was reduced and seemingly absent in Experiment 1B, $t(19) = 0.114$, $p = 0.910$ (OC: $M = 918$ ms, $SD = 103$; TC: $M = 917$ ms, $SD = 114$), for the potentially semantically related categories. These results provided initial evidence that when two nontarget categories were semantically related, they essentially could be treated as one category even though their distinctively different visual characteristics lacked any overlap. Templates for search in these displays seemed to have used semantic features to increase search efficiency.

One alternative possible explanation for our findings was that observers could not establish an effective search template for locks (in Experiment 1B) as efficiently as they could for clocks (in Experiment 1A). If adequate search templates for locks could be constructed in either the OC or TC conditions of Experiment 1B, this would have acted to minimize the two-category cost in that study. To check that our lock stimuli did not intrinsically yield much smaller two-category costs than our clock stimuli (rather than only in combination with a related category—keys—as we supposed), we ran an additional 10 observers (five females, five males; aged 23–30 years) on an altered version of Experiment 1A in which the two categories of nontargets were the clocks and locks of Experiments 1A and 1B. A two-way repeated measures ANOVA on those observers' RTs, with the within-observer factors of categories (OC vs. TC) and image type (clocks vs. locks), yielded a main effect of categories, $F(1, 9) = 12.574$, $p = 0.006$, and a marginal main effect of image type, $F(1, 9) = 5.099$, $p = 0.05$, but, crucially, no hint of an interaction between categories and image type, $F(1, 9) = 0.387$, $p = 0.549$. This last term indicated that when either object category is paired with a semantically unrelated object category, the two-category cost is similar for clock OC ($M = 754$ ms, $SD = 55$) and TC ($M = 812$ ms, $SD = 97$) trials, $t(9) = -2.638$, $p = 0.027$, versus lock OC ($M = 778$ ms, $SD = 76$) and TC ($M = 824$ ms, $SD = 82$) trials, $t(9) = -4.148$, $p = 0.002$. The results of Experiments 1A and 1B could not be explained in terms of locks yielding smaller two-category costs than clocks but rather only in terms of the relationships between locks and keys.

We supposed that our findings reflected semantic overlap of categories in Experiment 1B; however, an alternative possibility was that our locks and keys stimuli may have overlapped in some physical features that our clocks and keys did not and that observers' templates may have used visual information, not semantic information. One method of countering this criticism might be to use multiple related and unrelated object category pairs to dilute any such effects. However, such potential criticisms could be fully excluded only if the semantically unrelated objects were physically identical to the semantically related ones in two experiments.

In Experiment 2 we attempted to manipulate the cognitive set of observers to emphasize the visual distinctiveness of keys and locks. We hoped that holding the visual information constant across the nontarget object categories of two experiments (keys and locks in both cases) could train observers either (a) to consider keys and locks as semantically distinct (Experiment 2A) and hence reinstate a two-category cost or (b) to consider locks and keys as semantically similar (Experiment 2B) and hence find no cost (as we had in Experiment 1B). Such a result would bolster our view that the semantic relationship between nontarget categories in Experiment 1B had allowed for the efficient treatment of two categories of objects as one semantic category.

Experiments 2A and 2B

Experiment 2 examined the source of differing two-category costs in Experiments 1A and 1B. Although keys and locks share little more in common visually than keys and clocks, observers exhibited no two-category costs for keys and locks. We believed this was because observers' TFR specified not the dissimilar visual characteristics of keys and locks but rather semantic information that overlaps for those categories. To test this idea further, Experiment 2 manipulated perceived semantic relationships between keys and locks with differing training sessions. For Experiment 2A, training emphasized the differing characteristics of keys and locks. Experiment 2B directed emphasis to the shared semantic features of keys and locks. If two-category costs were to arise for Experiment 2A as they did for Experiment 1A and remain absent for Experiment 2B, this would provide further evidence that semantic relationships between keys and locks, rather than any overlapping visual features, supported reduced two-category costs in Experiment 1B versus Experiment 1A. The stimuli in Experiments 2A and 2B were identical, so any difference in two-category cost could not reflect visual feature overlap alone.

Observers

A total of 40 paid observers (Experiment 2A: six males, 14 females, aged 19–35 years; Experiment 2B:

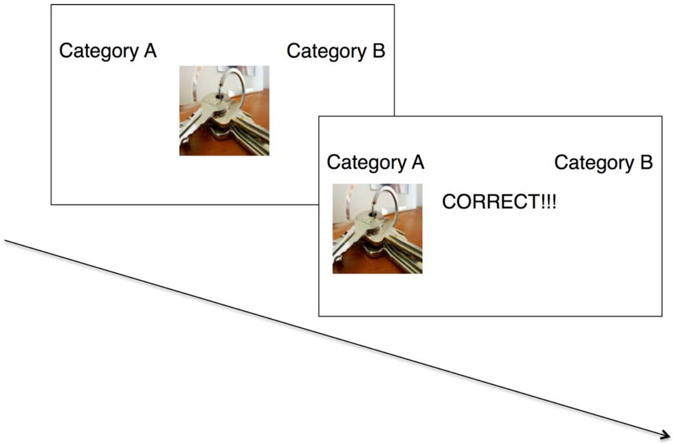

Figure 3. A typical trial in the training phase of Experiments 2A and 2B. Observers responded by pressing the "X" (left; Category A) or "M" (right; Category B) keys to indicate to which category they believed the stimulus belonged.

eight males, 12 females, aged 18–32 years) from the University of Cambridge and the local area participated after giving informed written consent.

Materials

All equipment used in Experiment 2 was identical to that used in Experiment 1. For the training phases of Experiment 2, observers viewed 152 images of objects, half of which depicted objects into which another object is inserted (38 locks, 38 miscellaneous electrical sockets, USB ports, and so on). The other half depicted images of objects that are inserted into other objects (38 keys, 38 miscellaneous electrical plugs, USB sticks, and so on). These images were presented individually in boxes (150 × 150 pixels) in a randomized order at the center of the display. Short 1-s sound clips of crowd applause or crowd booing were also played at a comfortable volume on the external computer speakers. In Phase 2, the observers in Experiments 2A and 2B performed the same task (with exactly the same stimuli) as described for Experiment 1B (see Figure 1). None of the images from the training phase of Experiment 2 were used in the test portion.

Procedure

Experiment 2A training

Stimuli were assigned to two categories that would determine the observer's correct response: female and male. The *female* category consisted of items containing niches for corresponding parts, including locks, computer USB ports, wall sockets, seatbelt containers, and other sundry objects that have niches for insertion. The *male* category consisted of objects that fit inside their corresponding counterparts, including keys, USB sticks, plugs, seatbelts, and other sundry objects that are inserted into other objects. Therefore, locks and keys were assigned to different response categories in the Experiment 2A training phase, drawing attention to their semantic differences.

Experiment 2B training

The same stimuli from Experiment 2A were differently arranged into two new categories. The first category (locks and keys) consisted of all the images comprising locks or keys, and the second category (others) consisted of all the other male and female objects described above that were not locks or keys. Hence, in Experiment 2B, lock and key images were assigned to the same response category to bring attention to the shared semantic features of locks and keys.

The training phases for Experiments 2A and 2B began with an instruction screen detailing the events to occur as well as instructions for how to complete each trial. Observers were instructed to categorize images into one of two undefined categories presented on the screen and to learn through trial and error the correct categorization. The attending experimenter confirmed that the observer fully understood the instructions before launching the experiment. The text *Category A* was centered 43 mm by 90 mm above and left of fixation, and the text *Category B* was centered at the corresponding position to the right of fixation (both in size 24 font), and both remained on the screen for the entirety of the training phase. Observers were instructed to classify the centrally presented image as either Category A or Category B by pressing "X" for Category A and "M" for Category B (keys pertaining to the boxes on the left and the right of the screen, respectively). Observers received feedback to their response in the form of the words *correct* or *incorrect* presented in the center of the screen in size 48 font with a corresponding positive (a crowd cheering) or negative (crowd booing) sound for 1 s each. Figure 3 illustrates a typical trial's sequence of displays from the training phase. Observers were instructed to try to respond as accurately as possible.

In Experiment 2A, categorizing female objects into Category A and male objects into Category B received positive feedback and other responses received negative feedback. In Experiment 2B, responses assigning images comprising locks and keys into Category A or other stimuli into Category B received positive feedback, and other responses received negative feedback. Images remained on the screen until the observer responded and received feedback. Upon completion of this phase of the experiment, observers completed a

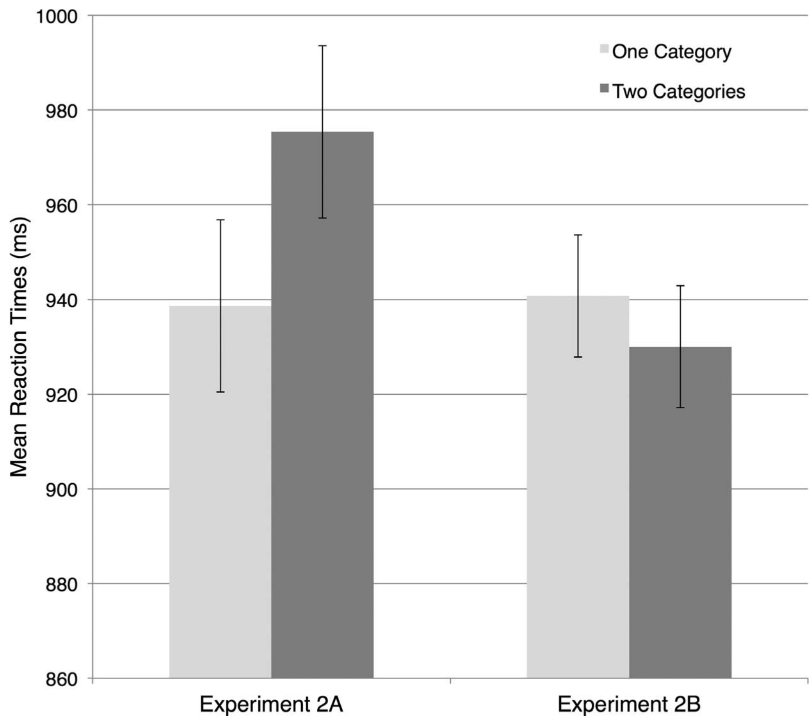

Figure 4. Mean RTs for OC and TC trials in Experiments 2A (training with emphasis on physical difference between keys and locks) and 2B (training with emphasis on semantic relationship between keys and locks). Error bars = 1 SEMpaireddiffs.

short questionnaire asking "What images were categorized into Category A?" and "What connection do you make between these images that would categorize them together?" The same questions were asked of Category B.

Following the training phase, observers each performed the same trials as were used in Experiment 1B. If the training phases of Experiments 2A and 2B had elicited a cognitive set of keys and locks as semantically similar in Experiment 2B but semantically dissimilar in Experiment 2A, the semantic TFR view would predict two-category costs in Experiment 2A but not in Experiment 2B.

Results and discussion

Observers again made very few errors ($M = 98.58\%$, $SD = 1.45$). A stem-and-leaf plot highlighted that two observers' response times from Experiment 2B were outliers: Their mean RTs were found to exceed 3 *SD* from the mean for the sample. They were excluded from analysis. (With them included the key terms discussed below remained significant, but those data likely distorted our estimate of the true mean of those scores.) We plotted (see Figure 4) mean RTs for Experiments 2A and 2B (using the same format as Figure 2) separately for OC and TC conditions.

Visual inspection of the plot suggested that there was a TC cost only for Experiment 2A and that it was reduced or nonexistent for Experiment 2B. These impressions were confirmed in a two-way mixed ANOVA with a between-observers factor of experiment (Experiment 2A vs. Experiment 2B) and a within-observer factor of categories (OC vs. TC). This yielded no main effect of experiment, $F(1, 36) = 0.274$, $p = 0.604$, and no main effect of categories (OC vs. TC), $F(1, 36) = 1.56$, $p = 0.219$, but a significant interaction between these two factors, $F(1, 36) = 5.16$, $p = 0.029$.

To investigate the source of this interaction, we performed *t* tests on OC versus TC means separately for Experiments 2A and 2B. Paired-samples *t* tests revealed a significant two-category cost in Experiment 2A between OC ($M = 939$ ms, $SD = 131$) and TC ($M = 975$ ms, $SD = 123$); $t(19) = 2.28$, $p = 0.034$. As we expected, however, the corresponding *t* test for Experiment 2B revealed no significant difference in RTs between OC ($M = 941$ ms, $SD = 141$) and TC ($M = 930$ ms, $SD = 300$) trials, $t(17) = -0.84$, $p = 0.414$.

In Experiment 2A, training observers to categorize keys and locks as two separate semantic categories (male and female) based on their differing visual features seems to have elicited a two-category cost—greater difficulty in rejecting two types of nontarget than one type—much like that seen in Experiment 1A. This effect of the training phase could not be due to fatigue or greater exposure as these aspects were equated in Experiments 2A and 2B.

Experiment 2B successfully replicated the findings of Experiment 1B, increasing our confidence in that

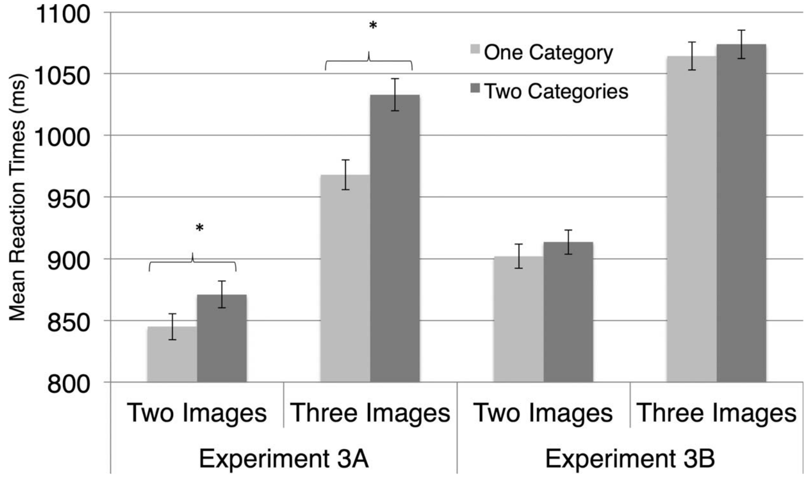

Figure 5. Mean RTs for OC versus TC trials for two-image and three-image trials in Experiments 3A (semantically unrelated nontargets) and 3B (semantically related nontargets). Error bars = 1 SEMpaireddiffs.

unusual, perhaps counterintuitive, result. Although with conventional statistics one cannot make strong claims from a null result, finding a similar absence of evidence of a two-category cost in Experiment 2B certainly suggested that any such effect was small and not robust if it is present at all in those conditions. These results further bolstered the conclusion that semantic templates for rejection can be used in the type of search examined here.

Experiments 3A and 3B

Experiments 1 and 2 established the existence of performance costs for ignoring two semantically unrelated categories of objects versus one category (or two semantically related categories). In Experiment 3, we wanted to discover whether this cost would increase with increasing numbers of nontargets or remain constant. Similar to Experiment 1, Experiment 3 comprised two studies—Experiments 3A (semantically unrelated nontargets: clocks and keys) and 3B (semantically related nontargets: locks and keys—to reveal the effect of increasing nontargets in the presence and absence of a potential semantic relationship. Additionally, each study comprised trials presenting one nontarget with the neutral target and others presenting two nontargets of the same category with the neutral target.

Observers

A total of 40 observers (Experiment 3A: 11 males, nine female, aged 18–37 years; Experiment 3B: seven males, 13 females, aged 18–33 years) from the University of Cambridge and the local area gave written informed consent and were paid for participating.

Materials and procedure

Experiments 3A and 3B replicated the materials and procedures of Experiments 1A and 1B, respectively, but with the following exception. Each observer in Experiment 3A performed both a complete set of trials identical to those in Experiment 1A (unrelated categories of nontargets) and a second condition, which was identical other than presenting three images (two nontarget images from the same category and one target) per display. Half of the observers performed the two-image trials first and the other half performed the three-image trials. The same was true for observers in Experiment 3B, except that the nontarget categories were semantically related (locks and keys).

Results and discussion

Observers made few errors in Experiment 3 ($M = 94.42\%$, $SD = 4.46$). A stem-and-leaf plot suggested that two observers from Experiment 3A were outliers in terms of average RTs. Their mean RTs were found to exceed 3 *SD* from the mean, so they were excluded from analysis. Their exclusion did not affect the pattern of results. Figure 5 plots the mean RTs for one nontarget and two nontarget trials separately for Experiments 3A and 3B.

Visual inspection of the plot suggested that there was a two-category cost only for Experiment 3A and that this was reduced or nonexistent for both parts of Experiment 3B. Moreover, the size of the effect seemed to increase with the number of nontarget images for Experiment 3A. A mixed three-way ANOVA with a between-observers factor of experiment (Experiment 3A vs. Experiment 3B) and within-observer factors of categories (OC vs. TC) and set size (two versus three images) revealed no main effect of experiment, $F(1, 36) = 1.893$, $p = 0.177$. However, there was a main effect of categories, $F(1, 36) = 22.359$, $p < 0.001$, and a main effect of set size, $F(1, 36) = 123.76$, $p < 0.001$. Crucially, there was also an interaction between categories and experiment, $F(1, 36) = 8.813$, $p = 0.005$. To investigate the source of this interaction, we performed repeated measures ANOVAs on RTs from Experiments 3A and 3B separately with the within-observer factors of categories and set size.

Results from Experiment 3B, with the semantically related nontargets, showed a marginal main effect of categories, $F(1, 19) = 4.224$, $p = 0.054$, and a significant effect of set size, $F(1, 19) = 52.581$, $p < 0.001$; RTs increased with increasing numbers of nontargets. The interaction between these factors was not significant, $F(1, 19) = 0.041$, $p = 0.842$. The results for Experiment 3A revealed a main effect of categories, $F(1, 17) = 16.701$, $p = 0.001$, and set size, $F(1, 17) = 93.895$, $p < 0.001$, and an interaction between these two factors, $F(1, 17) = 7.935$, $p = 0.012$, indicating that the number of nontarget images significantly impacted the observers' ability to inhibit multiple categories. TC trials yielded slower RTs than OC trials for both two-image trials, $t(17) = -3.190$, $p = 0.005$, and three-image trials, $t(17) = -3.908$, $p = 0.001$, but this cost was larger in three-image trials, as reflected in the interaction term.

The trials of Experiment 3 at Set Size 2 were a replication of Experiment 1; thus, we should expect to see the same pattern of results in an analysis of those condition as in Experiment 1 (our thanks to an anonymous reviewer for noting this). To assess this, we performed *t* tests to compare the OC and TC trials at Set Size 2 in Experiments 3A and 3B. These revealed a difference between RTs for OC versus TC trials in Experiment 3A (OC: $M = 968$, $SD = 123$; TC: $M = 1033$, $SD = 165$), $t(17) = -3.908$, $p = 0.001$, but not Experiment 3B (OC: $M = 1064$, $SD = 159$; TC: $M = 1074$, $SD = 168$), $t(19) = -1.167$, $p = 0.258$. The mean difference in Experiment 3A was larger than that in Experiment 3B, $t(36) = 3.112$, $p = 0.004$.

The results from Experiment 3 suggested that, for semantically unrelated target categories, two-category costs increase with the number of nontargets presented. This pattern was not found for semantically related categories. This pattern is consistent with the notion that two-category costs reflect a less efficient guidance of attention away from nontargets in TC conditions relative to OC conditions or two semantically unrelated categories. However, these results may also be accounted for by assuming that the conditions did not affect the efficiency with which attention is guided but rather affected the speed of the observer's decision about whether or not each particular individual image is a nontarget. If, for example, observers used serial, self-terminating search and attended to each image individually in a display, and if the speed of their decision about each image (nontarget or target) were reduced for two unrelated categories relative to one category (or two related ones), that might also account for the patterns of two-category costs we observed in Experiments 1, 2, and 3.

Accordingly, in two further experiments, we investigated potential two-category costs for unrelated categories in displays that had only one image (minimizing the need for attention guidance) versus two or four images. We expected that in the one-image displays, as little attention guidance would be necessary, any two-category costs should largely reflect less efficient decisions about whether that single item was a target or a nontarget. In contrast, for two- or four-image displays, attentional guidance could play a role. Together, these conditions should provide an indication of the relative contributions to two-category costs of attentional guidance versus decision making about individual images. Unfortunately, this comparison required an alteration to our method. Rather than clicking on the target image in each trial, observers in Experiment 4 simply responded regarding whether a target was present or absent by pressing one of two keys accordingly. This change to the method was necessary because when only one image is present, if it is a target on all trials (as would have been the case using the task from Experiments 1, 2, and 3), observers could have clicked on the image without analyzing it. We no longer included semantically related categories for these conditions because we had already established the minimal two-category costs for those conditions to our satisfaction.

Experiment 4

Experiment 4 replicated the conditions of Experiment 3A but with the following exceptions. First, the observers' task was to determine whether a target image was present in each display comprising one, two, or four images. Second, to accommodate this variation in the task, half of the trials contained a target image and the other half contained only nontargets (i.e., no target image).

Observers

A total of 20 observers (three males, 17 females, aged 18–32 years) from the University of Cambridge and the local area gave written informed consent and were paid for participating.

Materials

All forms and computer materials for Experiment 4 were identical to those for Experiments 1, 2, and 3 except that responses were made via an Apple keyboard with the word *yes* printed on the "C" key and the word *no* printed on the "N" key.

Procedure

The procedure followed that of Experiment 3A with the following exceptions. In this experiment, one, two, or four images were presented on each trial and only half of the trials in each set size contained a neutral target image, while the other half presented only nontarget images (comprising keys or clocks). Observers responded by pressing a *yes* key as quickly and accurately as possible to indicate there was a neutral image (not of a key or clock) on the screen or a *no* key to indicate that there was no neutral image present. Incorrect responses were signaled by a brief beep. There was an intertrial interval of 500 ms, followed by the next image(s). Observers completed 384 trials, took a 5-min break, and repeated the experiment. The images used for the second round of the experiment were from the same pool as the first round of the experiment but were randomly allocated to the various trials. Thus, observers viewed the same images in a new, randomized order. The full experiment consisted of 768 trials, which were completed in approximately 30 min. Although the study comprised more trials than Experiment 3, the faster responses allowed the study to be completed in nearly the same amount of time as Experiments 1 and 2. This allowed us to increase the number of trials to maintain power now that there were three set sizes without increasing the number of observers.

Single-image trials could present the target neutral image, clock, or key in any of the four possible locations around the central fixation cross. Two-image presentations presented two keys, two clocks, or a neutral target with a clock or key presented to (a) either side of the horizontal midline or (b) above and below the vertical midline. In four-image trials, observers were presented with four clocks, four keys, or three clocks (or three keys) and one neutral target.

The run order for Experiment 4 followed that of the previous experiments with four blocks of trials. Two blocks had two sets of 48 TC trials, where key and clock stimuli were presented in a randomized order; the other two blocks were OC trials containing two segregated sets of 48 trials, where each set contained only one category of nontarget stimuli. The number of trials per block was increased to 48 so that observers would view each possible combination (of target presence and item locations) twice in each block. The intermixed and segregated nontarget blocks were presented in a counterbalanced ABBA sequence, where "A" represented intermixed blocks for half the observers and segregated blocks for the other half. Note that one-, two-, and four-image trials were all presented within the same blocks of trials.

Results and discussion

Observers again made very few errors ($M = 95.60\%$, $SD = 4.24$). A stem-and-leaf diagram highlighted one outlier observer; they were excluded from analyses because their RTs were more than 3 *SD* from the mean. Figure 6 plots mean RTs for TC versus OC trials separately for Set Sizes 1, 2, and 4 and for target-present and target-absent trials. Visual inspection of the plot suggested that there was no influence of TC versus OC condition at Set Size 1, an effect at Set Size 2, and a tendency toward a reversed effect at Set Size 4 (in each case, the same tendencies arising for target-present and target-absent trials). A three-way repeated measures ANOVA with within-observer factors of categories (OC vs. TC), set size (one, two, four images), and target presence (present, absent) yielded a main effect of set size, $F(2, 17) = 68.294$, $p < 0.001$, and a significant interaction between categories and set size, $F(2, 17) = 4.409$, $p = 0.029$. The interaction confirmed that the two-category cost differed as a function of set size. There was no main effect of target presence, $F(1, 18) = 2.283$, $p = 0.148$, and it did not feature in any significant term involving categories, maximum $F(1, 18) = 1.428$, $p = 0.248$.

To reveal the source of the interaction between categories (OC vs. TC) and set size, we examined differences between OC and TC means at each set size (collapsing across target-present and target-absent conditions). At Set Size 1 there was no detected two-category cost, $t(18) = 0.333$, $p = 0.743$ (TC: $M = 769$ ms, $SD = 90$; OC: $M = 774$ ms, $SD = 94$); conditions showed very similar RTs. At Set Size 2, an effect was observed as in the other studies for TC ($M = 905$ ms, $SD = 111$) versus OC ($M = 879$ ms, $SD = 126$) trials, $t(18) = -2.475$, $p = 0.023$. However, at Set Size 4, where we expected the effect would be larger, there was no effect, $t(18) = 1.694$, $p = 0.108$, with any tendency in the opposite direction to that observed at Set Size 2 between OC ($M = 1129$ ms, $SD = 204$) and TC ($M = 1104$ ms, $SD = 173$) trials.

If the two-category cost (when it occurred) were due to decision making about whether individual stimuli were targets or nontargets, it should have been

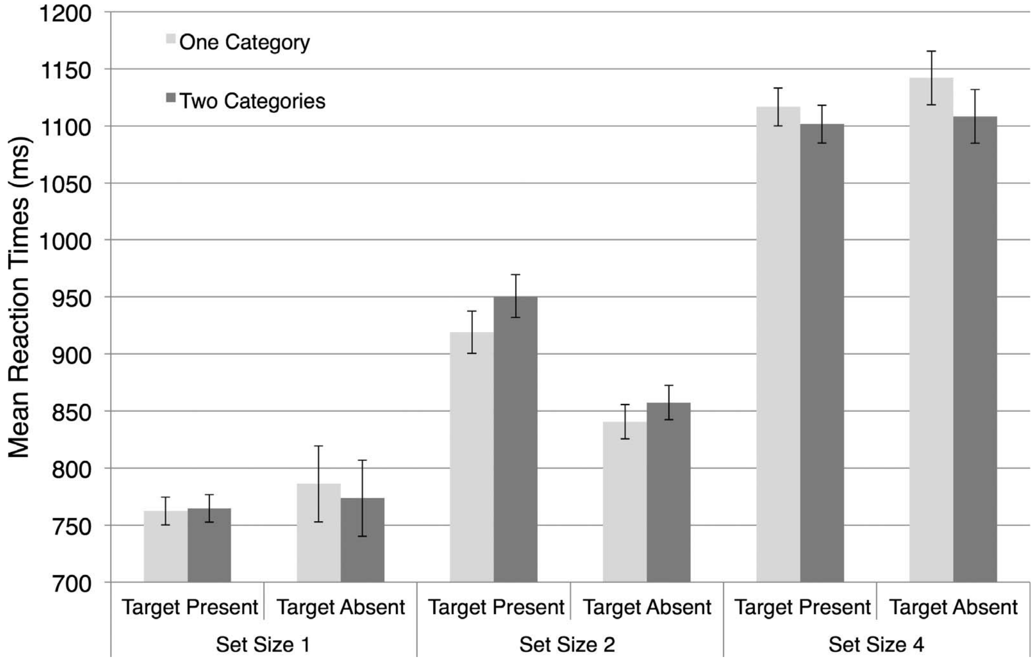

Figure 6. Mean RTs for OC versus TC trials at Set Sizes 1, 2, and 4 in Experiment 4 across target-present and target-absent trials. Error bars = 1 SEMpaireddiffs to facilitate comparison across the within-observer factors in Experiment 4 (Franz & Loftus, 2012).

observed at Set Size 1, but it was not. The null result at Set Size 1 suggested that this was not the case. Rather, in combination with the observed effect at Set Size 2, this finding strongly suggested that two-category costs in our experiments reflected impoverished attention guidance away from nontargets in TC conditions relative to OC conditions.

Given the effect sizes for the categories factor (OC vs. TC) in Experiments 1A and 3A (Cohen's $d = 0.75$ and 0.76 respectively), our power to detect an effect of similar size in Experiment 4 at Set Size 1 was around 0.94 with the original sample of 20 observers. While this power was more than adequate, the key conditions of Experiment 4 (Set Sizes 1 and 2) were key to this part of our argument. Accordingly, to increase our power to see smaller effects, we combined the data from Experiment 4's key conditions (Set Sizes 1 and 2) with those from 30 additional observers who ran the experiment under conditions identical to those of the original sample but without Set Size 4. We estimated that 50 observers would yield a power of approximately 0.98 to see an effect at Set Size 1 similar to that observed repeatedly for Set Size 2; given a lenient alpha of 0.1, we expected power to be around 0.90 to see an effect of half the size consistently observed for Set Size 2. Once outliers (whose mean RTs fell more than 3 *SD* from the sample mean) were removed, 46 observers' data remained, but power should still have been high. Calculating the two-category cost across target-present and target-absent trials for each observer, we then ran a mixed two-way ANOVA with a between-observers factor of sample (original Experiment 4 sample versus the 30 additional observers) and a within-observer factor of set size (one vs. two images). This yielded a main effect of set size, $F(1, 44) = 7.386$, $p = 0.009$, but no main effect of sample, $F(1, 44) = 2.018$, $p = 0.163$. Additionally, there was no interaction between these factors, $F(1, 44) = 0.556$, $p = 0.460$. Irrespective of sample, there remained a larger difference between OC and TC conditions at Set Size 2 than at Set Size 1. Importantly, despite increased power to see any effect, no two-category cost was observed at Set Size 1 (OC: $M = 837$, $SD = 146$; TC: $M = 844$, $SD = 145$); $t(45) = -0.563$, $p = 0.576$. In contrast, there remained a significant two-category cost at Set Size 2 (OC: $M = 977$, $SD = 176$; TC: $M = 1026$, $SD = 200$); $t(45) = 3.403$, $p = 0.001$. It therefore appeared that decision-making processes operating at Set Size 1 (when observers determined whether the single presented image was a target or nontarget) did not yield robust two-category costs; such hypothetical costs would not therefore be suitable to explain our very robust effects at Set Size 2. Rather, our OC versus TC effects seem specific to displays with more than one image and likely reflected differences in attentional guidance across conditions.

Only one aspect of the results seemed inconsistent with the semantic TFR view: the absence of an effect at Set Size 4. In fact, the numerical tendency, though not significant, was in the direction opposite to that we expected, primarily for target-absent trials. One obvious explanation for this was that perceptual load suppresses semantic processing and its effects at larger set sizes. Indeed, initially this was the authors' favored explanation, given that high perceptual load tends to suppress late selection processes on the basis of semantics (for a recent discussion, see, e.g., Lavie, Beck, & Konstantinou, 2014). In opposition to this view, however, Belke et al. (2008) recently demonstrated that increased perceptual load at larger set sizes need not hinder the action of semantic target templates. On the basis of this result one might assume that TFR too would be unaffected and that load cannot therefore provide an explanation for the absence of a categories effect at Set Size 4. Such a conclusion would point instead to an alternative interpretation on the basis of our Set Size 4 displays containing three nontarget objects of the same category. These may powerfully bias attention toward themselves because of their similarity irrespective of a top-down search template. However, though the relationship of semantic TFR to perceptual load merits further study, speculation regarding isolated, unexpected findings such as this risks overfitting an explanation to one or two data points. We hope to address this finding in future work but do not discuss it further here.

In Experiment 5 we sought to clarify a further issue relating to this last experiment: whether the same pattern of findings using set sizes of one, two, and four stimuli would arise if we used the target-localization measure from Experiments 1, 2, and 3. The one-image trials might not yield meaningful data because they always comprised only a target and might therefore be clicked without being categorized as a target (though observers were not explicitly coached that this was the case). However, we still hoped to observe the same pattern of findings as in Experiment 4 for Set Sizes 2 and 4.

Experiment 5

Experiment 5 replicated the conditions of Experiment 3A but with the following exceptions. First, set sizes of one, two, and four images per display were intermixed within blocks of trials as in Experiment 4. Second, we ran an increased sample relative to the previous studies. In an initial ANOVA (with factors of categories and set size) based on 20 observers, we found that the key term of interest—the interaction between categories and set size—was marginal, $F(2, 17) = 2.530$, $p = 0.109$. To clarify whether this pattern of results was indeed similar to Experiment 4 and reflected reduced power to see an effect or, alternatively, whether it was evidence of a distinct pattern, we increased the sample size to 30 observers. With hindsight, we should have predicted that more observers would have been required given the expected reduced power of this study (fewer trials per condition per observer than our previous studies).

Observers

A total of 30 observers (18 females, 12 males, aged 18–24 years) from the University of Cambridge and the local area gave informed consent and participated in this study. One observer's data were excluded because after the experiment the observer reported using the nondominant hand to complete the task due to a recent injury to the dominant hand. Their data were recorded but not analyzed.

Materials and procedure

All forms and computer materials for Experiment 3 were identical to those for the previous experiments. Displays were identical to target-present displays from Experiment 4.

Results

Observers again made very few errors ($M = 97.31\%$, $SD = 2.92$). A stem-and-leaf diagram did not highlight any outliers for the study. Figure 6 plots mean RTs for TC and OC conditions separately for Set Sizes 1, 2, and 4. Note that the pattern for Set Sizes 1 and 2 looks similar to Experiment 4, though the nonsignificant trend toward a reversed two-category cost at Experiment 4, Set Size 4, is not evident here (see Figure 7). A two-way repeated measures ANOVA with within-observer factors of categories (OC vs. TC) and set size (one, two, four images) yielded main effects of categories, $F(1, 28) = 4.830$, $p = 0.036$, and set size, $F(2, 27) = 199.271$, $p < 0.001$. More importantly, there was a significant interaction between these two factors, $F(2, 27) = 3.946$, $p = 0.031$.

To investigate this interaction, we performed paired-samples t tests for the means of the OC and TC at each set size. Paralleling our findings in Experiment 4, there was no two-category cost at Set Size 1 (OC: $M = 648$ ms, $SD = 99$; TC: $M = 651$ ms, $SD = 90$), $t(28) = -0.323$ (*ns*), but a cost was evident at Set Size 2 (OC: $M = 847$ ms, $SD = 128$; TC: $M = 881$ ms, $SD = 142$), $t(28) = -3.477$, $p = 0.002$. There was no such effect at Set Size 4, $t(28) = -1.345$, $p = 0.189$, though the mean RTs fell in the same direction as Set Size 2 (OC: $M = 1099$ ms, $SD = 214$; TC: $M = 1121$ ms, $SD = 205$).

Overall, Experiment 5 showed a pattern similar to that of Experiment 4: no two-category cost at Set Sizes 1 or 4 and a clear effect at Set Size 2. There were nonsignificant trends in the opposite directions at Set Size 4 in the two studies, which may simply indicate no overall effect at that set size. Indeed, any trend toward a reversed two-category cost arose only in the target-absent trials of Experiment 4 that do not relate directly

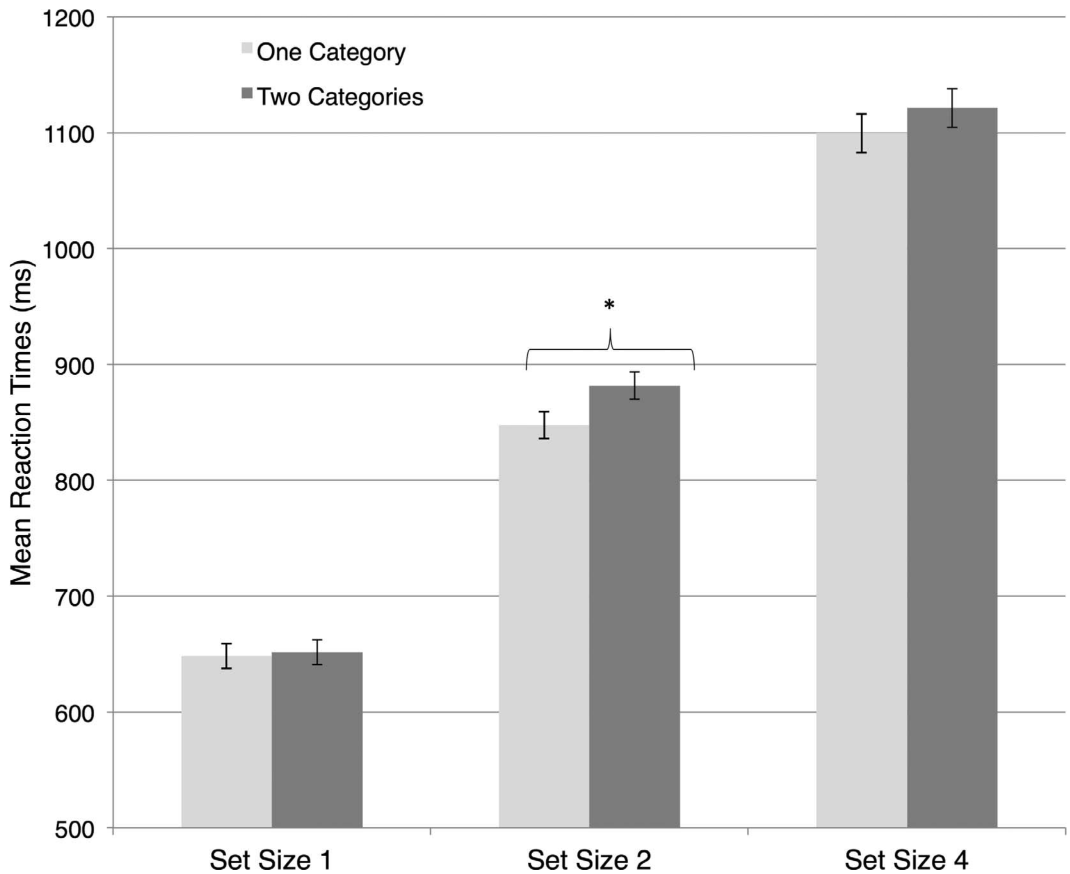

Figure 7. Mean RTs for OC and TC trials at Set Sizes 1, 2, and 4 in Experiment 5. Error bars = 1 SEMpaireddiffs.

to the other experiments. That single condition may yet yield interesting avenues for further research, but on the basis of the evidence reported here there seems to be little reason to assume that effect is real.

Experiments 6A and 6B

One final concern that might be raised regarding the pattern of findings in our experiments is that observers may not have used TFR to ignore nontarget stimuli but rather may have used standard search templates to guide attention to the nontarget before clicking the other image. This strategy has been termed *search and destroy* (Moher & Egeth, 2012) and has hindered interpretation of previous work on TFR. On the search-and-destroy view, our two-category costs would emerge from either less efficient search (guidance of attention to nontargets) or less efficient decisions to destroy (to reject the selected item as an irrelevant nontarget). We already had evidence against the latter possibility because such decision-making effects should have given rise to two-category costs in the Set Size 1 condition of Experiment 4, which it did not. However, less efficient target-search components could not be addressed by Experiments 1 through 5. Accordingly, in a final experiment, we reversed the task of Experiments 1A and 1B, asking observers to search for the nontarget items they had ignored in those original experiments. If the pattern of results in Experiment 1 had reflected observers searching for nontargets as targets and then simply clicking on the other image, we should expect to find similar patterns of results in this new task.

Observers

A total of 40 observers (Experiment 6A: five males, 15 females, aged 18–44 years; Experiment 6B: nine males, 11 females, aged 18–30 years) from the University of Cambridge and the local area gave written informed consent and were paid for participating.

Stimuli, apparatus, and procedure

Stimuli, apparatus, and procedure were all identical to those in Experiments 1A and 1B except that

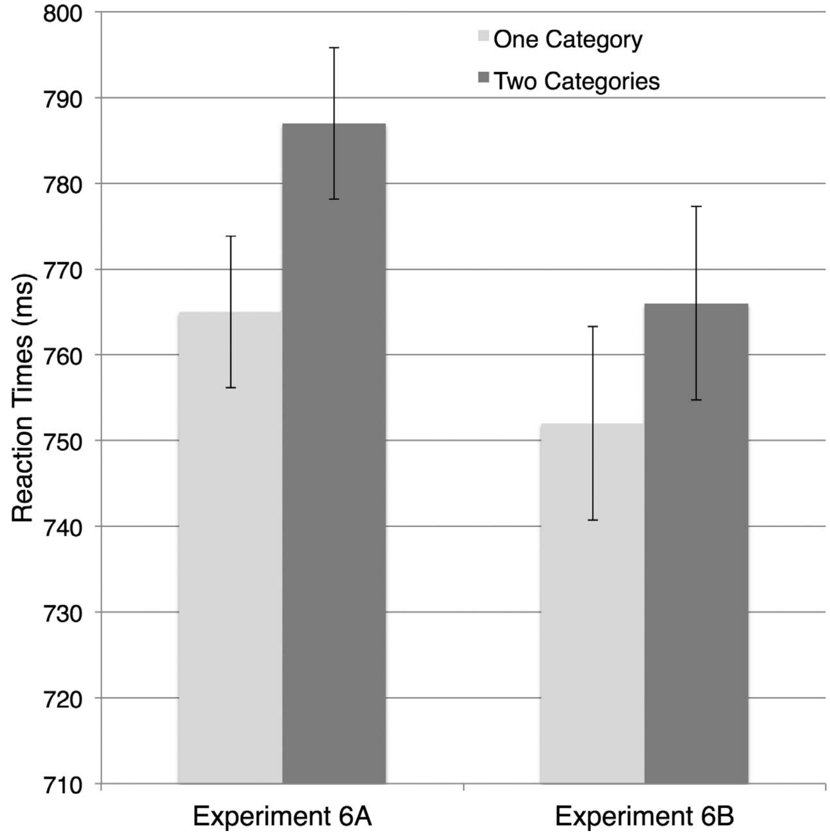

Figure 8. Mean RTs for OC and TC trials for Experiments 6A and 6B. Error bars = 1 SEMpaireddiffs.

observers were asked to move the cursor over the image comprising specific categories of objects (and then press a button) rather than to ignore them as nontargets.

Results

Accuracy was high ($M = 99.25\%$, $SD = 1.04$), and a stem-and-leaf diagram identified no observers with extreme RTs. Figure 8 plots mean RTs across observers separately for Experiments 6A and 6B. In contrast to the results of Experiments 1A and 1B, visual inspection of the plot suggested that weak two-category costs of roughly equal magnitude were evident in Experiments 6A and 6B.

Confirming this initial impression, a mixed two-way ANOVA with a between-observers factor of experiment (Experiment 6A vs. Experiment 6B) and a within-observer factor of categories (OC vs. TC) yielded no main effect of experiment, $F(1, 38) = 0.350$, $p = 0.558$, a main effect of categories, $F(1, 38) = 6.465$, $p = 0.015$, and, crucially, no hint of the interaction that so clearly characterized Experiments 1, 2, and 3, $F(1, 38) = 0.389$, $p = 0.537$. Another similar study not described here showed this same pattern. Experiment 6A showed a weak but significant two-category cost, $t(19) = -2.395$, $p = 0.027$; Experiment 6B showed a slightly weaker trend in the same direction, $t(19) = -1.314$, $p = 0.204$.

Any apparent difference between the two-category costs for Experiment 6A versus Experiment 6B reflected an outlier score (>3 SD above the mean of the other 19 observers), without which the two means differed by only 3 ms. With this outlier score removed, we compared the two-category costs in Experiment 2B 1A and 1B versus Experiment 2B 6A and 6B in a two-way between-observers ANOVA with factors of template (TFR in Experiment 2B 1A and 1B vs. standard template in Experiment 2B 6A and 6B) and semantic relatedness of categories (unrelated in Experiment 2B 1A and 1B vs. related in Experiments 1B and 6B). This yielded no main effect of template, $F(1, 75) = 0.306$, $p = 0.582$, a main effect of semantic relatedness, $F(1, 75) = 5.122$, $p = 0.027$, and, crucially, an interaction between the two factors, $F(1, 75) = 3.809$, $p = 0.055$. Despite the greatly reduced power to detect an interaction in this analysis relative to all the others described here, there was evidence that the effect of semantic relatedness on the two-category cost was different (greater) in Experiment 1 versus Experiment 6. This finding effectively precluded any explanation of our basic findings from Experiments 1 through 5 in terms of search-and-destroy processes; on that view, the results of Experiments 1 and 6 should mirror one another closely.

One further difference between Experiments 1 and 6 was overall faster response times in Experiment 6. This result is in line with previous findings using simple stimuli (e.g., Arita et al., 2012) suggesting that slower response times may be a general characteristic of search using TFR as opposed to standard search templates. However, the current results do not offer any ground for speculation about the cause(s) of this effect, which was not our focus here.

General discussion

Models of visual search typically incorporate top-down search templates, which bias attention away from irrelevant nontargets and toward relevant target items. The vast majority of previous work in this area has focused exclusively on target templates that enhance processing of features associated with target items, typically using large arrays of abstract items. Those studies have tended to conclude that target templates' efficacy is determined by how proficiently targets can be specified at the level of visual features available in the retinal image, not high-level or semantic features.

Other recent work has begun to examine target templates in search displays and use tasks that resemble some aspects of real-world search. Some of these studies have reported that in search for categories of naturalistic stimuli, attention can stray from target items to semantically related items. This is consistent with target templates acting to bias attentional processes at higher category-level vision or in terms of semantic properties. However, as we discussed earlier, there is a limitation to using those procedures that make interpretation of their findings difficult. Those findings provide suggestive evidence of semantic influences of templates, but convergent evidence from other approaches is needed.

TFR have received much less attention than target templates. Prior to the current work there had been no direct assessment of the relative importance of semantic limitations versus visual (retinal image) feature limitations on TFR in natural search. Accordingly, we opted to extend the logic of a paradigm previously used to study target templates.

Previous work had already suggested that observers would reject nontargets less efficiently when they belonged unpredictably to either of two categories than when they belonged to one prespecified category. However, our approach required both this and an additional assumption: that if there was sufficient overlapping information (at the level of semantic properties or visual features) between the two categories, the observer's TFR could specify this information and reject nontargets as efficiently when they belonged to either of two categories compared with just one prespecified category. This assumption, strongly supported by intuition, was subsequently borne out clearly in our results.

In Experiment 1, observers were presented with a nontarget from one of two known categories and an unknown target. Observers were instructed to ignore images from the nontarget categories and to select the neutral image as quickly and accurately as possible. The two categories in each study were chosen to be physically dissimilar, offering little overlap in visual features. Accordingly, we expected that if there was also little semantic overlap between the two categories to be inhibited, search should be less efficient (and RTs therefore slower) when nontargets could be either of two categories versus when they belonged to a single prespecified category. Such was indeed the case: A clear two-category cost was evident for Experiment 1A's unrelated nontarget categories. However, in Experiment 1B, the semantic overlap between the nontarget categories would allow for both types of nontargets to fall under a single rejection template, provided that the search template could influence performance via its semantic properties rather than its visual features. (Findings extended to three-image displays in Experiment 3.)

Experiment 2 further explored the idea of grouping by semantic relationships, controlling absolutely for visual similarities between the two categories of nontargets. In one condition (Experiment 2A), we aimed to break down the semantic relationship between keys and locks using a training phase to encourage observers to focus on semantic features of each category that did not overlap, treating them as belonging to distinct categories. In Experiment 2B's training phase we provided the same exposure to the stimuli, but the task encouraged observers to treat locks and keys as a single category, highlighting the semantic overlap between the two categories. In Experiment 2A, we found a clear two-category cost, but did not do so in Experiment 2B. This provided the strongest evidence yet of search templates operating at the level of semantic rather than visual features.

While the main difference between Experiment 2A and Experiment 1B was the trained emphasis on semantic distinctiveness in Experiment 2A versus the natural semantic distinctiveness in Experiment 1B, Experiment 2A mimicked the results found in Experiment 1A. These results suggested (a) that we were successful at redirecting the cognitive organization from a top-down overarching category to a focus on the bottom-up individual visual parts and (b) that the shared semantic relationship between keys and locks was used to create a single search template in Experiments 1B and 2B.

Experiments 4 and 5 distinguished two possible sources of two-category costs: whether they reflected less efficient guidance of attention away from the nontarget stimuli in TC versus OC conditions or whether they instead reflected less efficient decisions about whether or not each individual scene was a nontarget. Experiments 4 and 5 included single-image trials that minimized the need for high-level guidance of attention yet required decision making about the single search item. The two-category costs of Experiments 1, 2, and 3 were still observed in two-image displays but not in one-image displays. We concluded that the two-category cost did not reflect speed of decision making about any individual image (as this should have been expressed in the one-image scenes) but rather must have reflected changes in the efficiency with which attention was guided.

The absence of clear two-category costs in Experiments 4 or 5 for four-image displays (evident for three-image displays in Experiment 3) initially suggested that such semantically mediated guidance of attention might have been suppressed once four scenes were present, perhaps because a capacity limit was exceeded in those conditions (Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2001). However, we also entertained a second explanation—that the presentation of three nontargets of the same category may have provided a sufficiently strong influence on attention to override the influence of the search template. Determining which—if either—of these accounts is true will require further study beyond the scope of the current article.

Finally, in Experiments 6A and 6B we examined search for the same nontarget items that observers had ignored in Experiments 1A and 1B. The pattern of results in Experiment 6 was markedly different from that in Experiment 1. We concluded that the pattern of results in Experiment 1 was unlikely to be best accounted for in terms of the processes (including, in particular, standard search templates, specifying target items' features) that operated in Experiment 6.

Why there should be no evidence of semantic influences on target templates in Experiment 6 (and in a further similar experiment) remains unclear. It is the focus of our next planned project. Our favored explanation is that two-category costs are substantially weaker for target templates (at least under the conditions described here), and thus there may be relatively little benefit of strategically using a template to specify semantic overlap between two categories. Again, though, this question is beyond the scope of the current article.

To conclude, the experiments reported here suggest that two-category costs can provide useful insights into the structure and functions of search templates, including their semantic properties. Our findings strongly suggest that TFR can specify semantic properties of nontargets in order to bias attention away from objects with those attributes. In so doing, they have also provided new evidence for the existence of TFR in general as distinct entities from search templates. TFR appear to direct attention to relevant target items by excluding known irrelevant nontargets from search. The findings we ascribed to TFR do not appear to be reducible to actions of standard search templates in a search-and-destroy manner but rather are distinct entities. Given their likely importance in supporting behavioral inhibition (preventing repeated attention to stimuli that the observer wishes to ignore), they merit further study in their own right.

Keywords: template for rejection, search template, two-category cost, semantic, natural scenes, real-world search

Acknowledgments

We are grateful to Stephen and Linda Daffron for paying the page costs for this article.

Commercial relationships: none.
Corresponding authors: Jennifer L. Daffron; Greg Davis.
Email: jd597@cam.ac.uk; gjd1000@cam.ac.uk.
Address: Department of Psychology, University of Cambridge, Cambridge, UK.

References

Arita, J. T., Carlisle, N. B., & Woodman, G. F. (2012). Templates for rejection: Configuring attention to ignore task-irrelevant features. *Journal of Experimental Psychology*: *Human Perception and Performance, 38,* 580–584, doi:10.1037/a0027885. [PubMed] [Article]

Belke, E., Humphreys, G. W., Watson, D. G., Meyer, A. S., & Telling, A. L. (2008). Top-down effects of semantic knowledge in visual search are modulated by cognitive but not perceptual load. *Perception & Psychophysics, 70,* 1444–1458, doi:10.3758/PP.70.8.1444. [PubMed]

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97,* 523–547, doi:10.1037//0033-295X.97.4.523. [Article]

Bundesen, C., Habekost, T., & Kyllingsbaek, S. (2005).

A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review, 112,* 291–328, doi:10.1037/0033-295X.112.2.291. [PubMed]

Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision, 9*(1):34, 1–9, doi:10.1167/9.1.34. [PubMed] [Article]

Burge, J., Fowlkes, C. C., & Banks, M. S. (2010). Natural-image statistics predict how the figure-ground cue of convexity affects human depth perception. *The Journal of Neuroscience, 30,* 7269–7280, doi:10.1523/JNEUROSCI.5551-09.2010. [PubMed] [Article]

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. *Psychonomic Bulletin & Review, 15,* 795–801, doi:10.3758/PBR.15.4.795. [PubMed]

Cohen, D. J. (2009). Integers do not automatically activate their quantity representation. *Psychonomic Bulletin & Review, 16,* 332–336, doi:10.3758/PBR.16.2.332. [Article]

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers, 25*(2), 257–271.

Cohen, M. A., Konkle, T., Rhee, J. Y., Nakayama, K., & Alvarez, G. A. (2014). Processing multiple visual objects is limited by overlap in neural channels. *Proceedings of the National Academy of Sciences, USA, 111,* 8955–8960, doi:10.1073/pnas.1317860111. [PubMed]

Çukur, T., Nishimoto, S., Huth, A. G., & Gallant, J. L. (2013). Attention during natural vision warps semantic representation across the human brain. *Nature Neuroscience, 16,* 763–770, doi:10.1038/nn.3381. [PubMed]

Cunningham, C. A., & Wolfe, J. M. (2014). The role of object categories in hybrid visual and memory search. *Journal of Experimental Psychology: General, 143,* 1585–1599, doi:10.1037/a0036313. [PubMed]

Delorme, A., Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural images. *Cognitive Brain Research, 19,* 103–113, doi:10.1016/j.cogbrainres.2003.11.010. [PubMed]

Dent, K., Allen, H. A., Braithwaite, J. J., & Humphreys, G. W. (2012). Parallel distractor rejection as a binding mechanism in search. *Frontiers in Psychology, 3,* 278, doi:10.3389/fpsyg.2012.00278. [PubMed]

Duncan, J. (1983). Category effects in visual search: A failure to replicate the "oh-zero" phenomenon. *Perception & Psychophysics, 34,* 221–232. [PubMed]

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458. [Article]

Fabre-Thorpe, M., Delorme, A., Marlot, C., & Thorpe, S. (2001). A limit to the speed of processing in ultra-rapid visual categorization of novel natural scenes. *Journal of Cognitive Neuroscience, 13,* 171–180. [PubMed]

Franz, V. H., & Loftus, G. R. (2012). Standard errors and confidence intervals in within-subjects designs: Generalizing Loftus and Masson (1994) and avoiding the biases of alternative accounts. *Psychonomic Bulletin & Review, 19,* 395–404. [PubMed]

Friedman-Hill, S., & Wolfe, J. M. (1995). Second-order parallel processing: Visual search for the odd item in a subset. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 531–551. [PubMed]

García-Orza, J., Perea, M., Abu Mallouh, R., & Carreiras, M. (2012). Physical similarity (and not quantity representation) drives perceptual comparison of numbers: Evidence from two Indian notations. *Psychonomic Bulletin & Review, 19,* 294–300, doi:10.3758/s13423-011-0212-8. [PubMed]

Godwin, H. J., Hout, M. C., & Menneer, T. (2014). Visual similarity is stronger than semantic similarity in guiding visual search for numbers. *Psychonomic Bulletin & Review, 21,* 689–695, doi:10.3758/s13423-013-0547-4. [PubMed]

Hwang, A. D., Wang, H.-C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world images. *Vision Research, 51,* 1192–1205, doi:10.1016/j.visres.2011.03.010. [Article]

Kaiser, D., Stein, T., & Peelen, M. V. (2014). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proceedings of the National Academy of Sciences, USA, 111,* 11217–11222, doi:10.1073/pnas.1400559111. [PubMed]

Kim, M., & Cave, K. R. (1999). Top-down and bottom-up attentional control: On the nature of interference from a salient nontarget. *Perception & Psychophysics, 61,* 1009–1023. [PubMed]

Koshino, H. (2001). Activation and inhibition of stimulus features in conjunction search. *Psycho-*

nomic Bulletin & Review, 8, 294–300, doi:10.3758/BF03196164. [PubMed]

Kristjánsson, Á., Jóhannesson, Ó. I., & Thornton, I. M. (2014). Common attentional constraints in visual foraging. *PLoS One*, *9*(6), e100752, doi:10.1371/journal.pone.0100752. [PubMed]

Lavie, N., Beck, D. M., & Konstantinou, N. (2014). Blinded by the load: Attention, awareness and the role of perceptual load. *Philosophical Transactions of the Royal Society B: Biological Sciences*, doi:10.1098/rstb.2013.0205. [PubMed]

Lupyan, G. (2008). The conceptual grouping effect: Categories matter (and named categories matter more). *Cognition, 108,* 566–577, doi:10.1016/j.cognition.2008.03.009. [PubMed]

Mack, S. C., & Eckstein, M. P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision, 11*(9):9, 1–16, doi:10.1167/11.9.9. [PubMed] [Article]

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9*(11):8, 1–13, doi:10.1167/9.11.8. [PubMed] [Article]

Menneer, T., Cave, K. R., & Donnelly, N., (2009). The cost of search for multiple T\targets: Effects of practice and target similarity. *Journal of Experimental Psychology: Applied, 15*(2), 125–139, doi:10.1037/a0015331. [PubMed]

Moher, J., & Egeth, H. E. (2012). The ignoring paradox: Cueing distractor features leads first to selection, then to inhibition of to-be-ignored items. *Attention, Perception, & Psychophysics, 74,* 1590–1605, doi:10.3758/s13414-012-0358-0. [PubMed]

Moores, E., Laiti, L., & Chelazzi, L. (2003). Associative knowledge controls deployment of visual selective attention. *Nature Neuroscience, 6,* 182–189, doi:10.1038/nn996. [PubMed]

Müller, H. J., Humphreys, G. W., & Donnelly, N. (1994). SEarch via Recursive Rejection (SERR): Visual search for single and dual form-conjunction targets. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 235–258. [PubMed]

Potter, M. (1976). Short-term conceptual memory for pictures. *Psychology: Human Learning and Memory, 2,* 509–522. [PubMed]

Reeder, R. R., & van Peelen, M. (2013). The contents of the search template for category-level search in natural images. *Journal of Vision, 13*(3):13, 1–13, doi:10.1167/13.3.13. [PubMed] [Article]

Reeder, R. R., van Zoest, W., & van Peelen, M. (2015). Involuntary attentional capture by task-irrelevant objects that match the search template for category detection in natural scenes. *Attention, Perception, & Psychophysics, 77,* 1070–1080, doi:10.3758/s13414-015-0867-8. [PubMed]

Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural images. *Journal of Vision, 3*(6):5, 440–455, doi:10:1167/3.6.5. [PubMed] [Article]

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology, 62,* 1904–1914, doi:10.1080/17470210902853530. [PubMed]

Schwarz, W., & Eiselt, A.-K. (2012). Numerical distance effects in visual search. *Attention, Perception, & Psychophysics, 74,* 1098–1103, doi:10.3758/s13414-012-0342-8. [PubMed]

Seidl-Rathkopf, K. N., Turk-Browne, N. B., & Kastner, S. (2015). Automatic guidance of attention during real-world visual search. *Attention, Perception, & Psychophysics, 77,* 1881–1895, doi:10.3758/s13414-015-0903-8. [PubMed]

Sigman, M., Cecchi, G. A., Gilbert, C. D., & Magnasco, M. O. (2001). On a common circle: Natural images and Gestalt rules. *Proceedings of the National Academy of Sciences, USA, 98,* 1935–1940, doi:10.1073/pnas.031571498. [PubMed] [Article]

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2015). Disentangling the effects of spatial inconsistency of targets and distractors when searching in realistic scenes. *Journal of Vision, 15*(2):12, 1–21, doi:10.1167/15.2.12. [PubMed] [Article]

Telling, A. L., Kumar, S., Meyer, A. S., & Humphreys, G. W. (2010). Electrophysiological evidence of semantic interference in visual search. *Journal of Cognitive Neuroscience, 22,* 2212–2225. [PubMed]

Tkacik, G., Prentice, J. S., Victor, J. D., & Balasubramanian, V. (2010). Local statistics in natural images predict the saliency of synthetic textures. *Proceedings of the National Academy of Sciences, USA, 107,* 18149–18154, doi:10.1073/pnas.0914916107. [PubMed]

Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision, 5*(1):8, 81–92, doi:10:1167/5.1.8. [PubMed] [Article]

Walker, S., Stafford, P., & Davis, G. (2008). Ultra-rapid categorization requires visual attention: Images with multiple foreground objects. *Journal of*

Vision, 8(4):21, 1–12, doi:10.1167/8.4.21. [PubMed] [Article]

Wichmann, F. A., & Gegenfurtner, K. R. (2010). Animal detection in natural images: Critical features revisited. *Journal of Vision, 10*(4):6, 1–27, doi:10.1167/10.4.6. [PubMed] [Article]

Wolfe, J. M. (1994). Vision search in continuous naturalistic stimuli. *Vision Research, 34,* 1187–1195. [PubMed]

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real images. *Attention, Perception, & Psychophysics, 73,* 1650–1671, doi:10.3758/s13414-011-0153-3. [PubMed]

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: An alternative to the feature integration model for visual search. *Journal of Experimental Psychology*: *Human Perception and Performance, 15,* 419–433. [PubMed]

Woodman, G. F., & Luck, S. J. (2007). Do the contents of visual working memory automatically influence attentional selection during visual search? *Journal of Experimental Psychology*: *Human Perception and Performance, 33,* 363–377. [PubMed]

Wu, C.-C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world images. *Frontiers in Psychology,* 5, 54, doi:10.3389/fpsyg.2014.00054. [PubMed]

Wyble, B., Folk, C., & Potter, M. C. (2013). Contingent attentional capture by conceptually relevant images. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 861–871, doi:10.1037/a0030517. [PubMed]

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

Appendix

Example images of clock stimuli

Example images of key stimuli

Examples of lock stimuli