

Categorical templates are more useful when features are consistent: Evidence from eye-movements during search for societally important vehicles

Michael C. Hout¹, Arryn Robbins¹, Hayward J. Godwin², Gemma Fitzsimmons² & Collin Scarince¹

1. Department of Psychology at New Mexico State University, USA
2. Department of Psychology at University of Southampton, United Kingdom

Author Note: Correspondence regarding this article should be addressed to Michael C. Hout, New Mexico State University, Department of Psychology, Las Cruces, New Mexico, 88003, P.O. Box 30001 / MSC 3452. Telephone: 412-983-5290. Email: mhout@nmsu.edu

Acknowledgements: We thank Garrett Bennett and Alexis Lopez for their assistance in data collection.

Keywords: eye-movements, target templates, categorical search

Figures: 8

Appendix: 1 table, 2 figures

Abstract

Unlike in laboratory visual search tasks – wherein participants are typically presented with a pictorial representation of the item they are asked to seek out – in real world searches, the observer rarely has veridical knowledge of the visual features that define their target. During “categorical search,” observers look for any instance of a categorically-defined target (e.g., helping a family member look for their mobile phone). In these circumstances, people may not have information about non-critical features (e.g., the phone’s color), and must instead create a broad mental representation using the features that define (or are typical of) the category of objects they are seeking out (e.g., modern phones are typically rectangular and thin). In the current investigation (Experiment 1), using a categorical visual search task, we add to the body of evidence suggesting that categorical templates are effective enough to conduct efficient visual searches. When color information was available (Experiment 1a), attentional guidance, attention restriction, and object identification were enhanced when participants looked for categories with consistent features (e.g., ambulances), relative to categories with more variable features (e.g., sedans). When color information was removed (Experiment 1b), attention benefits disappeared, but object recognition was still better for feature consistent target categories. In Experiment 2, we empirically validated the relative homogeneity of our societally important vehicle stimuli. Taken together, our results are in line with a “category-consistent” view of categorical target templates (Yu, Maxfield, & Zelinsky, 2016), and suggest that when features of a category are consistent and predictable, searchers can create mental representations that allow for the efficient guidance and restriction of attention, as well as swift object identification.

Imagine that a family member asks you to help search for their mobile phone; you've not seen it before, and they don't offer a description. As you look around, you may be expecting to encounter a common "smartphone" (i.e. a flat, rectangular, sleek device). Smartphones are ubiquitous, so it is reasonable to create a mental representation of your target category using the features commonly shared by these devices. However, if the object in question happens to be an older model (e.g., a "flip-phone"), the target might share few perceptual features with your expectation, making your "target template" inexact (Hout & Goldinger, 2015). The older device is unambiguously a mobile phone, but it has features that are now uncommonly encountered (e.g., an antenna, rounder shape, clamshell rather than touchscreen), and so it may prove difficult to guide your attention swiftly to its location. You may even examine it directly and yet be perceptually unprepared to recognize it given its unexpected appearance (see Hout, Walenchok, Goldinger, & Wolfe, 2015). Such a scenario can be contrasted with common laboratory-based search tasks wherein observers are either cued with veridical representations of what they are asked to find (e.g., "find *this* picture of a bicycle"), or are given verbal descriptions of simple stimuli with precisely defined features (e.g., "search for a blue square").

We have learned a great deal about visual search and attention through the use of standard, precisely-cued search tasks (Treisman & Gelade, 1980). However, a gap exists in the literature: Real world searchers are rarely afforded the benefit of having such precise mental representations (Hout & Goldinger, 2015). Search behavior for vaguely defined targets is less well understood than precisely-guided searches. Even when we search for very familiar targets (e.g., our favorite coffee mug), the precise appearance of our target in the world may be difficult to anticipate (e.g., it may be occluded by other objects, its color may have faded, its handle may have broken). Categorical search is an important and exceedingly common behavior (e.g.,

looking for *any* coffee mug, searching for *any* writing utensil), and our intuition about the ease with which categorical search can be conducted belies its complexity. Categorical searches call on all the mental processes required for precisely cued search (e.g., visual working memory, the movement of attention, object recognition), but also require the ability to call to mind (from long-term memory) a mental representation derived from our experience learning, for instance, what is and is not a coffee mug. Here, we conducted a pair of experiments to better understand how mental representations are constructed and acted upon during these tasks.

Is categorical search guided?

During categorical search, viewers attempt to find any object from a cued category without having complete information about the perceptual details of their target. Instead, they must construct a more general target template; one that is presumably comprised of the features that are common to the class of objects they are tasked with finding (Maxfield, Stadler, & Zelinsky, 2014; Robbins & Hout, 2015). To that end, it is possible that during categorical search, observers simply move their attention along passively, fixating on new objects in their environment (and determining if each new viewed item is their target), without using a guiding representation to “push” their attention around in space (Wolfe, et al., 2004). However, recent evidence suggests that, to the contrary, categorical search can be guided. For instance, Yang and Zelinsky (2009) had people search for the category “teddy bears” while having their eye movements tracked; their findings indicated that targets were fixated far sooner than would be expected by chance (i.e., sooner than a no-guidance model would predict). Schmidt and Zelinsky (2009) conducted a similar investigation, but varied the amount of detail specified by a categorical cue (e.g., “footwear” vs “boots” vs “brown boot”), and found that guidance improved as more verbally supplied information was added to the categorical cue.

As evidence has accumulated suggesting that categorical search is indeed guided, researchers have begun to turn their attention to modeling the ways in which observers construct categorical mental representations during search. Yu, Maxfield, and Zelinsky (2016) recently created a generative model of category representation, drawing on methods from computer vision. This “category-consistent features” model captures the (largely) unsupervised way in which people tend to learn the feature structure of various categories. Their model was trained on thousands of images (spanning many dozens of categories), and extracted the features that were common to categories at multiple levels of a hierarchy (e.g., taxis, cars, vehicles). After training, their model successfully captured the eye movement behavior of real participants performing categorical search with cues from varying levels of the hierarchy (e.g., by recreating a subordinate-level advantage for guidance and a basic-level advantage for verification).

A central idea of their model is that searchers create templates using features that are typical of (or consistent with) a category, rather than those that are necessarily helpful in discriminating one category from another. For instance, “small” and “rectangular” are not features that necessarily define mobile phones, as many non-phones are small and rectangular. However, those features are extremely common to this category, and thus using them to guide one’s attention is likely a fruitful way to quickly find targets.

The present investigation: Not all categories are created equal

Here, we address the argument from Yu, Maxfield, and Zelinsky (2016) that category consistent features may be used to construct categorical target templates. If this contention is correct, then search performance should be better for categories that are comprised of objects with consistent or predictable features, relative to categories that are comprised of objects with features that are more heterogeneous or variable. This is because consistent-category objects

comprise features that have been learned by viewers over time in such a way that searchers are better able to predict which features will appear during search. As a consequence, searchers can form a more precise (or accurate) target template for consistent-category objects than inconsistent-category objects (“precise” here refers to the close match between the template and the objects). Put simply, we sought to tap into the inherent differences between different categories to better understand categorical templates in visual search. To do this, we relied on a variety of vehicle categories for stimuli.

Taken together, the goal of the current study was to examine whether more predictable categories yield more precise target templates, by examining search behavior for consistent and variable categories. We had people search for societally important (referred to later as “SI” or simply, “important”) vehicles (e.g., police cars), which, by design, have features that make them readily identifiable and consistent. We compared search performance for these SI vehicles to search behavior for civilian vehicles, which do not need to be distinctive to drivers, and thus are freer to vary in appearance. We hypothesized that attentional guidance (i.e., the ability to push one’s attention around in space), attention restriction (i.e., the ability to hone in on relevant information and avoid inspection of irrelevant information), and object identification (i.e., the ability to perceive and recognize an object) would be better for SI vehicles than civilian, because the more consistent features of SI vehicles would engender more useful categorical target templates. In Experiment 1, we compared search for SI and civilian vehicles while tracking eye movements, finding converging evidence in favor of our hypotheses. In Experiment 2, we verified that superior search performance for SI vehicles was due to the consistent nature of the objects’ features.

Experiment 1

In Experiment 1, we examined search performance for two types of vehicles: Societally important and civilian, and assumed that (by design) SI vehicles (i.e., ambulances) had more predictable features (e.g., color patterns, shape) than civilian vehicles (i.e., sedans). Participants received a word cue and then searched for a matching target. In Experiment 1a, all pictures were presented in color. In Experiment 1b, stimuli were identical, but the images had been converted to grayscale to eliminate color as a guiding feature (see Wolfe & Horowitz, 2017). For brevity, we do not report accuracy results, but they can be found in Appendix Table A1. We report reaction time (RT), and several measures derived from eye-tracking, which provided information about 1) attentional guidance to the targets (and away from distractors), 2) attention restriction (i.e., the extent to which attention was directed only to relevant target categories), and 3) object recognition behaviors (i.e., target recognition and distractor rejection). The various measures we report require some degree of explanation, so rather than elaborate upon them here, we provide detail about the nuances of each measure, our predictions, and our data-cleaning procedures in each of the sub-sections of our results.

Method

Participants. Fifteen students from New Mexico State University participated in the experiment (eight in Experiment 1a, and seven in Experiment 1b). All self-reported normal or corrected-to-normal vision.

Equipment. The experiment was controlled by EPrime vs.2 (Psychology Software Tools, Pittsburgh, PA). The stimuli were presented on a 17" CRT monitor with refresh rate of 60Hz and screen resolution of 1920 x 1200. Monocular eye tracking was recorded at 500Hz using an Eyelink 1000.

Design. Participants were randomly assigned to complete search in color (Experiment 1a) or grayscale (Experiment 1b). After completing 10 practice trials, participants completed two blocks of 80 experimental trials, including a short break between blocks. Selection of targets (defined by vehicle type and viewpoint) randomly varied across trials in equal proportions. Data were analyzed separately for each experiment as a function of the sole within-subjects variable of target vehicle type.

Procedure and Stimuli. Participants took part in the study individually, in a room with floor lamps, rather than overhead fluorescent lighting. Participants used a chin rest during all trials and were initially calibrated (using a nine-point system) to ensure accurate tracking. The chin rest was adjusted so each participant's gaze landed centrally on the computer screen when the participant looked straight ahead. The calibration procedure establishes a map of the participant's known gaze position relative to the tracker's coordinate estimate of that position. The routine proceeded by having participants fixate a black circle as it moved to nine different positions (randomly) on the screen. Calibration was accepted if the mean error was less than 0.5° of visual angle, with no error exceeding 1.0° of visual angle. Periodic recalibrations ensured accurate recording of gaze position throughout the experiment. For analysis purposes, object *interest areas* were defined as the smallest rectangular area that encompassed any given image. An eye movement was classified as a saccade when its distance exceeded 0.5° and its velocity reached $35^\circ/\text{s}$ (or acceleration reached $9,000^\circ/\text{s}^2$). Viewing was binocular, but only the right eye was recorded.

During search, participants were told to locate and identify the target as quickly and accurately as possible. At the start of each trial, a word cue was displayed, signaling the type of vehicle the participant was to search for and from which viewpoint (see Figure 1). This cue

remained on screen until the participant pressed the spacebar, indicating that they were ready to begin search. Next, the participant saw a gaze-contingent fixation cross. When the fixation cross appeared, participants had to maintain gaze on it for 500 ms, which triggered the search display to appear. The participant then searched for the target (which was set in a random location on each trial) in an array containing 20 images, one of which was the target. Targets never appeared at central fixation, so as to require at least one eye-movement to complete search.

The images consisted of one exemplar of each of the six vehicles and three viewpoint (front, back, side) combinations (e.g., ambulance + front, sedan + back). See Figure 2 for examples of our stimuli. There were six exemplars for each combination of vehicle and viewpoint (e.g., six different police cars shown from the front), which were randomly selected for use across trials. Additionally, two road signs (stop and yield) were shown, but were never targets. These items were included so that each quadrant of the display would contain exactly five images, pseudo-randomly arranged to ensure no overlap of the objects. Inclusion of the road signs thus allowed us to create displays with no regions being any more (or less) populous than another. All images were resized (maintaining original proportions) to a maximum of 150x150 pixels. Our “societally important” vehicles included ambulances, school buses, and police cars. “Civilian” vehicles included sedans, sport-utility vehicles (SUVs), and delivery trucks.

In our analyses, we distinguish between “relevant” and “irrelevant” distractors, defined relative to the target being searched for. “Relevant” distractors were those that matched the target category, but not orientation (e.g., when the target was “police+front,” relevant distractors were police cars shown from the side or back). “Irrelevant” distractors were those that did not

match the category of the target (e.g., an ambulance or sedan when the target was a police car).

In each trial, therefore, there were 2 relevant distractors and 15 irrelevant distractors.

Once participants located the target, they pressed the spacebar. The items in the search array were then replaced by numbers that remained on screen for two seconds in the same spatial locations as the items. Participants were told to remember the number that corresponded to the spatial location of the target item. They were then shown a two-alternative forced choice, and indicated which of the two numbers corresponded to the target's location (see Hout & Goldinger, 2010, 2012 for similar display creation and search confirmation procedures).

Results

We report results only for trials in which a correct response was made (94% and 95% of trials in Experiments 1a and 1b, respectively). Results are reported for each experiment in turn. RTs are first presented, followed by eye-tracking derived measures, grouped by attentional guidance, attention restriction, and object recognition. Unless otherwise noted, dependent variables were analyzed using a repeated measures ANOVA (equivalent to a two-tailed paired-samples t-test), with a single factor of target vehicle type (important, civilian).

Reaction time. RT was measured from display onset to spacebar press (see Figure 3). In Experiment 1a, we found an effect of target vehicle type, $F(1, 7) = 46.876, p < .001, \eta_p^2 = .870$, with faster RTs for important vehicles (2496 ms), relative to civilian (3431 ms). In Experiment 1b, we found a marginally significant effect, $F(1, 6) = 4.14, p = .088, \eta_p^2 = .408$, with faster RTs for important vehicles (3085 ms), relative to civilian (3421 ms).

Attentional guidance. For attentional guidance, we began by examining time to fixate, measured from trial onset to the time the participant first fixated the target (see Figure 4). In Experiment 1a, we found an effect of target type, $F(1, 7) = 37.304, p < .001, \eta_p^2 = .842$, with

faster first fixation times for important vehicles (1435 ms), relative to civilian (2121 ms). In Experiment 1b, the effect was not significant, $F(1, 6) = 1.280, p = .301$ (2104 ms, and 2278ms for important and civilian targets, respectively).

We next report the number of distractor items visited (i.e., visually inspected) prior to target fixation. “Visits” are similar, but not identical, to an eye fixation. We defined a visit as a period of time in which the eyes remained within the object’s interest area, regardless of whether that visit involved a single fixation, or a primary fixation followed by additional fixations that did not leave the interest area. This analysis was performed using a 2 (target type; important, civilian) x 2 (distractor relevance; relevant, irrelevant) repeated measures ANOVA. As a reminder, relevant distractors were those that matched the target category but not orientation, and irrelevant distractors were those that did not match the category of the target.

In Experiment 1a, we found a main effect of target type, $F(1, 7) = 25.442, p = .001, \eta_p^2 = .784$, with fewer distractors being visited prior to fixation of important vehicle targets (2.23), relative to civilian ones (3.05). There was also a main effect of distractor relevance, $F(1, 7) = 433.425, p < .001, \eta_p^2 = .984$, with fewer visits to relevant distractors (0.93), relative to irrelevant distractors (4.35); bear in mind, however, that only two relevant distractors were present, and 15 irrelevant distractors were present. Importantly, there was also a two-way interaction, $F(1, 7) = 45.023, p < .001, \eta_p^2 = .865$, indicating that relevant targets were fixated with equal frequency regardless of target type, but that irrelevant distractors were visited less frequently when looking for important targets (verified by a two-tailed paired-samples t-test; $t(7) = 5.999, p = .001$). This analysis tells us that people were equally good at directing their attention to the relevant target categories, regardless of whether they were searching for important or civilian targets, but that when people were looking for important vehicles, they looked at fewer irrelevant distractors,

relative to when they were looking for civilian targets (we examine attention restriction in greater detail, below). In Experiment 1b, there was no main effect of target type, $F(1, 6) = 0.160$, $p = .703$. There was a main effect of distractor relevance, $F(1, 6) = 357.402$, $p < .001$, $\eta_p^2 = .983$, indicating again that fewer relevant distractors (0.88) were visited, relative to irrelevant distractors (5.34). The two-way interaction was not significant, $F(1, 6) = 0.581$, $p = .475$. This analysis tells us that people looked at fewer relevant distractors (relative to irrelevant ones), but that finding is more or less inevitable, again, because only two relevant distractors were present, and 15 irrelevant distractors were present. More importantly, the lack of an interaction effect tells us that the distribution of visits was no different depending on whether the searcher was looking for an important or civilian target.

It is important to also consider globally (i.e., regardless of distractor relevance) whether or not attentional guidance was significantly better than would be expected by chance. To do this, we added up the total number of visits on distractor vehicles, prior to fixation of the target. If there is significant attentional guidance to the target, people should visit fewer than 50% of the set size, prior to fixating the target. To be conservative, we ignored the road signs (i.e., we assumed they were so unlike the vehicles that they could easily be ignored), and considered half of the visual set size to be nine items, rather than ten. We then performed one-sample t-tests, comparing the number of visits made to nine. A significant t-test, therefore, would indicate the presence of attentional guidance, and non-significant t-tests (or mean number of visits significantly greater than nine) would indicate a lack of attentional guidance.

In Experiment 1a, the mean total number of visits (prior to fixation of the target) when looking for important targets was 4.45 (SD = .46), which was significantly less than nine; $t(7) = -27.72$, $p < .001$ (two-tailed). When looking for civilian targets, the mean was 6.11 (SD = .83),

which was also significantly less than nine; $t(7) = -9.85, p < .001$ (two-tailed). In Experiment 1b, the mean total number of visits when looking for important targets was 6.13 ($SD = .99$), which was significantly less than nine; $t(6) = -7.68, p < .001$ (two-tailed). When looking for civilian targets, the mean was 6.30 ($SD = .71$), which was also significantly less than nine; $t(6) = -10.12, p < .001$ (two-tailed). Thus, in both experiments, attentional guidance to the target was significantly better than predicted by chance in all cases (i.e., when looking for societally important targets and when looking for civilian targets).

Attention restriction. For attention restriction, we report two measures that afford a more nuanced picture than the number of visits alone can provide. First, we report the proportion of items in the display that were viewed on each trial. This analysis is different from the analyses examining the number of visits in that the visit count could include a single item having been visited multiple times. Reporting the proportion of items viewed is therefore a measure of attention restriction rather than attentional guidance (e.g., an item viewed twice before the target would count as two visits, but would only be counted as 1/17th of the total distractors vehicles present). It should also be noted that unlike the visit count, this analysis also allows for items having been viewed *after* the target was viewed, which did happen on occasion (see the object recognition analyses for more detailed discussion of targets requiring more than a single visit to be recognized). As with the number of visits, here we used 2 (target type; important, civilian) x 2 (distractor relevance; relevant, irrelevant) repeated measures ANOVAs. For simplicity, we limited our analyses to vehicle distractors and ignored the two traffic sign distractors, as we had no *a priori* reason to suppose that they were more (or less) similar to any particular vehicle category.

In Experiment 1a, there was no main effect of target type, $F(1, 7) = 2.912, p = .132$ (.46 and .52 of the total distractors were viewed for important and civilian vehicles, respectively). There was a main effect of distractor relevance, $F(1, 7) = 158.191, p < .001, \eta_p^2 = .958$, indicating that relevant distractors (.60) were more likely to be viewed than irrelevant ones (.38). Importantly, the two-way interaction was significant, $F(1, 7) = 36.381, p = .001, \eta_p^2 = .839$. Two-tailed paired-samples t-tests showed that target type did not affect the likelihood of viewing relevant targets ($t(7) = 1.284, p = .240$), but that irrelevant items were less likely to be examined when searching for important vehicles, relative to civilian vehicles ($t(7) = 5.635, p = .001$). In Experiment 1b, there was no main effect of target type, $F(1, 6) = 2.223, p = .187$ (.47 and .51 of the total distractors were viewed for important and civilian vehicles, respectively). There was a main effect of distractor relevance, $F(1, 6) = 32.571, p = .001, \eta_p^2 = .844$, with relevant distractors (.55) being more likely to be viewed, relative to irrelevant ones (.44). The two-way interaction was not significant, $F(1, 6) = 1.861, p = .221$. Visual inspection of the means suggests that when searching for important targets, people were less likely to view irrelevant distractors than when searching for civilian targets, but this difference was only marginally significant (two-tailed paired-samples t-test, $t(6) = 2.203, p = .070$), and thus should be interpreted with caution. See Figure 5.

Though not as pure a measure of attention restriction as the proportion of items viewed, we next report the frequency with which irrelevant distractors were visited/revisited, provided they were looked upon at least once. This gives us a sense of how often attention was poorly restricted, giving rise to unhelpful features needing to be reexamined before the target was found. Here, we used a two-way chi-square analysis, with target type (important, civilian) and the number of visits (1-5) as factors. In Experiment 1a, we found a significant chi-square, $\chi^2(4, N =$

3049) = 21.559, $p < .001$, indicating that the distribution of the number of visits was not equivalent when searching for important versus civilian targets. Instead, multiple visits were more common when searching for civilian targets, relative to important ones. In Experiment 1b, we also found a significant chi-square, $\chi^2(4, N = 3164) = 34.694$, $p < .001$, indicating non-equivalent distributions. Once more, multiple visits were more common when targets were civilian than when they were important.

Object recognition. For target object recognition, we first report decision time, measured from the time the target was fixated to the depression of the spacebar. Here, we restricted our analyses to trials in which the target was examined only once or twice. Previous research has shown that a large proportion of revisitations are directed at search targets. Participants seem to plan more than one saccade at a time; they often fixate the target, move on to another item, and then quickly return having realized the target was seen (Peterson et al., 2001). Restricting decision time analyses to these trials alone allowed us to avoid contamination of our results by trials in which the target was viewed but not recognized, and was only appreciated after a subsequent visit with multiple intervening fixations between the first and last view upon it (which would artificially inflate the decision time). To confirm that our targets were commonly recognized after two or fewer visits, see Appendix Figure A1. It shows that 88% and 91% of targets were recognized with 1-2 visits in Experiment 1a and 1b, respectively. Chi-square analyses (like those performed for the distribution of irrelevant distractor visits) further confirmed that the distribution of the number of target visits was largely equivalent between important and civilian targets. In Experiment 1a, the analysis was marginally significant, $\chi^2(4, N = 519) = 9.082$, $p = .059$, showing that civilian targets tended slightly to

require more visits, relative to important targets. In Experiment 1b, frequency distributions were equivalent between important and civilian targets, $\chi^2(4, N = 468) = 3.660, p = .454$.

In Experiment 1a, we found an effect of target type, $F(1, 7) = 9.281, p = .019, \eta_p^2 = .570$, with faster decision times on important targets (1066 ms), relative to civilian targets (1326 ms). In Experiment 1b, there was no effect of target type, $F(1, 6) = 2.251, p = .184$ (decision time was 974 ms and 1199 ms for important and civilian targets, respectively). See Figure 6.

We next report distractor recognition (i.e., rejection) time as the average amount of time that the distractors were dwelled upon, per visit. As before, we used 2 (target type; important, civilian) \times 2 (distractor relevance; relevant, irrelevant) repeated measures ANOVAs. In Experiment 1a, we found a main effect of target type, $F(1, 7) = 14.935, p = .006, \eta_p^2 = .681$, indicating that distractors were rejected more quickly when searching for important targets (293 ms), relative to civilian targets (336 ms). We also found a main effect of distractor relevance, $F(1, 7) = 21.724, p = .002, \eta_p^2 = .756$, indicating that irrelevant distractors (272 ms) were more easily dismissed than relevant ones (358 ms). The two-way interaction was not significant, $F(1, 7) = 0.298, p = .602$. In Experiment 1b, we again found a main effect of target type, $F(1, 6) = 7.913, p = .031, \eta_p^2 = .569$, indicating faster distractor rejection when searching for important targets (303 ms), relative to civilian targets (340 ms). There was also a main effect of distractor relevance, $F(1, 6) = 62.893, p < .001, \eta_p^2 = .913$, showing that irrelevant distractors (282 ms) were dismissed more quickly than relevant ones (361 ms). The two-way interaction was not significant, $F(1, 6) = 0.044, p = .841$.

Discussion

The primary goal of Experiment 1 was to examine whether more predictable categories yield more useful target templates than categories that are more variable in nature. We compared

search performance for predictable categories (SI vehicles) to search performance for less predictable categories (civilian vehicles).

To quickly summarize our findings, we found that when searchers could use color information (Experiment 1a), mental representations for SI vehicles afforded faster RTs, better attentional guidance, enhanced attention restriction, and superior object recognition (relative to search for civilian vehicles). When the color feature was removed (in Experiment 1b), however, many of these effects disappeared. What remained was a marginally significant benefit for SI vehicles on RTs, the finding that more revisits occur to civilian vehicles, and crucially, swifter rejection of distractors when searching for SI vehicles. Put simply, predictable categories yield seemingly universal benefits when color can be used as a distinguishing feature. When this important feature is unavailable, however, mental representations for predictable categories still afford a benefit, but it is primarily in the speeding up of non-target rejection.

Another aspect of our data warrants brief commentary. It may seem counter-intuitive that relevant distractors were visited less frequently than irrelevant items, prior to fixation of the target. However, it is important to remember that that finding is mathematically inevitable, because there was a greatly uneven distribution of relevant to irrelevant distractors (i.e., two relevant distractors and 15 irrelevant ones, plus the two road signs). The important thing to note about that analysis is that, in Experiment 1a, irrelevant visits were less frequent when looking for SI vehicles. And, that in both experiments, we found evidence that categorical guidance was better than would be expected by chance. If attentional guidance was altogether absent, we would expect half of the vehicles (i.e., 9) to be visited prior to fixating the target, and Figure 4 clearly shows that the average number of visits was no greater than seven (a contention confirmed by the one-sample t-tests). This suggests indeed that categorical templates are capable

of guiding attention, in keeping with prior research (e.g., Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009).

Clearly, color is a useful feature, and it is not at all surprising that the use of color enhanced attentional guidance (see, for instance, Itti & Koch, 2000; 2001). But it is important to note that both SI and civilian vehicles were composed of a range of colors, and that participants were never given explicit information about what colors to expect or search for. Presentation of a categorical cue, it would seem, is enough to give rise to a target template that incorporates (from learned expectations about the likely colors of, for instance, a police car) color information. What is perhaps more interesting is that while color information was particularly useful for the guidance and restriction of attention, other features (e.g., shape) were particularly useful for deciding when incoming visual information did not match the to-be-searched for category (we return to this point in the General Discussion). For instance, knowledge that police vehicles often have lights mounted on the roof (impacting their shape), that ambulances tend to have alternating red and white patterns (which would create contrast even when images were grayscale), and that school buses have not just a distinct color but a distinct shape, seems useful enough to aid in distractor rejection even when color information is unavailable. Taken together, our findings replicate those of Yu, Maxfield, and Zelinsky (2016) that category consistent features can be used to construct more beneficial target templates. However, we extend their findings by showing that not all features are equally useful (e.g., color information is particularly helpful for attentional guidance), and that even when seemingly crucial features like color are not available, secondary features like shape can be used to enhance object recognition behaviors.

Experiment 2

In Experiment 1, we found evidence that more predictable categories yield better search performance than less predictable ones. However, a critical assumption in Experiment 1 was that the societally important vehicles we chose were indeed more predictable than our civilian vehicles, and that assumption was not empirically tested.

If SI vehicle features are more consistent, then they are by definition, a more homogeneous group than civilian vehicles. Search for specific exemplars should therefore be hindered by the presence of similar (i.e., category-matched) distractors more so for SI vehicles than civilian ones (see Duncan & Humphreys, 1989). The purpose of Experiment 2 thus was simply to verify the homogeneity (consistency) of SI vehicle features, and put the critical assumption made in Experiment 1 to the test. Experiment 2 was nearly identical to Experiment 1, except participants were shown picture cues and were asked to search for a specific target among distractors selected from the same vehicle category.

Method

Participants. Twenty-one new participants from New Mexico State University took part in the study. One participant was removed from analysis due to a technical problem that prevented the data from being recorded correctly.

Design & Procedure. The design, procedure, and stimuli were identical to Experiment 1, except that picture cues were used (in place of word cues), and non-targets now belonged to the same category as the target. Vehicle type was again randomly varied across trials (in equal proportions). All stimuli were shown in color; there was no grayscale condition.

Results

Again we report results only for trials in which a correct response was made (96% of trials; see Appendix Table A1 for group means). Results were analyzed in similar fashion to

Experiment 1, using a repeated measures ANOVA with a single factor of target vehicle type (important, civilian), unless otherwise noted. Prior two-way ANOVAs with distractor relevance as a factor were no longer applicable, as all distractors were now relevant. We again report RTs, attentional guidance (time to fixate, the number of visits prior to target fixation), attention restriction (the proportion of items viewed, distractor revisits), and object recognition behaviors (target decision time, and distractor dwell time).

If SI vehicles are more consistent/homogeneous than civilian vehicles, then the pattern of effects we observed in Experiment 1 should flip, yielding better performance to civilian vehicles instead. Plots of all DVs (with the exception of distractor visits) can be found in Figure 7; see Figure 8 for distractor visits.

Reaction time. We found an effect of target type, $F(1, 19) = 188.206, p < .001, \eta_p^2 = .908$, with faster RTs to civilians vehicles (1795 ms), relative to important ones (2928 ms).

Attentional guidance. For time to fixate, we found an effect of target type, $F(1, 19) = 103.400, p < .001, \eta_p^2 = .845$, with faster first fixation times to civilian vehicles (1000 ms) than important ones (1737 ms). For the number of distractor visits, we also found an effect of target type, $F(1, 19) = 89.752, p < .001, \eta_p^2 = .825$, with fewer visits when searching for civilian targets (3.719), relative to important ones (5.59). As in Experiment 1, we performed one-sample t-tests to determine whether or not attentional guidance was significantly better than expected by chance. The mean total number of visits (prior to fixation of the target) when looking for important targets was 5.59 (SD = .88), which was significantly less than nine; $t(19) = -17.41, p < .001$ (two-tailed). When looking for civilian targets, the mean was 3.71 (SD = .70), which was also significantly less than nine; $t(19) = -33.71, p < .001$ (two-tailed). Therefore, significant

attentional guidance was present when looking for both types of targets (which in this case, is not terribly surprising, given the pictorial nature of the target cue).

Attention restriction. For the proportion of distractors viewed, we found an effect of target type, $F(1, 19) = 160.112, p < .001, \eta_p^2 = .894$, with fewer distractors viewed for civilian (.23), relative to important vehicles (.37). As in Experiment 1, for the number of distractor visits, we used a chi-square analysis to determine if the distribution of visits was the same for important vehicles as it was for civilian vehicles. We found a significant chi-square, $\chi^2(4, N = 7911) = 85.201, p < .001$, indicating that the distribution of the number of visits was not equivalent when searching for important versus civilian targets. Here, multiple visits were more common when searching for important targets, relative to civilian ones.

Object recognition. As in Experiment 1, we wanted to confirm that our targets were usually recognized after two or fewer visits. See Appendix Figure A2, which shows that 95% of targets were recognized with two or fewer visits. A significant chi-square analysis showed that the distribution of the number of target visits was uneven between important and civilian targets, $\chi^2(4, N = 1342) = 36.631, p < .001$. Here, multiple visits were more common when searching for important vehicles, relative to civilian ones.

More importantly, for target decision times, we found an effect of target type, $F(1, 19) = 28.104, p < .001, \eta_p^2 = .597$, with faster recognition of civilian targets (696 ms), relative to important ones (1037 ms). We also found an effect of target type on distractor dwell times, $F(1, 19) = 73.408, p < .001, \eta_p^2 = .794$, with shorter dwell times on civilian distractors (203 ms), relative to important ones (225 ms).

Discussion

Participants in Experiment 2 were afforded the benefit of having perfectly precise target representations, in that picture cues perfectly matched to-be-located information. Despite this template precision, search for SI vehicles now suffered universally, relative to search for civilian vehicles. RTs, attentional guidance, attention restriction, and object recognition behavior were all now superior when searching for civilian vehicles. These findings strongly suggest that our contention (from Experiment 1) that SI vehicles are more homogeneous is indeed a correct assumption, making individual SI exemplars more difficult to find among category-matched distractors.

General Discussion

Categorical search is an important and exceedingly common activity (e.g., search for any car) that, despite its subjective ease, is not well explained by current models of search. Previous research on categorical templates posited that a general template may be constructed from typical features (Maxfield, Stalder & Zelinsky, 2014; Robbins & Hout, 2015), perhaps akin to a category prototype rather than a specific exemplar (Yang & Zelinsky, 2009). More recently, it has been argued that successful categorical search takes advantage of category-consistent features (Yu, Maxfield, & Zelinsky, 2016): That is, features that are shared across many objects within a category. However, as the category-consistent features idea is quite new, there has been little direct behavioral evidence in support of this notion.

Here, our goal was to see if we could find such evidence by examining whether more predictable categories yielded more precise (and therefore more useful) target templates. We compared performance for societally important and civilian vehicles. The SI vehicles formed a more perceptually homogenous category than the civilian vehicles (evidenced by Experiment 2), and so can be presumed to have a greater number of category-consistent features. We found

universally improved search performance for the societally important vehicles in Experiment 1a (and better distractor rejection in Experiment 1b), when distractors were from other vehicle categories, and a complete reversal of performance in Experiment 2, wherein distractors were sampled from the same vehicle category as the target.

The findings from Experiment 1b – that without the feature of color, SI vehicles only lead to better distractor rejection – should not be overlooked. Object recognition is another aspect of visual search that seems to be understudied, and is often discarded as a bit of a “black box” in the stream of search processes. Indeed, object recognition is sometimes discussed as if it happens instantaneously, or rather that object recognition processes are largely the same, regardless of the constraints of the search task.

However, object recognition can vary greatly, depending on task demands. For instance, Godwin et al. (2015) found that in multiple-target search, complex distractors are rejected more efficiently than simpler ones, because complex objects afford a greater number of features that can be used to reject a non-target (i.e., non-targets only have to have a single “wrong” feature in order to be classified as such, and complexity gives rise to a wider range of such features to choose from). Walenchok, Hout, and Goldinger (2016) found that the phonological similarity between the names of potential targets and distractors can create interference during search, which primarily manifests as difficulty in rejecting distractors when the name of the target(s) and distractors share a phonological onset (e.g., a distractor *bee* when the target is a *beachball*). And Hout and Goldinger (2015) found that – by manipulating the similarity between the cue and the actual target that was encountered – more precise target templates lead to faster target recognition times, relative to when templates are less precise (see also Schmidt & Zelinsky, *in press*, for evidence that people can flexibly add details to their target representation).

Here, our findings show that when categories are comprised of consistent features, it becomes easier to quickly disregard non-targets as an inappropriate match to what the searcher intends to find. This has important ramifications, in that category consistency seems not only to affect the first aspect of search, attentional guidance to the target, but also its partner process, the recognition of targets and rejection of distractors.

Taken together, our results strongly suggest that during categorical search, observers can create more precise mental representations from categories that contain more consistent features. Because various categories share differing numbers of features, this means that not all categories are equal in terms of how precise the templates based upon those categories can be. As discussed, template precision is known to be an important driver of search performance, reducing RTs by enabling searchers to more rapidly fixate and identify targets (Hout & Goldinger, 2015). In the current study, we found that categories with more consistent features (SI vehicles) allowed participants to develop a precise template that more effectively guided their attention to the target, and helped avoid non-target features. Precise templates also permitted targets to be swiftly identified once attention landed on them, and allowed searchers to reject non-targets more effectively when they happened to be viewed.

It is worth noting that our findings should be replicated using other stimuli. As with almost any experiment, it's possible that the effects we observed were driven by stimulus idiosyncrasies (in this case related to the particular categories we selected). This seems unlikely, given the range of categories we used, given the number of exemplars for each category and viewpoint combination, and given that Experiment 2 was designed precisely to address the featural variability of the categories and to check that our assumptions were correct. Nonetheless, to be safe, future work should explore different categories of objects, and perhaps

employ a wider range of categories than used in the current study. A particularly interesting future direction may even be to create artificial stimuli in which categories are created with more (or fewer) consistent features, and then (following training of the new stimuli) have participants search for these artificial items.

In sum, our results provide preliminary behavioral evidence supporting the recent modelling approaches to understanding categorical target templates (Yu, Maxfield, & Zelinsky, 2016), which posit that mental representations are constructed using features that are common to a category, rather than those that necessarily define it. Our findings also go an important step further by demonstrating that the degree of featural variation of these category-consistent features has a direct impact on target template precision during search.

REFERENCES

- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96(3), 433.
- Godwin, H. J., Walenchok, S., Hout, J. W., Hout, M. C., & Goldinger, S. D. (2015). Faster than the speed of rejection: Object identification processes during visual search for multiple targets. *Journal of Experimental Psychology: Human Perception & Performance*, 41, 1007-1020. doi: 10.1037/xhp0000036
- Hout, M. C., Godwin, H. J., Fitzsimmons, G., Robbins, A., Menneer, T., & Goldinger, S. D. (2015). Using multidimensional scaling to quantify similarity in visual search and beyond. *Attention, Perception, & Psychophysics*, 78, 3-20. doi: 10.3758/s13414-015-1010-6.
- Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception & Psychophysics*, 72, 1267-1282. doi: 10.3758/APP.72.5.1267.
- Hout, M. C., & Goldinger, S. D. (2012). Incidental learning speeds visual search by lowering response thresholds, not by improving efficiency: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, 38, 90-112. doi: 10.1037/a0023894.
- Hout, M. C., & Goldinger, S. D. (2015). Target templates: the precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics*, 77(1), 128-149. doi: 10.3758/s13414-014-0764-6.
- Hout, M. C., Walenchok, S. C., Goldinger, S. D., & Wolfe, J. M. (2015). Failures of perception in the low-prevalence effect: Evidence from active and passive visual search. *Journal of*

- Experimental Psychology: Human Perception & Performance*, 41, 977-994. doi: 10.1037/xhp0000053.
- Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research*, 40(10), 1489-1506.
- Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature reviews neuroscience*, 2(3), 194-203.
- Maxfield, J. T., Stadler, W.D., & Zelinsky, G. J. (2014). Effects of target typicality on categorical search. *Journal of Vision*, 14, 1. doi:10.1167/14.12.1.
- Peterson, M. S., Kramer, A. F., Wang, R. F., Irwin, D. E., & McCarley, J. S. (2001). Visual search has memory. *Psychological Science*, 4, 287-292.
- Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from <http://www.pstnet.com>.
- Robbins, A., & Hout, M. C. (2015). Categorical templates: Typical category members are found and identified quickly during word-cued search. *Summary published in Object Perception, Attention, and Memory (OPAM) 2015 Conference Report, Visual Cognition*, 23, 817-821. doi: 10.1080/13506285.2015.1093247.
- Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *The Quarterly Journal of Experimental Psychology*, 62, 1904–1914.
- Schmidt & Zelinsky, G. J. (*in press*). Adding details to the attentional template offsets search difficulty: Evidence from contralateral delay activity. *Journal of Experimental Psychology: Human Perception and Performance*.
- Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97-136. doi: 10.1016/0010-0285(80)90005-5.

- Walenchok, S. C., Hout, M. C., & Goldinger, S. D. (2016). Implicit object naming in visual search: Evidence from phonological competition. *Attention, Perception, & Psychophysics*, 78, 2633–2654. doi: 10.3758/s13414-016-1184-6
- 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. doi: 10.1037/0096-1523.15.3.419.
- Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature: Human Behavior*, 1. doi: 10.1038/s41562-017-0058.
- Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision research*, 44, 1411-1426.
- Zelinsky, G. J. (2008). A theory of eye movements during target acquisition. *Psychological Review*, 115, 787–835.
- Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision research*, 49, 2095-2103.
- Yu, C., Maxfield, J. & Zelinsky, G. (2016; *online first edition*). Searching for category-consistent features: A computational approach to understanding visual category representation. *Psychological Science*, doi: 10.1177/0956797616640237.

FIGURES

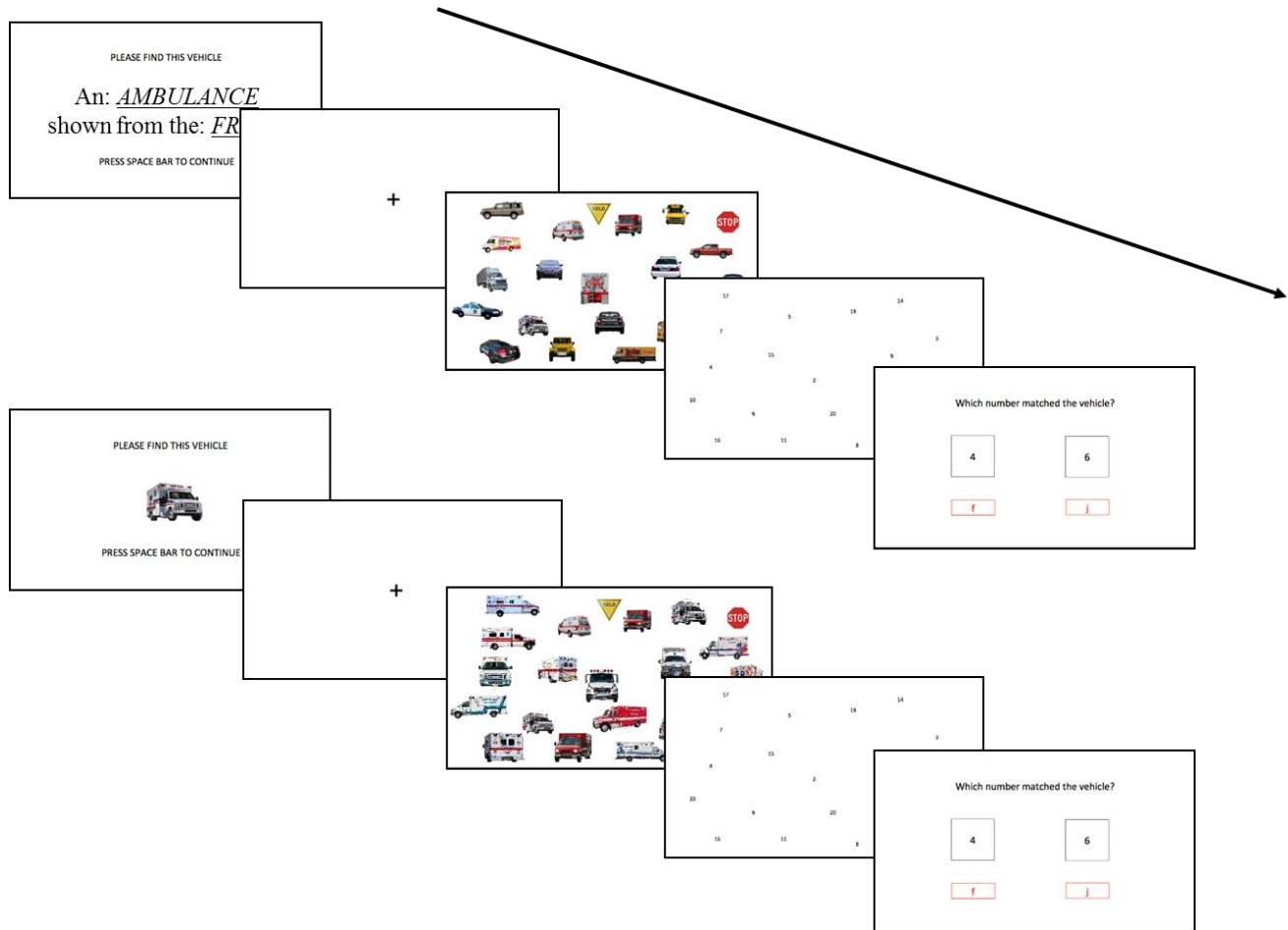


Figure 1. Progression of events in a trial for Experiments 1 (top) and 2 (bottom). Stimuli in Experiment 1a were shown in full color (as depicted), and were shown in grayscale in Experiment 1b.







Societally Important Vehicles		Civilian Vehicles	
Police Car		Sedan	
School Bus		Delivery Truck	
Ambulance		SUV	

Figure 2. Sample stimuli, showing exemplars from each of the three societally important categories, and each of the three civilian categories.

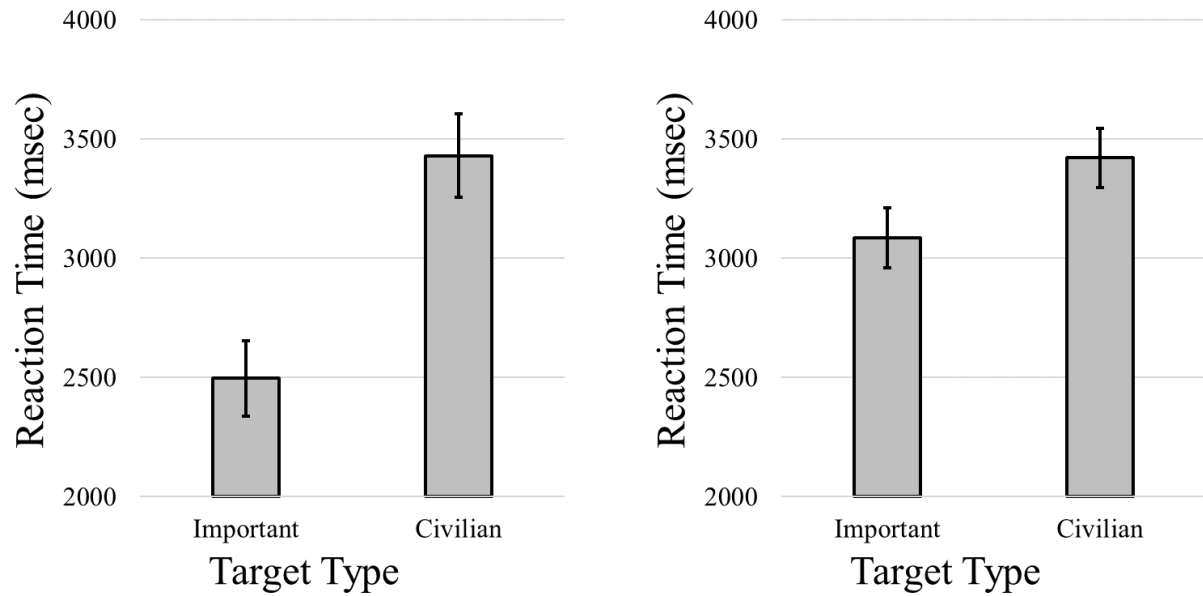


Figure 3. RTs from Experiments 1a (left) and 1b (right), presented as a function of target type (important, civilian). Error bars represent one standard error of the mean (SEM).

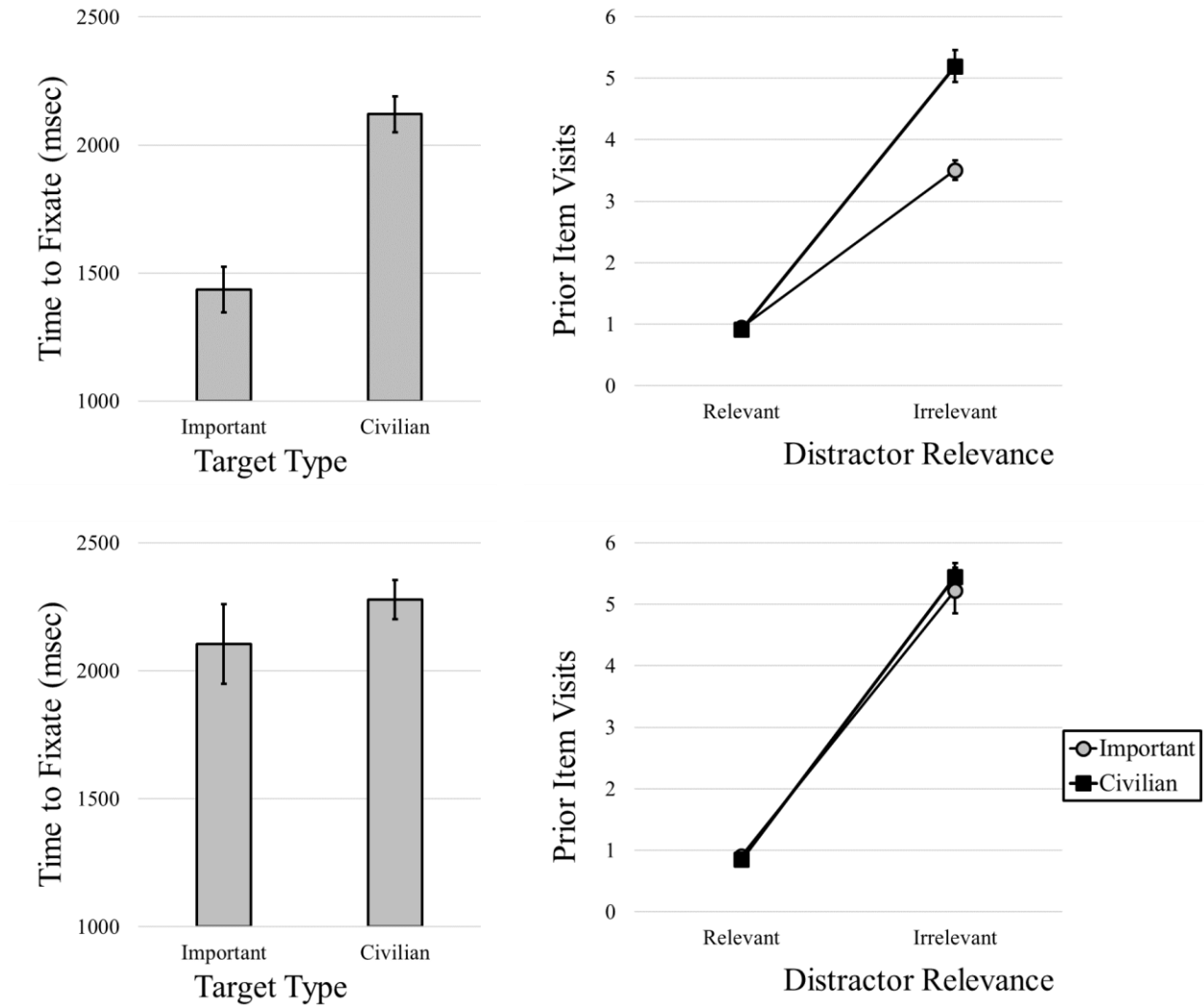


Figure 4. Attentional guidance measures from Experiments 1a (top) and 1b (bottom). Left panels show time to fixate the target as a function of target type. Right panels show the number of distractor items visited prior to fixation of the target; results are presented as a function of target type (gray circles and black squares for important and civilian targets, respectively) and distractor relevance. Error bars represent one standard error of the mean (SEM).

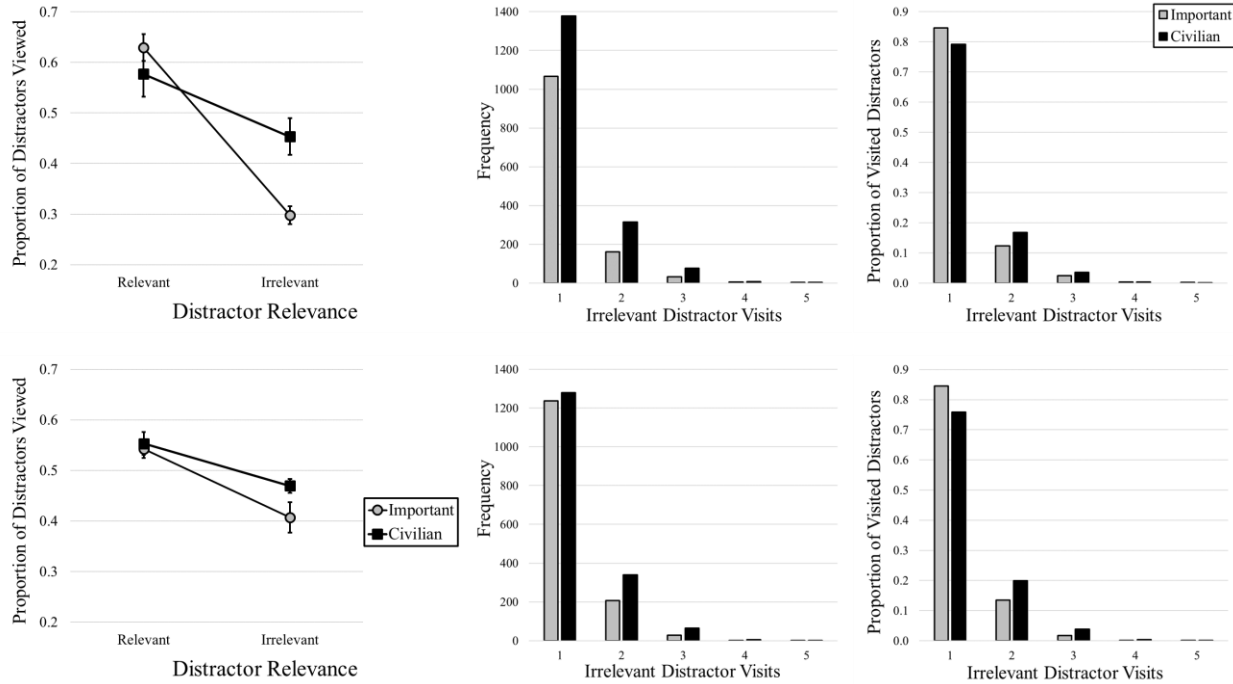


Figure 5. Attention restriction measures from Experiments 1a (top) and 1b (bottom). Left panels show the proportion of distractors viewed, presented as a function of distractor relevance and target type (gray circles and black squares for important and civilian targets, respectively). Error bars represent one standard error of the mean (SEM). Center and right panels display the frequency of irrelevant distractor visits. Gray bars show results for societally important targets, and black bars show civilian targets. Center panels show raw frequency data. Frequencies differed across experiments and conditions due to differences in the number of participants and accurate search trials. To make the conditions more directly comparable, right panels show raw frequency data converted to a proportion of total trials per condition.

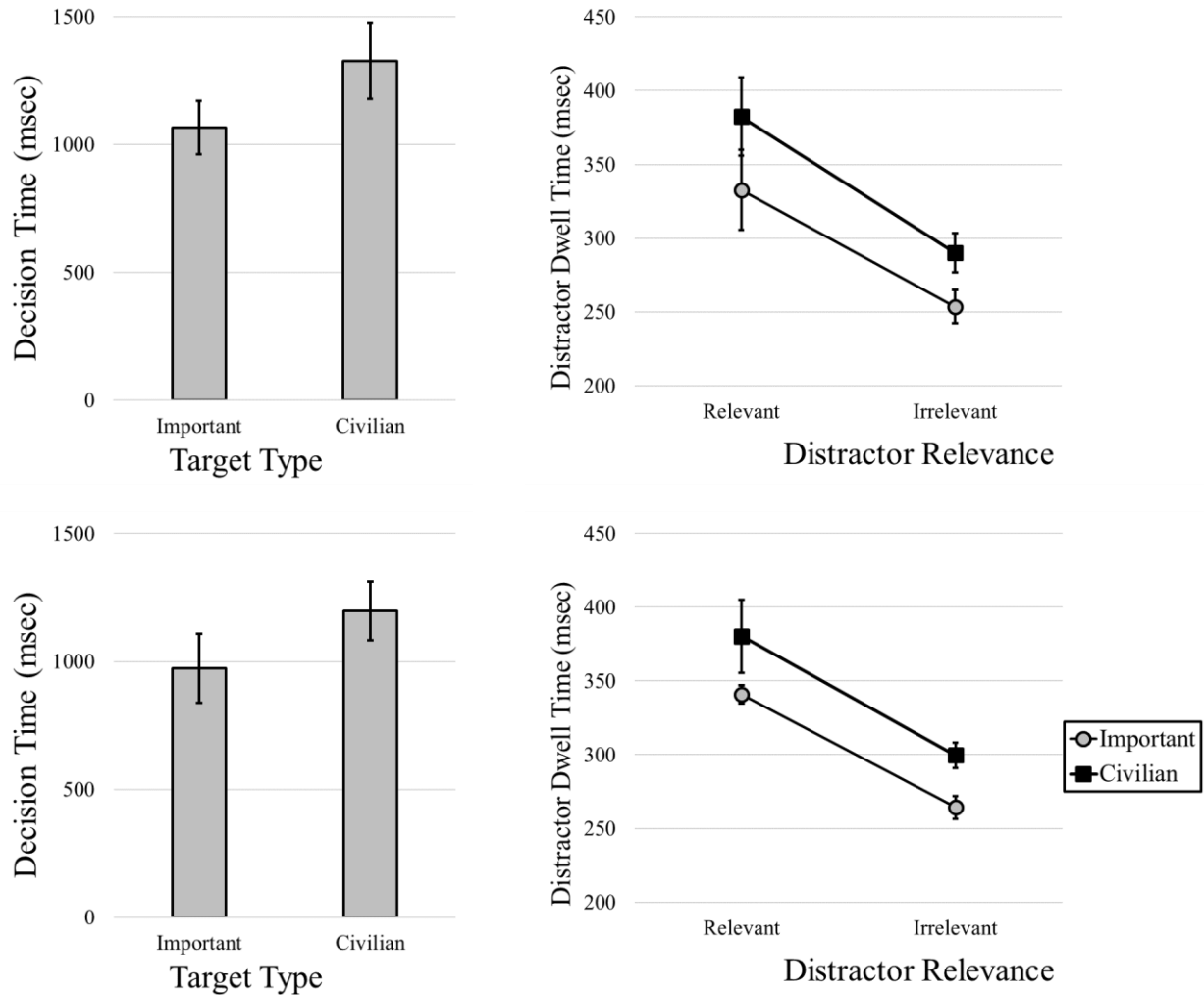


Figure 6. Object recognition measures from Experiments 1a (top) and 1b (bottom). Left panels show decision time (i.e., target recognition) as a function of target type. Right panels show the distractor dwell time (i.e., distractor rejection); results are presented as a function of target type (gray circles and black squares for important and civilian targets, respectively) and distractor relevance. Error bars represent one standard error of the mean (SEM).

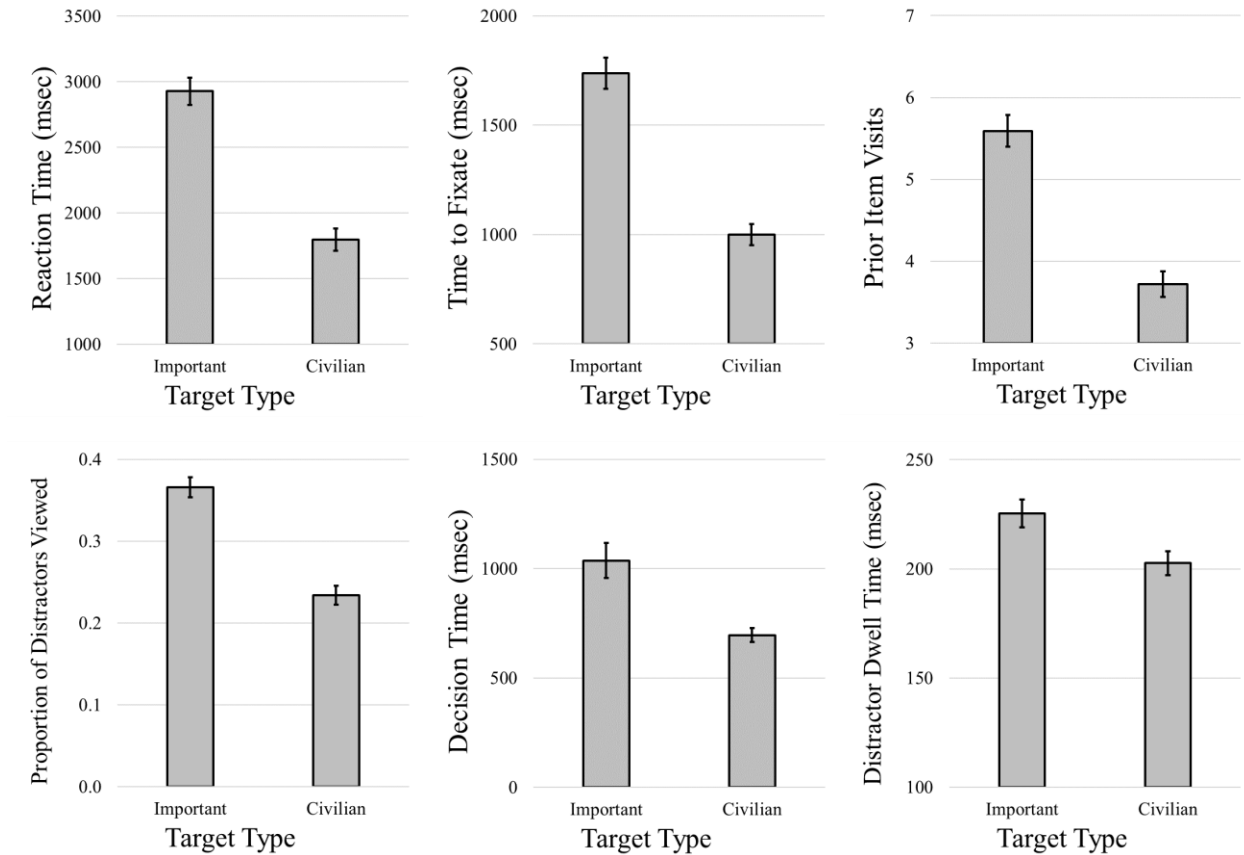


Figure 7. Various DVs from Experiment 2, plotted as a function of target vehicle type. On the top row, from left to right: Reaction time, time to fixate, and prior item visits. On the bottom row, from left to right: proportion of distractors viewed, decision time, and distractor dwell time. Error bars represent one standard error of the mean (SEM).

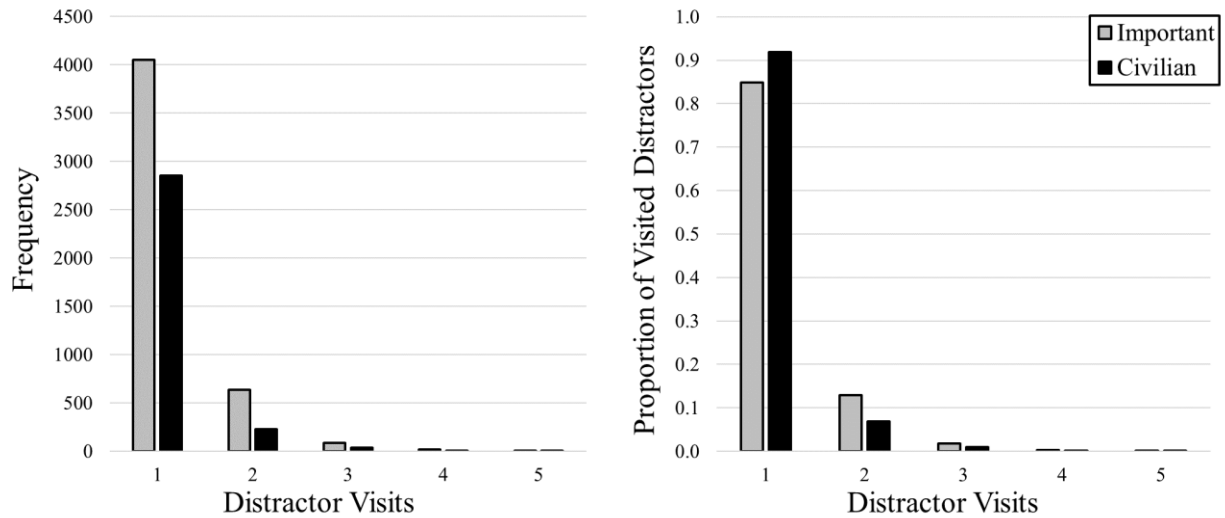


Figure 8. Frequency of irrelevant distractor visits, from Experiment 2. Gray bars show results for societally important targets, and black bars show civilian targets. The left panel shows raw frequency data. Frequencies differed across conditions due to differences in the number of accurate search trials. To make the conditions more directly comparable, the right panel shows raw frequency data converted to a proportion of total trials per condition.

APPENDIX

Table A1

Mean visual search accuracy (and standard error of the mean), presented by target type, from Experiments 1 and 2. Also reported are the results of one-way ANOVAs, examining search accuracy as a function of target type.

Experiment	Target Type	Search Accuracy					
		Mean	SE	<i>df</i>	<i>F</i>	<i>p</i>	η_p^2
1a	Societally Important	0.986	0.010	1, 7	12.600	0.009	0.643
	Civilian	0.903	0.029				
1b	Societally Important	0.956	0.019	1, 6	0.300	0.604	0.048
	Civilian	0.952	0.014				
2	Societally Important	0.954	0.014	1, 19	2.813	0.110	0.129
	Civilian	0.974	0.007				

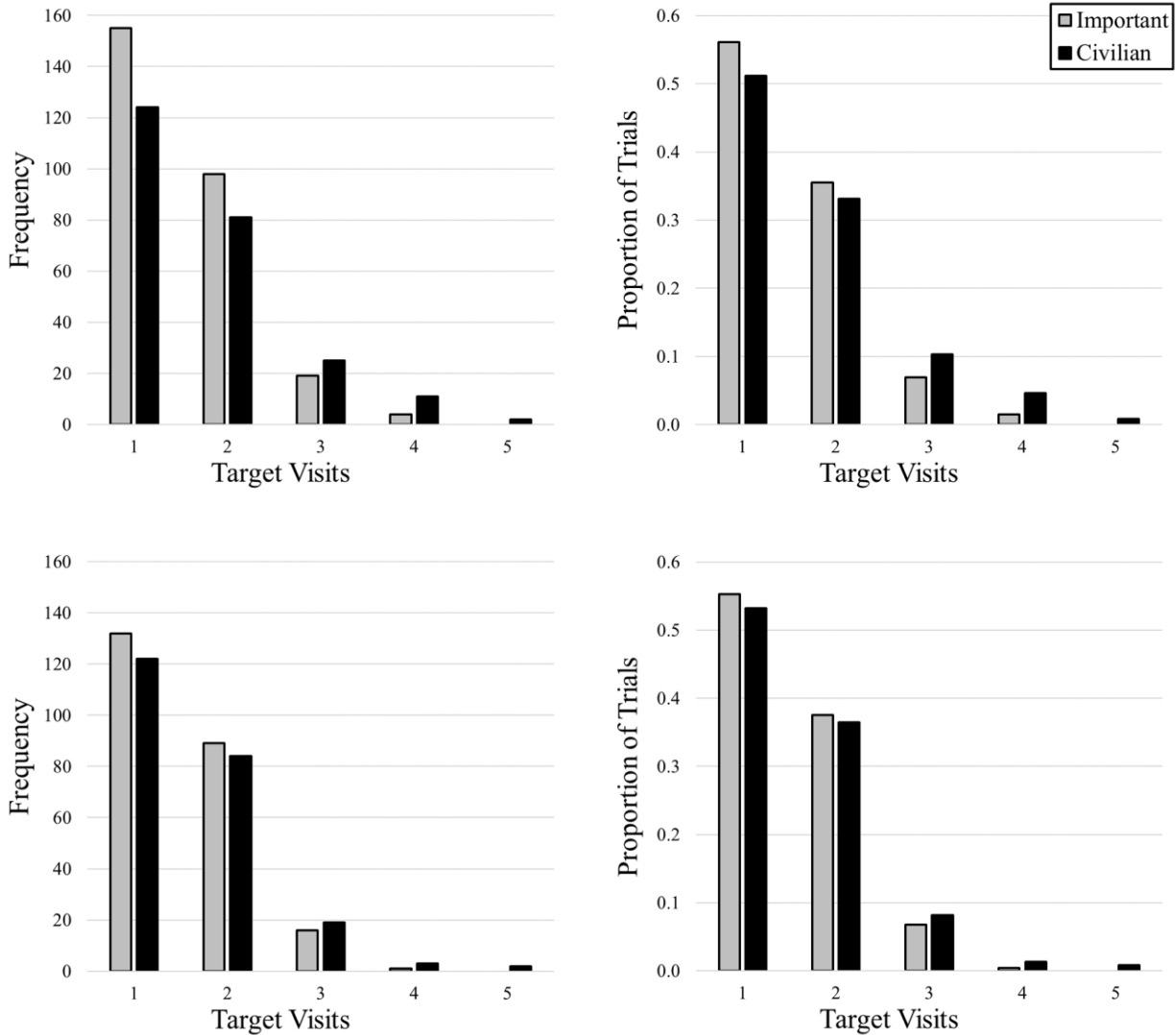


Figure A1. Frequency of target visits from Experiments 1a (top) and 1b (bottom). Gray bars show results for societally important targets, and black bars show civilian targets. Left panels show raw frequency data. Frequencies differed across experiments and conditions due to differences in the number of participants and accurate search trials. To make the conditions more directly comparable, the right panel show raw frequency data converted to a proportion of total trials per condition.

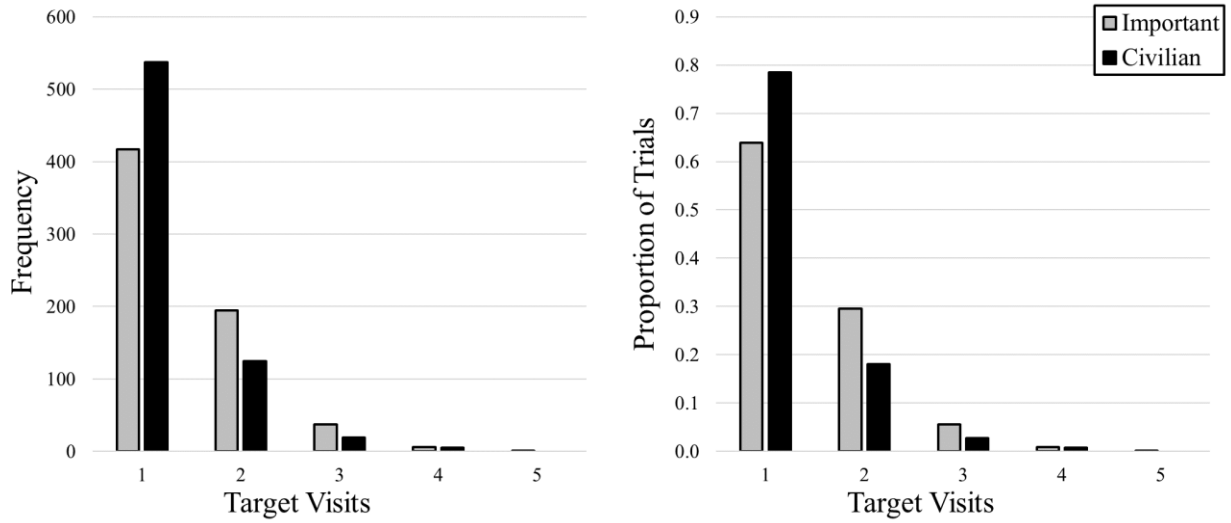


Figure A2. Frequency of target visits from Experiment 2. Gray bars show results for societally important targets, and black bars show civilian targets. Left panes shows raw frequency data. Frequencies differed across conditions due to differences in the number of accurate search trials. To make the conditions more directly comparable, the right panel show raw frequency data converted to a proportion of total trials per condition.