



The detail is in the difficulty: Challenging search facilitates rich incidental object encoding

Juan D. Guevara Pinto¹ · Megan H. Papesh² · Michael C. Hout²

Published online: 19 June 2020
© The Psychonomic Society, Inc. 2020

Abstract

When searching for objects in the environment, observers necessarily encounter other, nontarget, objects. Despite their irrelevance for search, observers often incidentally encode the details of these objects, an effect that is exaggerated as the search task becomes more challenging. Although it is well established that searchers create incidental memories for targets, less is known about the fidelity with which nontargets are remembered. Do observers store richly detailed representations of nontargets, or are these memories characterized by gist-level detail, containing only the information necessary to reject the item as a nontarget? We addressed this question across two experiments in which observers completed multiple-target (one to four potential targets) searches, followed by surprise alternative forced-choice (AFC) recognition tests for all encountered objects. To assess the detail of incidentally stored memories, we used similarity rankings derived from multidimensional scaling to manipulate the perceptual similarity across objects in 4-AFC (Experiment 1a) and 16-AFC (Experiments 1b and 2) tests. Replicating prior work, observers recognized more nontarget objects encountered during challenging, relative to easier, searches. More importantly, AFC results revealed that observers stored more than gist-level detail: When search objects were not recognized, observers systematically chose lures with higher perceptual similarity, reflecting partial encoding of the search object's perceptual features. Further, similarity effects increased with search difficulty, revealing that incidental memories for visual search objects are sharpened when the search task requires greater attentional processing.

Keywords Incidental memory · Visual search · Search difficulty

Introduction

While searching through the environment, observers gather large amounts of visual information, some more relevant than others. For example, while navigating a department store in search of jackets, observers likely view other clothing categories. Although irrelevant clothing sections (e.g., pants, shoes) can be quickly dismissed without much attention, research consistently shows that observers incidentally retain information about nontarget objects encountered in real-world scenes (e.g., Hollingworth & Henderson, 2002; Hollingworth, 2004, 2006). These memories are more than mere guesses: Across several experiments, Castelhana and Henderson (2005) found

that incidentally encoding visual scene objects produced comparable memory performance to intentionally encoding those objects. Similar results have been reported elsewhere (e.g., Draschkow, Wolfe, & Võ, 2014; Josephs, Draschkow, Wolfe, & Võ, 2016; Williams, 2010), suggesting that observers clearly encode (and later recognize) objects encountered during visual search, even if they are not explicitly attempting to do so.

Not all searches are created equal: Many factors can affect the speed and accuracy of search tasks, such as the amount of visual clutter (Whitney & Levi, 2011), target-nontarget similarity (Alexander & Zelinsky, 2011, 2012; Duncan & Humphreys, 1989), cue precision or accuracy (Hout & Goldinger, 2015; Schmidt & Zelinsky, 2009; Yang & Zelinsky, 2009), and the number of searched-for targets (Menneer, Barret, Phillips, Donnelly, & Cave, 2007; Menneer, Cave, & Donnelly, 2009). Search difficulty also influences what observers incidentally remember from the search task. For example, Hout and Goldinger (2010, 2012) found that following challenging, multiple-target search, observers correctly recognized 11% more nontarget objects than

✉ Michael C. Hout
mhout@nmsu.edu

¹ Department of Psychology Rollins College, Winter Park, FL, USA

² Department of Psychology, New Mexico State University, Las Cruces, NM, USA

they did following easier, single-target search. This pattern reflects a trade-off between search efficiency and incidental object encoding, because challenging searches were slower, but yielded better incidental memory. Thomas and Williams (2014) replicated this trade-off in a different search task: Even when observers could not know the target details in advance, and instead had to slowly locate the unique item in search displays, observers had better memory for nontarget objects, relative to speedier conditions with predefined targets.

Because challenging searches are conducted more slowly, observers have a greater likelihood of looking at nontargets, and for longer durations, which could potentially explain difficulty-enhanced nontarget memory. Indeed, Hout and Goldinger (2012) found that the number of fixations made during search predicted subsequent nontarget recognition accuracy, with more fixations occurring in difficult search conditions. However, in one task, Hout and Goldinger (2010) eliminated the need for eye movements by using Rapid Serial Visual Presentation (RSVP) search: Twenty objects appeared in sequential order for 250 ms each, and observers' target present/absent responses were only possible after *all* objects had been presented. In this way, item exposure durations were equated across all conditions. Although viewing time was equated for all pictures, observers' incidental memory for nontargets was again enhanced following the more difficult multiple-target search trials, relative to the easier single-target search trials.¹ Guevara Pinto and Papesch (2019) recently replicated these results and suggested that enhanced nontarget memory is due to attention-allocation strategies. When target discrimination is expected to be difficult, they suggested, observers allocate more attention to visually "scrutinizing" each search item. Across several experiments, difficult searches forced observers to focus attention on the central search stream, impairing their ability to spot items in the periphery, but improving nontarget encoding. To rule out the possibility that the effects were an artifact of simply searching for more items in the difficult conditions, they included conditions in which search cues were well or poorly specified words (e.g., *pie* vs. *dessert* or *food*). The results again replicated, and suggested that difficulty-enhanced nontarget memory, at least partially, derives from the manner in which observers process information: Difficult searches encourage a more exhaustive or attentive processing style to facilitate target detection, which incidentally enhances nontarget encoding.

What sort of information is incidentally stored in long-term memory during visual search, and how much can individuals retain? Early research on visual long-term memory (VLTm) consistently showed that people have high capacity for visual information, even when pictures were only studied for brief

durations (Shepard, 1967; Standing, 1973; Standing, Conezio, & Haber, 1970). More recently, studies have confirmed this large capacity, and have also shown that intentionally encoded visual objects yield highly-detailed representations (Brady, Konkle, Alvarez, & Oliva, 2013a; Brady, Konkle, Gill, Alvarez, & Oliva, 2013b; Konkle, Brady, Alvarez, & Oliva, 2010a, 2010b; Vogt & Magnussen, 2007). In a compelling demonstration of the capacity and precision of VLTm, Brady et al. (2008) had participants study 2,500 pictures for only 3 s each, and later select the studied item from two alternatives. When the alternatives were from different semantic categories (e.g., *bird* target, *table* lure), participants chose the correct picture 93% of the time. Even when both objects belonged to the same category, or when both images represented the same object in a different "state" (e.g., empty coffee mug, full coffee mug), participants correctly selected the old item more than 85% of the time. The large and detailed capacity for visual memories has been attributed to observers' appreciation of objects' conceptual and/or perceptual distinctiveness during encoding (e.g., Antonelli & Williams, 2017; Konkle et al., 2010a).

Although much has been learned about VLTm from intentional encoding procedures, memory formation in the real world is mostly incidental. In visual search tasks, for instance, processing and remembering nontargets is orthogonal to the goals of search. Moreover, studies with intentional encoding procedures typically employ 2-AFC memory comparisons, (e.g., Brady et al., 2008; Konkle et al., 2010a), old/new recognition (ONR) probes (e.g., Cunningham, Yassa, & Egeth, 2015), or continuous-report methods (e.g., Brady et al., 2013b) to assess the fidelity of object representations stored in VLTm. With the exception of continuous report methods, many procedures leave open the possibility that responses can be based on either perceptually rich representations or categorical gist-level representations (e.g., any *bird*, relative to a specific *bird*). In continuous report methods, participants are typically given a 360° color wheel, from which they either select the color of the studied object or move a cursor around the wheel until the color matches their memory (i.e., a method of adjustment). When observers select the wrong color, they often select colors perceptually close to the studied color (Brady et al., 2013b). By plotting the probability of errors across levels of perceptual similarity, researchers can gain insight into the precision of VLTm. Although continuous report methods are ideal for paradigms with distinctly colored objects, many visual search studies avoid using distinct or unnatural colors during visual search,² making continuous report difficult to implement with standard search paradigms.

¹ Previous work has demonstrated that multiple-target search incurs performance costs relative to single-target search (e.g., Menneer et al., 2007; Menneer, Cave, & Donnelly, 2009). However, it is possible that these costs derive from memory demands rather than from task parameters such as search difficulty.

² Color serves as a basic visual feature, which would make most search processes relatively easy, regardless of other manipulations (see Wolfe & Horowitz, 2017).

To measure whether incidentally encoded search objects are characterized by perceptual detail or categorical gist, we created a semi-continuous report method by expanding upon standard alternative forced-choice (AFC) paradigms. Specifically, we manipulated the number of within-category alternatives (four versus 16), and the similarity relationship between exemplars (previously seen objects) and lures (new within-category items). Whereas the continuous report color wheel provides a perceptually determined similarity space, our method relied on multidimensional scaling (MDS) to determine the similarity relationships between exemplars and lures. Since Shepard's (1987) foundational work demonstrating the utility of MDS for similarity assessment (for an earlier account, see Torgerson, 1952), it has been used to reveal the similarity structure of categories ranging from irregular shapes (e.g., Homa, Blair, McClure, Medema, & Stone, 2018) to man-made and naturalistic objects (e.g., Hout, Goldinger, & Brady, 2014; Horst & Hout, 2015), and even scenes of architecture and nature (Berman et al., 2014; Coburn et al., 2019). MDS is a statistical tool that examines observers' similarity ratings derived from many possible formats (e.g., pairwise Likert-scale ratings, spatial distance metrics, perceptual discrimination times). Using these data, algorithms model the psychological similarity between all items in the set, resulting in similarity "maps" for each set. These maps arrange items spatially, such that inter-item distance is proportional to perceived similarity, with shorter distances denoting greater perceived similarity (see Hout, Papesh, & Goldinger, 2012, and Hout et al., 2013, 2015, 2016, for reviews and tutorials). MDS distance has been used in cognitive research for decades, informing theoretical understanding of categorization (e.g., Goldstone, 1994; Nosofsky, 1984, 1992), face perception (Papesh & Goldinger, 2010; Pedelty, Cohen, Levine, & Shevall, 1985; Valentine, 1991), language (Goldinger,

1998), and visual search (Godwin, Hout, & Menneer, 2014; Hout & Goldinger, 2016), among others.

In the present study, MDS distances were used to quantify exemplar-lure similarity in surprise 4-AFC and semi-continuous 16-AFC memory tests, following easy and difficult search tasks. During the search phases of all experiments, participants encountered pictures of real-world objects from distinct categories as targets or nontargets. Only one exemplar from each category ever appeared during this phase. Search difficulty was manipulated via the number of potential targets held in working memory. Following all search tasks, we tested participants' memory for each exemplar against within-category lures, each of which had a measurable similarity relationship to the previously seen object. For example, if participants encountered a *butterfly* during incidental encoding, a test trial would present them with that butterfly and three or 15 other butterflies, with the goal of selecting the correct "old" (i.e., previously encountered) butterfly. We predicted that we would replicate the difficulty-enhanced memory effect observed in prior work (e.g., Hout & Goldinger, 2010, 2012; Thomas & Williams, 2014), with higher hit rates to old exemplars when search was more challenging. We also hypothesized that these memories would be characterized by rich perceptual detail (i.e., high-fidelity representations), rather than sparse or gist-like detail (i.e., low-fidelity representations), because incidental encoding during difficult searches benefits from exhaustive item processing (e.g., Guevara Pinto & Papesch, 2019). Although this would certainly manifest in higher hit rates to old exemplars, this hypothesis can also be tested by examining errors, as is commonly done in research on VLTm fidelity (e.g., Brady et al., 2013a, b). If observers store perceptually rich representations, false alarms during AFC tests should be predicted by the exemplar-lure similarity of each object, with decreasing false alarms as "distance" from the exemplar increases (Fig. 1, left panel). Conversely, if

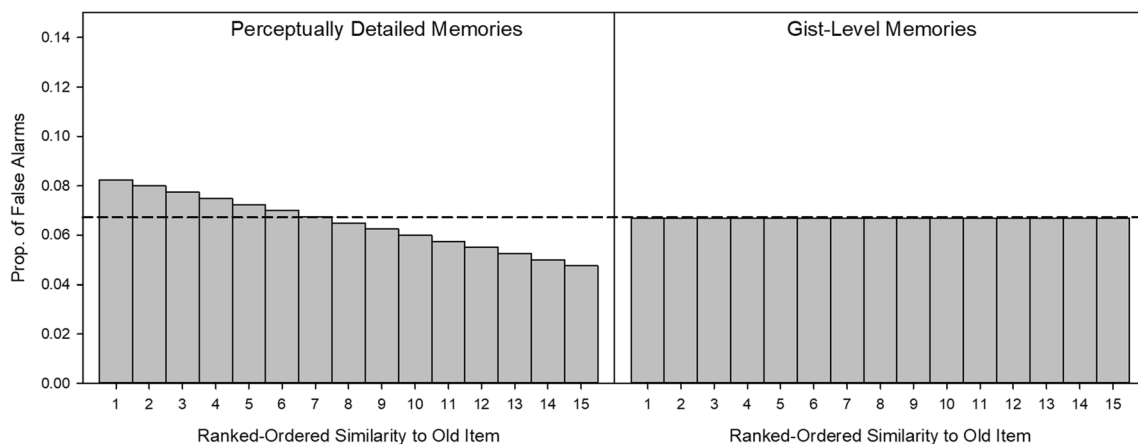


Fig. 1 Hypothetical false-alarm distribution for perceptually detailed memories (*left panel*) and gist-level memories (*right panel*). The x-axis reflects the rank-ordered exemplar-lure similarity to the old item, with 1 being the lure of highest similarity and 15 being the lure of lowest

similarity in a 16-AFC memory test. The dashed line indicates an equal distribution of false alarms across all levels of exemplar-lure similarity (6.67% each)

observers encode only sparse, gist-level representations, false alarms should be equally distributed across all levels of exemplar-lure similarity (Fig. 1, right panel).

Experiment 1

In Experiment 1, we explored the influence of search difficulty on incidental memory by manipulating observers' search loads. Although only one or two targets could ever be present in the search display, observers kept in mind two, three, or four *potential* targets. Prior research has shown that search conducted under higher working memory (WM) load (i.e., with more potential search targets in mind) yields slower, less accurate target detection (e.g., Menneer et al., 2007; Menneer, Cave, & Donnelly, 2009), but better incidental memory for nontargets (Hout & Goldinger, 2010, 2012). In Experiment 1, we used both 4-AFC (Experiment 1a) and semi-continuous 16-AFC (Experiment 1b) tests to determine whether difficulty-enhanced nontarget memory effects produce perceptually rich memories or categorical (gist-like) memories.

Method

Participants. A power analysis (within-subjects effect $\alpha = .05$, $\beta - 1 = .90$) conducted on the average effect size across all four experiments reported in Hout and Goldinger (2010; $\eta_p^2 = .16$) indicated that at least 44 participants were necessary to observe a reliable effect of search difficulty on nontarget recognition. Sixty-one observers (68.8% women) from New Mexico State University participated in Experiment 1a, and 67 new observers (70.1% women) participated in Experiment 1b (30 observers from NMSU; 37 from Louisiana State University). All observers participated in exchange for partial course credit.

Stimuli and apparatus Stimuli were images from the *Massive Memory Multidimensional Scaling* database (Hout, Goldinger, & Brady, 2014; images originally obtained from <http://cvcl.mit.edu/MM/stimuli.html> – see Brady et al., 2008), which includes nearly 4,000 color images of real-world objects spread across 240 different categories (with 16–17 exemplars per category; see Appendix A for a full list of categories). Within each category, the database contains values denoting every item's inter-item distance with all other same-category objects, which Hout et al. (2014) rank-ordered to classify as “close,” “mid,” or “far” from one another in MDS space (using a ternary split). These designations correspond to pairs of items that were similar, moderately similar, and dissimilar, respectively. To populate 4-AFC trials with three lures in Experiment 1a, one unseen nontarget from each similarity level was randomly selected, allowing us to label items as having high, medium, or low perceptual similarity to the studied exemplar. Experiment 1b, however, included all possible exemplars from each object category, allowing us to compare perceptual similarity more continuously (see Fig. 2). Images were presented in full color and resized to fit within a 100×100 pixel square, subtending 2.3° of visual angle (horizontally and vertically) at a viewing distance of 60 cm. Stimuli were presented on 21.5-in. monitors, with $1,920 \times 1,080$ screen resolution and 60-Hz sampling rates. Experimental procedures were controlled using E-Prime 2.0 (Psychology Software Tools, 2006).

Procedure After providing informed consent, participants completed all visual search trials, followed by a surprise memory test. The visual search phase comprised two blocks of trials, with a 2-min break between each block. Before each trial, participants studied two, three, or four potential targets (low, medium, and high WM Load conditions, respectively), knowing that up to two might appear in the upcoming display. Once participants were ready to begin the search, they clicked the left mouse button to dismiss the search cues, after which a 250-ms fixation cross appeared at the center of the screen.

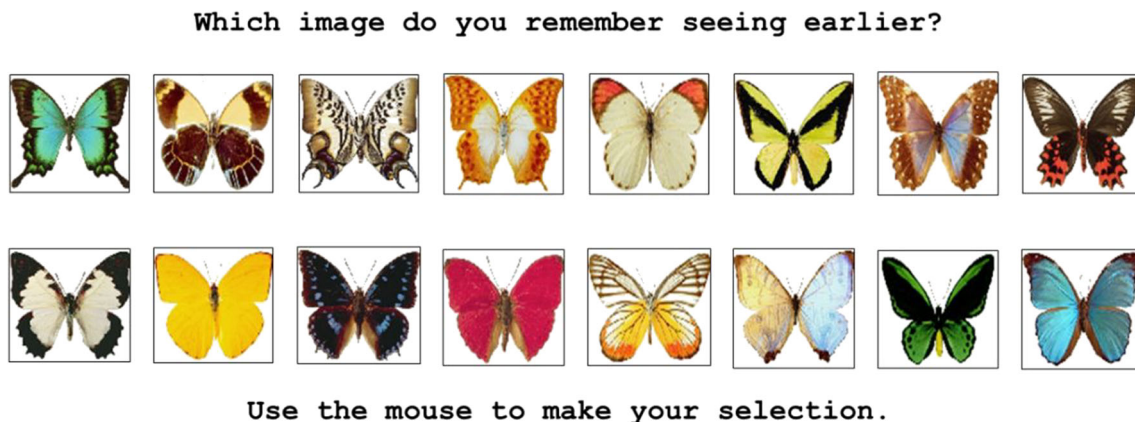


Fig. 2 Sample recognition test for all exemplars in the butterfly category. *Note:* Images were presented to participants in color. Full color version available online

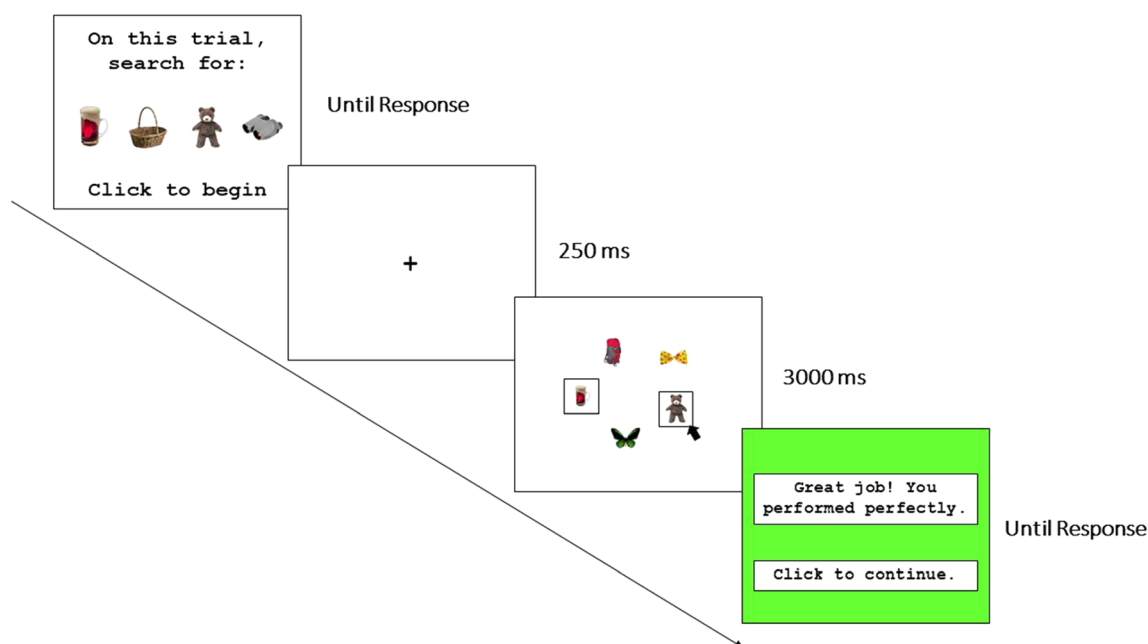


Fig. 3 Trial schematic following a High Working Memory (WM) Load trial in Experiments 1a and 1b. *Note:* Images were presented to participants in color. Full color version available online

Following the offset of the fixation cross, a five-object search array appeared and remained on the screen for 3 s. Search objects were presented in a circular array, spaced at equidistant locations from one another around an invisible circle whose radius was 10.3° of visual angle. In effect, this created an invisible circle that could be “rotated” randomly from one trial to the next, such that items did not always appear in precisely the same location from trial to trial; assignment of targets and nontargets to each of the five possible locations was randomized across trials.

Within each display, participants used the computer mouse to click on all targets as quickly as possible before the search array disappeared (i.e., all trials were target-present). As participants clicked objects, a black box appeared around each as visual confirmation (see Fig. 3). Half of the trials included one target, and half included two targets. To maximize nontarget encoding, one- and two-target trials were randomly intermixed within each block, encouraging participants to continue searching (and therefore to continue examining the items) after finding one target. Once the search array disappeared, participants received feedback on their search performance for both correct (i.e., all targets were selected) and incorrect responses (i.e., at least one target was not selected and/or at least one nontarget object was selected). Feedback was self-paced and participants clicked the left mouse button to start the next trial. Participants completed six practice trials (two per WM Load condition) before completing two 105-trial blocks of experimental trials. Within each block, there were an equal number of one- and two-target trials across WM Load conditions (35 each), presented in random order. Stimuli used in practice trials were not used elsewhere in the experiment.

At the beginning of each experiment, 217 categories were randomly selected from the 240 object categories in the database. Seventy categories were randomly designated as targets, and 49 were randomly assigned to each level of WM Load. From each category, one exemplar was randomly selected for use in the search task. Whereas the selected target exemplars could appear during any search trial, nontarget exemplars only appeared within their designated WM Load trials, and each appeared five times over the course of the search phase. This provided ample opportunity for encoding (as nontarget objects may not be directly viewed on all trials), but limited the possibility of ceiling effects in recognition. In this way, only one exemplar per category ever appeared during visual search.

After search, and before the surprise memory task, participants completed simple arithmetic problems (e.g., $9 \times 3 = ???$) for 2 min. Each problem was presented in the center of the screen and participants used the mouse to choose between a correct (i.e., 27) and incorrect (e.g., 25) answer, which were always within two digits of each other (so that incorrect answers were not obvious). Following the arithmetic task, participants completed a surprise AFC recognition memory test for all objects encountered during search. For each recognition decision, participants saw either four objects (Experiment 1a) or 16 objects (Experiment 1b) presented in horizontal lines: One previously seen object and three or 15 lures from the same object category (see Fig. 2). The lures in Experiment 1a were chosen by selecting category items rated as relatively similar, moderately similar, and relatively dissimilar to the studied exemplar (see *Stimuli and apparatus*), and the lures in Experiment 1b contained the entire similarity spectrum. Exemplars and lures appeared in randomly determined

locations on every trial. Participants clicked the object they remembered seeing earlier in the search task, and no feedback was provided. Both target and nontarget exemplars were tested, in random order, resulting in 217 recognition trials (70 targets and 147 nontargets, 49 from each WM Load condition).

Results

One participant from Experiment 1b was excluded from analyses due to inattention during the experimental session, as indicated by researcher observation. Two additional participants from Experiment 1b were removed from analyses for committing false alarms during visual search in more than 15% of trials. All proportion data were arcsine-square-root transformed prior to analysis to ensure normality. For clarity, we present raw values in our descriptives and graphs, but inferential statistics are based on the transformed data (inferential statistics on raw data do not differ from those based on transformed data, and are reported in Appendices E and F, respectively, along with descriptive statistics). Alpha level for all analyses was set at .05, and multiple comparisons were subjected to Bonferroni corrections. Greenhouse-Geisser-corrected degrees of freedom are reported for any sphericity violations. Traditional (i.e., frequentist) analyses of variance (ANOVAs) were supplemented with Bayesian ANOVAs conducted in JASP (JASP Team, 2018) using the JASP-recommended default parameters (Cauchy scale = .707, r fixed effects = .05, r random effects = 1) as suggested by Rouder, Morey, Speckman, and Province (2012; see also Wagenmakers et al., 2018). For Bayesian ANOVAs, the Bayes factor (BF_{10}) reflects how likely the data are under a model including each effect or interaction, relative to a null model without the effect or interaction. These analyses are provided to complement the frequentist analyses and allow readers a fuller appreciation of the strength of effects (both null and positive). Raw data and analysis files for all experiments are available on OSF.³

Overall, search accuracy was high, with observers detecting 98% of targets across both experiments and false alarming in fewer than 0.5% of trials. Analyses (reported in Appendix B) confirm that our search difficulty manipulation was effective: In both Experiments, observers were most accurate in Low-Load searches. For brevity, we also report analyses of memory for targets in Appendices C and D, as these items were intentionally encoded, and our hypotheses centered around incidental memory for nontarget objects.

Memory for nontargets To determine whether observers were better able to remember nontargets encountered in the context of challenging (relative to easy) search, we examined the

proportion of recognition hits in Experiments 1a and 1b in separate repeated-measures ANOVAs testing the effect of WM Load (low, medium, high). A reliable effect was observed in both Experiments (Exp. 1a: $F(2, 120) = 69.4$, $p < .001$, $\eta_p^2 = .54$, $BF_{10} = 3.612e+17$; Exp. 1b: $F(2, 126) = 24.11$, $p < .001$, $\eta_p^2 = .28$, $BF_{10} = 7.333e+6$). As shown in Fig. 4, nontarget objects encountered during High-Load searches were better recognized than nontargets encountered during Medium- and Low-Load searches, all $ps < .001$. Nontargets encountered during Medium-Load searches were also better recognized than those encountered during Low-Load searches, both $ps < .03$. The results of both experiments confirm our predictions, and replicate previous findings (e.g., Hout & Goldinger, 2010, 2012), showing that observers remember more objects encountered during challenging relative to easy, visual search, even when the memory test is more stringent (16-AFC, Exp. 1b).

Recognition errors for nontargets In Experiment 1a, each 4-AFC comparison included lures designated as relatively similar, moderately similar, and relatively dissimilar to the studied exemplar. Similarly, in Experiment 1b, lures were rank-ordered from 1 (*highest similarity*) to 15 (*lowest similarity*). False alarms could thus be examined as a function of the perceptual overlap between studied exemplars and categorically related lures in both experiments. To determine whether observers' incidental memories were characterized by perceptual richness or categorical gist (see Fig. 1), we examined error probabilities as a function of exemplar-lure similarity across levels of search difficulty.

For both Experiments, repeated-measures ANOVAs were used to analyze the impact of exemplar-lure similarity and search difficulty on the proportion of false alarms. Experiment 1a used a 3 (WM Load: Low, Medium, High) ×

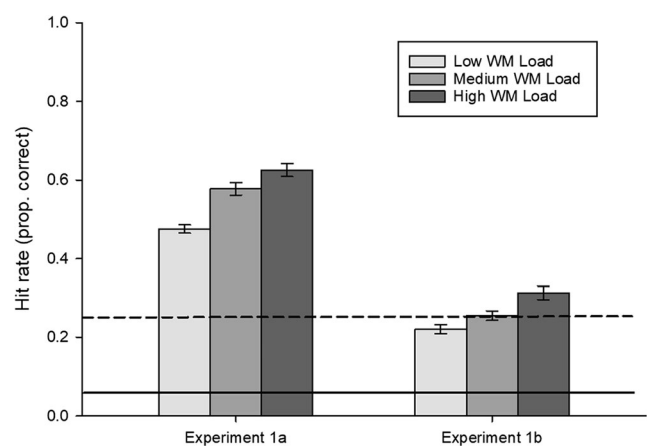


Fig. 4 Recognition hit rates for nontargets encountered in Experiments 1a and 1b as a function of working memory (WM) load. The dashed line represents chance-level performance in Experiment 1a (25%) and the solid line represents chance performance in Experiment 1b (6.25%). Error bars represent ± 1 SEM

³ <https://osf.io/p7yk2/>

Table 1 Statistics from omnibus analyses on the proportion of false alarms as a function of working memory load and exemplar-lure similarity in Experiments 1a and 1b

| Experiment | Effect | Frequentist | | | | Bayesian | | Evidence favoring effect? |
|-------------|----------------|-------------|----------|----------|------------|------------------|--|---------------------------|
| | | df | <i>F</i> | <i>p</i> | η_p^2 | BF ₁₀ | | |
| 1a (4-AFC) | | | | | | | | |
| | WM Load (WM) | 2,120 | 4.52 | 0.01 | 0.07 | 0.023 | | No evidence |
| | Similarity (S) | 2,120 | 18.44 | < .001 | 0.24 | 4.20e+7 | | Strong evidence |
| | WM × S | 4,240 | 2.09 | 0.08 | 0.03 | 1.782 | | Weak evidence |
| 1b (16-AFC) | | | | | | | | |
| | WM Load (WM) | 2,120 | 8.11 | < .001 | 0.12 | 0.011 | | No evidence |
| | Similarity (S) | 10.2, 640.9 | 5.46 | < .001 | 0.08 | 1.01e+7 | | Strong evidence |
| | WM × S | 28, 1764 | 0.89 | 0.63 | 0.01 | 0.0489 | | No evidence |

3 (Relative Exemplar-Lure Similarity: Similar, Moderate, Dissimilar) ANOVA and Experiment 1b used a 3 (WM Load: Low, Medium, High) × 15 (Lure Similarity) ANOVA. The statistics for both analyses are presented in Table 1. As shown in the Table, both Experiments yielded reliable effects of WM Load and Similarity, but no interactions, indicating that nontarget recognition errors were generally biased towards lures with high perceptual similarity to the incidentally encoded exemplars.

Although the main effects in the omnibus analyses on false recognition in both Experiments were generally consistent with our predictions, we failed to observe the predicted interaction. Specifically, we predicted that memory errors driven by exemplar-lure similarity would vary as a function of search difficulty, such that highly similar lures would be more likely to be falsely recognized following challenging, relative to easy, search. Although the interactions in both experiments were not reliable, we conducted simple effects analyses on

the effect of lure similarity on the proportion of total false alarms at each WM Load level to fully examine this a priori prediction. Statistics from these analyses can be found in Table 2.

In both Experiments, there was no relationship between exemplar-lure similarity and the probability of false alarms during Low-load search. During Medium- and High-load search, however, we observed an exemplar-lure similarity effect in both Experiments, characterized by an increased probability of false alarms as the lures increased in similarity to the exemplar. In Experiment 1a, all pairwise comparisons were reliable in Medium-load search (all *ps* < .05), but the effect in High-load search was driven exclusively by the high-similarity lures (moderate and dissimilar lures did not statistically differ, *p* > .05; see Fig. 5). Experiment 1b produced similar results, except that the effect of exemplar-lure similarity was strongest in the High-load search condition (see Fig. 6). These results, presented in Table 2, suggest that when

Table 2 Statistics from simple effect analyses of exemplar-lure similarity across levels of search difficulty in Experiments 1a and 1b

| Experiment | Search | Frequentist | | | | Bayesian | | Evidence for lure similarity effect? |
|-------------|----------------|-------------|----------|------------|------------|------------------|------------------|--------------------------------------|
| | | <i>F</i> | <i>p</i> | η_p^2 | CI | BF ₁₀ | BF ₀₁ | |
| 1a (4-AFC) | | | | | | | | |
| | Low WM Load | .812 | .446 | .013 | .000, .067 | .167 | 5.98 | No evidence |
| | Medium WM Load | 10.06 | < .001 | .144 | .040, .251 | 14850 | 6.734e-5 | Strong evidence |
| | High WM Load | 6.52 | .002 | .098 | 0.14, .197 | 210.44 | .005 | Strong evidence |
| 1b (16-AFC) | | | | | | | | |
| | Low WM Load | 1.69 | .053 | .026 | .000, .034 | .053 | 18.91 | No evidence |
| | Medium WM Load | 2.30 | .004 | .035 | .004, .047 | 1.51 | .662 | Weak evidence |
| | High WM Load | 2.93 | <.001 | .044 | .010, .058 | 47.42 | .021 | Strong evidence |

Note: BF₁₀ reflects strength of evidence for the alternative, BF₀₁ reflects strength of evidence for the null

CI reflect 95% confidence intervals for partial eta-square

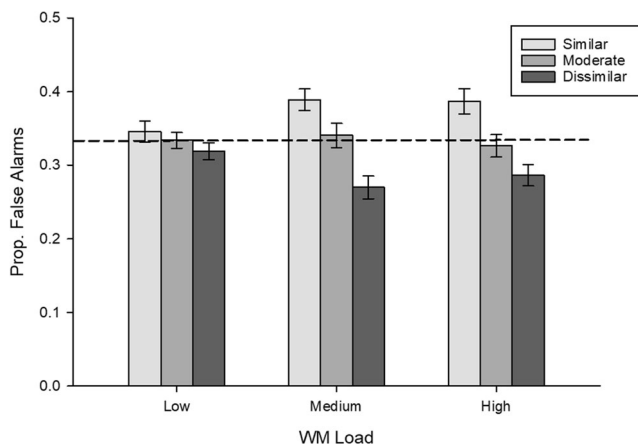


Fig. 5 Proportion of total false alarms for nontarget objects encountered in Experiment 1a as a function of exemplar-lure similarity. The dashed line indicates an equal distribution of false alarms across all levels of exemplar-lure similarity (33.33% each). Error bars represent ± 1 SEM

visual search is challenging, objects are processed more extensively, such that detailed perceptual representations are incidentally stored in memory. These detailed object representations allow observers to recognize more objects encountered during challenging, relative to easy, search, and also bias observers into selecting lures with high perceptual overlap with the encoded object when recognition fails.

Whereas the ANOVAs allowed us to determine whether search difficulty and exemplar-lure similarity predicted false alarm rates, they do not speak to whether the false alarm rates reliably differed from chance. As shown in the left panel of Fig. 1, if observers encode rich perceptual detail, the highest-similarity lures should produce false alarm rates reliably above chance, whereas lowest-similarity items should fall below chance. To evaluate this prediction, we conducted one-sample *t*-tests on the raw false alarm rates for items with the highest and lowest exemplar-lure similarity ratings, comparing each exemplar’s proportion of false alarms to chance (33.33% in Exp. 1a; 6.67% in Exp. 1b).⁴ Table 3 presents the outcome of these analyses.

For the Low-Load condition in Experiment 1a, neither the “similar” nor “dissimilar” lures reliably differed from chance, suggesting that memories were more gist-like. In the Medium- and High-Load conditions, however, both exemplars reliably differed from the 33.33% criterion, with the high similarity item above that criterion and the low similarity item below it. These analyses confirm the interpretation of the ANOVAs: Challenging search produces perceptually rich incidental memories. Similar analyses on false alarm rates for the highest- and lowest-similarity items in Experiment 1b replicated this finding: False alarm rates for the most similar and dissimilar lures differed from the criterion, and in the predicted directions, when the object was encountered during challenging searches (e.g., Medium- and High-Load; see Table 3).

Discussion

Experiments 1a and 1b examined incidental memory formation as a function of visual search difficulty, which we manipulated by varying the number of potential targets. Following the search phase, participants completed a surprise 4-AFC (Experiment 1a) or 16-AFC (Experiment 1b) recognition phase probing their memory for exemplars previously encountered as search targets and nontargets. Lures in the memory test consisted of within-category objects that varied in similarity to studied exemplars. This allowed us to examine the precision of incidental memory representations by looking at both true and false memories. In both experiments, we replicated the effects of search difficulty on search accuracy and incidental memory (e.g., Hout & Goldinger, 2010, 2012): Search was less accurate, but memory was better, for exemplars encountered during difficult, relative to easy, search. As a novel extension, we examined the effect of exemplar-lure similarity on the distribution of error probabilities when true recognition failed. When participants incorrectly “recognized” a lure on the memory test, they were biased toward selecting lures with higher similarity to the studied exemplar. Although we did not observe the predicted interaction between search difficulty and exemplar-lure similarity, simple effects tests of

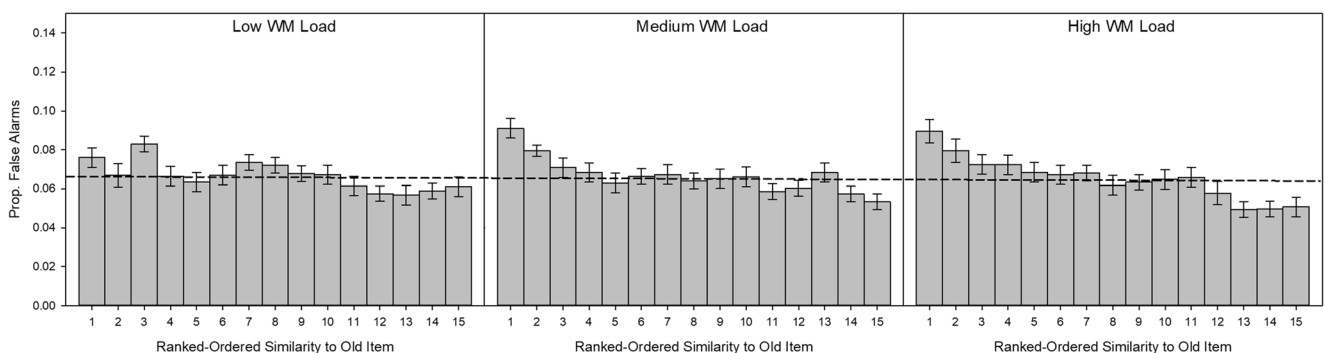


Fig. 6 Proportion of total false alarms for nontarget objects encountered in Experiment 1b as a function of exemplar-lure similarity (e.g., 1 = lure of highest similarity). The dashed line indicates an equal distribution of

false alarms across all levels of exemplar-lure similarity (6.67 % each). Error bars represent ± 1 SEM

Table 3 Statistics from one-sample *t*-tests comparing false alarm rates for exemplars with high and low similarity at each level of difficulty to equal distribution of false alarms in Experiments 1a and 1b

| Experiment | Exemplar similarity | Frequentist | | | | Bayesian | | |
|-------------|---------------------|-------------|----------|------------------|--------------|------------------|------------------|--------------------------------------|
| | | <i>t</i> | <i>p</i> | Cohen's <i>d</i> | CI | BF ₁₀ | BF ₀₁ | Evidence for lure similarity effect? |
| 1a (4-AFC) | | | | | | | | |
| | Low Load Search | | | | | | | |
| | Similar | .95 | .345 | .122 | -.131, .373 | .216 | 4.638 | No evidence |
| | Dissimilar | -1.23 | .221 | -.158 | -.410, .095 | .289 | 3.458 | No evidence |
| | Med. Load Search | | | | | | | |
| | Similar | 3.80 | <.001 | .487 | .219, .751 | 71.07 | .014 | Strong evidence |
| | Dissimilar | -4.03 | <.001 | .516 | -.781, -.247 | 140.55 | .007 | Strong evidence |
| | High Load Search | | | | | | | |
| | Similar | 3.13 | .003 | .400 | .138, .660 | 10.85 | .092 | Strong evidence |
| | Dissimilar | -3.19 | .002 | .409 | -.669, -.146 | 12.97 | .077 | Strong evidence |
| 1b (16-AFC) | | | | | | | | |
| | Low Load Search | | | | | | | |
| | 1st rank | 1.83 | .072 | .229 | -.020, .476 | .658 | 1.52 | No evidence |
| | 15th rank | -1.96 | .053 | .246 | -.494, .004 | .835 | 1.19 | No evidence |
| | Med. Load Search | | | | | | | |
| | 1st rank | 4.18 | <.001 | .522 | .259, .782 | 231.10 | .004 | Strong evidence |
| | 15th rank | -3.24 | .002 | .405 | -.659, -.149 | 14.78 | .068 | Strong evidence |
| | High Load Search | | | | | | | |
| | 1st rank | 3.44 | .001 | .431 | .173, .686 | 25.95 | .039 | Strong evidence |
| | 15th rank | -2.23 | .029 | .280 | -.528, -.029 | 1.39 | .719 | Weak evidence |

Note: The test value was set to .333 for Exp. 1a and .067 for Exp. 1b

BF₁₀ reflects strength of evidence for the alternative, BF₀₁ reflects strength of evidence for the null

CI reflect 95% confidence intervals for Cohen's *d*

this a priori prediction indicate that similarity exerted the biggest effect on items encountered during difficult searches, and that these effects were clearly evident by examining false-alarm rates for items with the highest versus lowest exemplar-lure similarity. Combined, these results provide partial support for the prediction that object representations formed during challenging conditions contain richer perceptual details than those formed during easy search.

Experiment 1 suggests that observers incidentally encode objects' perceptual details during challenging search, potentially due to greater attentional resources needed for visually scrutinizing nontarget objects, compared to easy searches (Guevara Pinto & Papesch, 2019). We suggest that, when cued with imprecise or challenging search targets (e.g., four potential targets), observers do not fully rely on target cues to guide search (Wolfe et al., 1989; cf. Theeuwes, 1994a, 1994b). Instead, they allocate greater attention to closely inspect each item in the display, with the goal of preventing target misses. The additional resources devoted to item processing incidentally benefit nontarget encoding, resulting in detailed object representations being stored in memory. Thus,

when observers searched under high WM load, they processed each nontarget to a greater extent than when searching under low WM load, resulting in "recognition" of highly similar lures in the subsequent AFC test.

Experiment 2

Experiment 1 provided partial evidence suggesting that incidental object memories encoded during challenging searches possess greater detail than those encoded during easier ones. However, this prediction cannot be fully supported without observing a reliable interaction between search difficulty and lure similarity in the proportion of false alarms. It is possible that this interaction was not observed in Experiment 1 due to the nature of the search task: Although search performance statistically differed across each level of difficulty, these differences may have been too small for observers to notice. For instance, in Experiment 1a, search hits ranged from 96% to 99% across conditions (in Experiment 1b, hits ranged from 97% to 99%; see Appendix B). Such small differences

between each level of search difficulty may have caused observers to perceive them all as equally challenging, limiting trial-by-trial changes to attention allocation strategies. We conducted a second experiment to allow observers to more easily appreciate the differences between conditions, examining the influence of exemplar-lure similarity on memory as a function of single- versus three-target search.

In addition to increasing the difficulty disparity across conditions, Experiment 2 also tested the possibility that detailed nontarget memories following difficult searches are due to oculomotor differences across conditions: Although search arrays in Experiment 1 were displayed for consistent durations (3,000 ms), it is possible that difficult search cues encouraged participants to inspect more objects (and/or for longer durations) than easy search cues, allowing for better incidental encoding. If so, then the results of Experiments 1a and 1b may be an artifact of eye movements (or be partially explained by them), rather than processing efficiency. Experiment 2 thus used a *passive*, Rapid Serial Visual Presentation (RSVP), search task. In RSVP, each search object briefly appears in serial sequence at the center of the screen, eliminating the need for overt eye movements. Observers' target present/absent responses are issued at the offset of the entire RSVP stream, ensuring that all objects are viewed for the same duration. If difficulty-enhanced incidental memories arise from greater attention to item processing, rather than the frequency or duration of eye movements to each item during search, then recognition errors should again be biased towards lures with higher exemplar-lure similarity.

Participants Sixty-six observers (65.2% women) from Louisiana State University participated in Experiment 2 in exchange for partial course credit.

Stimuli and apparatus All stimuli and viewing conditions were identical to those used in previous experiments.

Procedure Experiment 2 was similar to Experiment 1b, where participants completed a search phase, followed by an arithmetic distractor task, and then a surprise 16-AFC memory test. The search phase, however, used RSVP search in lieu of an object array, and observers searched for one (i.e., Low-Load) or three (i.e., High-Load) picture-cued targets during a 24-item serial sequence of objects. As in the previous experiments, observers self-initiated each trial after encoding the target(s), followed by a 250-ms central fixation cross. The 24-object RSVP sequence began at the offset of the fixation cross, with each object presented for 200 ms, with a blank 50-ms interstimulus-interval (ISI) between each object. At the end of the sequence, observers pressed the “F” key if a target had been present, or the “J” key if all targets were absent. Only one target could ever be present, if any. Responses were followed by a 1,500-ms feedback screen, ending the trial.

Participants completed four practice trials (two per WM Load condition), followed by two blocks of 60 RSVP search trials (target presence vs. absence was randomized with 50% overall prevalence) separated by a 2-min break. Observers were randomly presented with Low- and High-Load trials within each block (30 trials of each). As in previous experiments, each Load condition drew from its own set of nontarget objects. At the beginning of the experiment, 60 object categories were randomly selected for each participant to be used as targets across conditions throughout the experiment. For each Load condition, however, a unique set of 47 different nontarget object categories was randomly selected from the remaining categories. A random exemplar was then drawn from each category to be used as a nontarget object in the RSVP sequence. Given the very brief exposure of each object within the RSVP sequence (200 ms/per object), each nontarget exemplar was repeated 30 times over the course of the experiment (albeit in randomly selected serial positions within the sequence, and in different groupings of RSVP items confined to a single WM load condition), preventing floor-level recognition performance. Thus, each RSVP sequence included objects from 24 distinct categories. If present, the target object was randomly presented in any serial position within the sequence.

After working on the arithmetic distractor task for 2 min, participants then completed a surprise 16-AFC memory test for all nontargets, as described in Experiment 1b.

Results Experiment 2

Three participants were removed from analyses for committing more than 15% false alarms in RSVP search trials. As in Experiment 1, overall target detection was high (96.5% of targets detected) and the false alarm rate was low (only 7.1% of responses). Analyses (described in Appendix B) confirm that Low-Load search ($M = 97.1$, $SE = .01$) was more accurate than High-Load search ($M = 88.7$, $SE = .01$). Importantly, this difference in search hits was larger than those observed in Experiments 1a and 1b.

Memory for nontargets As shown in Fig. 7, we replicated the results of Experiment 1b, showing better incidental memory for nontargets encountered during High-Load ($M = .40$, $SE = .02$) than Low-Load searches ($M = .32$, $SE = .02$), $t(62) = 6.13$, $p < .001$, Cohen's $d = .77$, $BF_{10} = 201,002$.

Recognition errors for nontargets As in the previous experiment, our prediction was that when recognition fails, observers' responses should be biased toward lures with higher exemplar-lure similarity, particularly when search is challenging. This prediction was tested in a Load (Low, High) \times Lure Similarity repeated-measures ANOVA. A reliable effect of Lure Similarity was observed, $F(10.56, 654.99) = 6.18$, $p <$

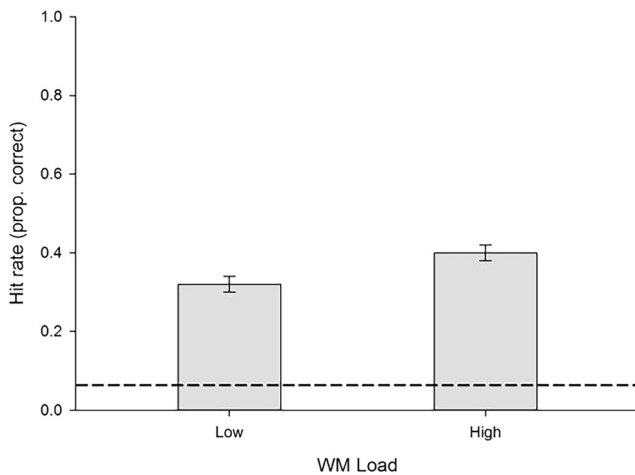


Fig. 7 Recognition hit rates for nontarget objects encountered in Experiment 2 as a function of Working Memory (WM) Load. The dashed line represents chance-level performance (6.25%). Error bars represent ± 1 SEM

.001, $\eta_p^2 = .09$, $BF_{10} = 1.41e+9$, but no effect of Load was found. More importantly, we observed the predicted Load \times Similarity interaction, $F(10.45, 647.95) = 1.87$, $p = .044$, $\eta_p^2 = .03$, $BF_{10} = .037$. As in Experiment 1, we conducted simple effects tests for each Load condition, comparing the effect of similarity on the proportion of false alarms at each level of search difficulty. A reliable effect was observed for objects encoded in the Low-Load condition, $F(14, 868) = 2.76$, $p < .001$, $\eta_p^2 = .04$, 95% CI = [.008, .056], $BF_{10} = 17.77$, but the effect was twice the size in the High-Load condition, $F(14, 868) = 5.16$, $p < .001$, $\eta_p^2 = .08$, 95% CI = [.034, .098], $BF_{10} = 1.19e+7$ (see Fig. 8). This further confirms that the difficulty-enhanced incidental memory effect observed in Experiment 1 was not a byproduct of eye movements or encoding time. Instead, when search difficulty forces observers to allocate greater attentional resources to visually inspecting search objects, they incidentally retain greater perceptual details about those objects.

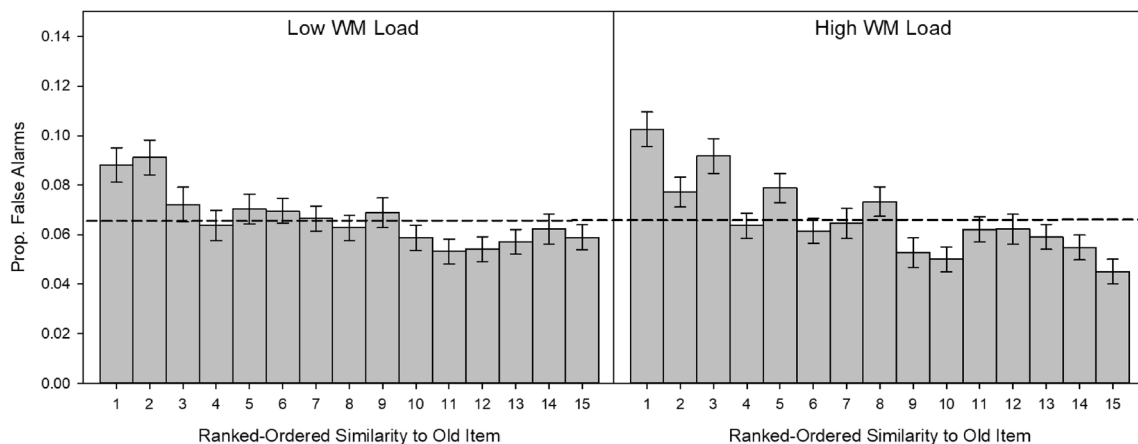


Fig. 8 Proportion of total false alarms for nontarget objects encountered in Experiment 2 as a function of exemplar-lure similarity (e.g., 1 = lure of highest similarity). The dashed line indicates an equal distribution of false

alarms across all levels of exemplar-lure similarity (6.67% each). Error bars represent ± 1 SEM

As in Experiment 1b, one-sample t -tests were conducted to determine whether the proportion of false alarms for lures with the highest and lowest exemplar-lure similarity differed from what gist-based memories would predict (i.e., an equal distribution of false-alarms at 6.67% across similarity levels). For the Low-Load condition, the proportion of false alarms for highest-similarity lure was greater than the criterion, $t(62) = 2.987$, $p = .004$, Cohen's $d = .367$, 95% CI = [.119, .631], $BF_{10} = 7.61$, but the lowest-similarity lure was not lower than the criterion, $t(62) = -1.459$, $p = .15$, Cohen's $d = .184$, 95% CI = [-.432, .066], $BF_{10} = .377$. This suggests that the gist-based account of incidental memories cannot be fully dismissed under easy search conditions. For High-Load searches, however, false alarm rates for both lures differed from criterion: The highest-similarity lure was falsely recognized more often than chance, $t(62) = 3.970$, $p < .001$, Cohen's $d = .501$, 95% CI = [.236, .760], $BF_{10} = 119.39$, and the lowest-similarity lure was selected less often than chance, $t(62) = -4.213$, $p < .001$, Cohen's $d = .531$, 95% CI = [-.793, -.265], $BF_{10} = .253.94$. These results suggest that observers retained some perceptual features, even when making incorrect memory decisions.

Discussion

Experiment 2 replicated and extended the results of Experiment 1, providing more conclusive evidence that higher search difficulty produces more perceptually rich incidental memories for nontarget items. Using a single- versus multiple-target RSVP search task, we found that veridical memories and the distribution of false alarms revealed enhanced memories under more challenging, multiple-target, search conditions. By using RSVP, Experiment 2 confirmed that difficulty-enhanced distractor memories are not an artifact of eye movements and fixation durations. Instead, these results suggest that representations incidentally encoded in

VLTM increase in detail as the attentional demands of the search task increase.

General discussion

In many visual search tasks, observers' goals are to quickly locate targets while minimizing the attention directed to distracting nontarget objects. Although processing nontargets decreases search efficiency, research often shows that observers nevertheless encode and store representations of nontarget objects, despite *actively* trying to ignore them (e.g., Castelhamo & Henderson, 2005; Hollingworth & Henderson, 2002; Williams, 2010; Williams, Henderson, & Zacks, 2005). This incidental encoding is exacerbated when search is challenging and nontarget objects are hard to ignore (e.g., Hout & Goldinger, 2010, 2012; Thomas & Williams, 2014). Across two experiments, we investigated the precision of incidentally encoded memory representations following easy and challenging visual search tasks. Specifically, we manipulated the number of search targets to produce different levels of search difficulty. We then tested participants' memories for the distracting (i.e., to-be-ignored) search information by requiring participants to choose the studied exemplar from a lineup of within-category lures, each of which had a documented similarity relationship to the old item. Across both experiments, participants were more likely to remember nontarget exemplars encountered in challenging search contexts, even when they could not self-select the search items to process (Experiment 2). Moreover, memory errors revealed that even when recognition failed, participants were more likely to select lures with higher similarity relationships to the previously-seen exemplar. These results replicate and extend prior research (Hout & Goldinger, 2010, 2012; Thomas & Williams, 2014), showing that memories incidentally formed during challenging visual search are characterized by detailed, rather than gist-like, representations.

The fidelity of long-term memory representations is often studied using continuous report methods in VLTM (e.g., Brady et al., 2013b). These methods allow researchers to quantify the precision with which observers remember the color of studied objects. Although continuous report methods are challenging with stimuli that are not color-coded (as in the present study), we used MDS similarity rankings and "lineups" comprised of up to 16 within-category objects to more closely approximate object-level precision in incidental memory. With these methods, we found that object representations encoded during challenging searches are not sparse in detail or characterized by category-only information. In addition to producing higher hit rates, challenging searches produced false alarm profiles consistent with precise memory representations: Observers were more likely to select highly

similar lures than those with weaker similarity relationships to the viewed exemplar, just as observers in VLTM studies are more likely to (incorrectly) select colors similar to those they studied.

Our results are consistent with recent findings from Antonelli and Williams (2017), who found that task-relevant perceptual features influence how object representations are structured in VLTM. Specifically, they found that when color was a task-relevant feature, incidentally encoded object representations that shared color tended to interfere with one another, but not with representations of a different color. Similarly, we observed that lures with high perceptual overlap, as operationalized via MDS rankings, were more likely to interfere with the correct recognition of incidentally encoded exemplars, and that this effect was greater for objects encountered during challenging visual search.

It is important to note, however, that although our study used a stringent memory test, particularly in Experiments 1b and 2, alternative forced-choice (AFC) tests have been criticized as a measure of VLTM storage capacity, as they may rely on familiarity processes rather than recollection (Yonelinas, 2002). Old-new recognition (ONR) probes, by contrast, may force observers to query memory for additional details of the encoded event, possibly reducing the likelihood of familiarity-based recognition judgments (Cunningham et al., 2015). For instance, Draschkow, Reinecke, Cunningham, and Vö (2018) recently used ONR probes to show that sensitivity decreased for "state" lures (e.g., an *old* object presented in a different physical state) relative to exemplar lures (e.g., a different category-exemplar relative to the *old* object) for incidentally encoded objects, although such a difference was not observed in previous work using 2-AFC probes (e.g., Brady et al., 2008; Cunningham et al., 2015). While it is possible that observers are more likely to query memory for a specific memory trace when probed by ONR relative to AFC probes, prior research suggests that AFC tests populated with within-category lures encourage participants to retrieve detailed memories (e.g., Guerin, Robbins, Gilmore, & Schacter, 2012). As such, we suggest that observers in the current study were not relying on familiarity processes to make their recognition decisions. Across both memory tests (i.e., 4-AFC and 16-AFC), observers made systematic errors that were biased towards lures that shared higher perceptual overlap with the old item. This exemplar-lure similarity bias suggests that the perceptual details of the object were retrieved even when a correct recognition was not made.

The present results suggest that search difficulty impacts attention allocation strategies, which themselves impact what is remembered from search events. Search difficulty, however, can be a relative concept. For example, although the results from Experiment 1 partially supported our prediction that difficult searches would improve incidental memory, this effect

only emerged in simple effects analyses; the predicted interaction was not reliable. This may be because even the challenging conditions were not terribly challenging: Search hits exceeded exceed 96%, even under High Load. By contrast, Experiment 2 had a wider range of search performance across difficulty conditions and we observed the predicted interaction. These results suggest that difficulty-enhanced distractor encoding requires some level of objective, and potentially subjective, difficulty before observers adapt their attention allocation strategies. For example, observers in Experiment 2 may have adopted a more “active” cognitive strategy during difficult trials (Smilek, Enns, Eastwood, & Merikle, 2006), eliciting close inspection of each item and, therefore, incidental encoding. Although research suggests that these strategies can be subject to cognitive control (Madrid & Hout, 2019), the present results are agnostic regarding the conscious control over trial-by-trial shifts in attention allocation.

Our findings are consistent with similar effects in the prospective memory (PM) literature. In PM, individuals complete an ongoing task (e.g., lexical decision) while simultaneously monitoring for target items (e.g., specific words). In these paradigms, perceptions of *global* “task-set” difficulty (i.e., how challenging it is to detect targets in the context of the ongoing task) have been extensively reported to alter *ongoing* task response times (e.g., Hicks, Marsh, & Cook, 2005; Marsh, Hicks, & Cook, 2005; Marsh, Cook, & Hicks, 2006). For instance, lexical decision times in PM tasks increase when participants *expect* PM target detection to be difficult, but not when they expect it to be easy, even if task difficulty is not directly manipulated (Lourenço, Hill, & Maylor, 2015). Similarly, it is possible that perceptions of task difficulty may influence how observers adopt search strategies prior to search. Active search strategies likely involve the allocation of additional attentional resources to process items in the search array (Smilek et al., 2006), and object encoding is incidentally benefited from these additional resources (Guevara Pinto & Papesh, 2019). This suggests that search strategies may be one potential mechanism responsible for difficulty-enhanced nontarget memory; observers may allocate greater attention during conditions that are subjectively expected to be difficult. Future research may examine how search strategies are affected by perceived search difficulty, and how such strategies subsequently impact incidental object encoding.

Lastly, researchers have recently debated the nature of object representations stored in VLTm. Two contrasting views have proposed that real-world objects are stored either as bound, unitary units (Balaban, Assaf, Mier, & Luria, 2019) or as a set of independent and separable features (Brady, Konkle, Alvarez, & Oliva, 2013a; Utochkin & Brady, 2019). While the present study was not designed to contribute to this debate, we suggest that our results are

consistent with the “independent features” account of object representations. Across each experiment, observers were more likely to falsely recognize lures with high, relative to low, perceptual overlap with the encoded object, indicating that some of the object’s features were stored in memory, while others were not (preventing correct recognition). This suggests that an object’s features were identifiable in memory even when the exemplar itself was not. Importantly, the exemplar-lure similarity effect was modulated by task difficulty: The effect size increased as the encoding task became more difficult. Task difficulty may also explain why some evidence favors bound representations, while other evidence favors independent features. For example, Brady et al. (2013a, Exp. 1) and Konkle et al. (2010a, 2010b) focused participants’ attention during encoding with exemplar repetition detection tasks, which is similar to an *n-back* task. Others (e.g., Exp. 2 in Brady et al., 2013a; Balaban et al., 2019), however, have used object size judgments, in which participants determine whether an object is larger/smaller than a predefined size. Differences in the difficulty or cognitive demand of these two tasks may impact encoding and explain disparate findings. Future research into the nature of object representations may benefit from directly manipulating the attentional demands at encoding, as it possible that *less demanding* tasks may yield bound object representations, while demanding tasks produce evidence for independent features.

In conclusion, observers retain a great deal of visual information from visual search, even if they are not intentionally trying to do so (Castelhano & Henderson, 2005; Draschko et al., 2014; Hollingworth, 2004, 2006; Hollingworth & Henderson, 2002; Josephs et al., 2016; Williams, 2010). These incidental memories are enhanced when searches are challenging, relative to easy (Hout & Goldinger, 2010, 2012; Thomas & Williams, 2014). In the present study, we replicated this finding, and showed that memory representations for objects encoded during difficult visual search are characterized by richly detailed traces. Rather than simply guessing when recognition fails, observers systematically “recognized” lures that shared higher similarity relationships with encoded objects, particularly those encoded during difficult search trials. We suggest that observers adopt more active, attentionally demanding cognitive strategies during challenging search, forcing observers to process search items more exhaustively. Although challenging search should force observers to focus more attention on the targets, our results suggest that, ironically, observers become more likely to process the nontargets too.

Acknowledgements We thank Emily Green for her efforts at helping conceptualize this project, and for overseeing all data collection at NMSU as part of her NMSU Honors College Thesis project. We also thank Jessica Budd and Cameron Grimball for their assistance with data collection.

Appendix A

Table 4 Object categories used across experiments with average multidimensional vector inter-item distances

| Category name | Avg. inter-item distance | Category name | Avg. inter-item distance | Category name | Avg. inter-item distance |
|----------------|--------------------------|--------------------|--------------------------|-------------------|--------------------------|
| Abacus | 0.68 | Ball | 0.70 | Computer Key | 0.67 |
| Airplane | 0.55 | Balloon | 0.75 | Cookie | 0.54 |
| Apple | 0.80 | Barbie Doll | 0.79 | Cooking Pan | 0.75 |
| Toy Soldier | 0.67 | Baseball Card | 0.57 | Cooking Pot | 0.70 |
| Axe | 0.68 | Basket | 0.70 | Baby Crib | 0.61 |
| Russian Doll | 0.63 | Bathing Suit | 0.70 | Cup and Saucer | 0.64 |
| Baby Carriage | 0.76 | Teddy Bear | 0.48 | Cushion | 0.64 |
| Backpack | 0.79 | Beanbag Chair | 0.62 | Decorative Screen | 0.72 |
| Baby Playpen | 0.69 | Bed | 0.68 | Desk | 0.68 |
| Bird House | 0.70 | Beer Mug | 0.74 | Dog | 0.62 |
| Bowling Pin | 0.62 | Bell | 0.76 | Doll House | 0.51 |
| Carabiner | 0.63 | Bench | 0.66 | Domino | 0.70 |
| Chocolate | 0.67 | Bicycle | 0.73 | Donut | 0.70 |
| Circuit Board | 0.66 | Dollar Bill | 0.63 | Doorknob | 0.75 |
| Coffemaker | 0.75 | Bird | 0.67 | Dresser | 0.70 |
| Cooler | 0.73 | Bongo | 0.77 | Dumbbell | 0.70 |
| Door Knocker | 0.71 | Bonsai Tree | 0.55 | Earrings | 0.62 |
| Feather | 0.66 | Boot | 0.65 | Easter Egg | 0.62 |
| Fruit Parfait | 0.63 | Bottle | 0.69 | Gym Equipment | 0.78 |
| Board Game | 0.67 | Bowl | 0.59 | Electric Fan | 0.71 |
| Eyeglasses | 0.69 | Bowtie | 0.65 | Fish Hook | 0.62 |
| Plant | 0.58 | Bread | 0.67 | Flag | 0.64 |
| Hair Dryer | 0.68 | Broom | 0.77 | Flashlight | 0.68 |
| Baby Chair | 0.66 | Bucket | 0.73 | Painting Frame | 0.70 |
| Highlighter | 0.69 | Butterfly | 0.47 | Frisbee | 0.74 |
| Horseshoe | 0.65 | Button | 0.76 | Video Game | 0.63 |
| Ice Skates | 0.72 | Cake | 0.76 | Garbage Can | 0.70 |
| Laptop | 0.75 | Calculator | 0.67 | Gift | 0.64 |
| Leatherman | 0.65 | Camcorder | 0.79 | Glove | 0.64 |
| Mailbox | 0.78 | Camera | 0.71 | Goggles | 0.62 |
| Protractor | 0.68 | Candle Holder | 0.70 | Golfball | 0.65 |
| Mattress | 0.67 | Car | 0.78 | Grill | 0.70 |
| Microphone | 0.74 | Cat | 0.66 | Guitar | 0.65 |
| Pasta | 0.72 | Ceiling Fan | 0.68 | Hair Brush | 0.67 |
| Printer | 0.67 | Cell Phone | 0.76 | Hammer | 0.68 |
| TV Controller | 0.71 | Chair | 0.71 | Hand Bag | 0.54 |
| Toy Robot | 0.64 | Cheese | 0.66 | Hand Gun | 0.72 |
| Router | 0.67 | Cheese Grater | 0.80 | Clothes Hanger | 0.67 |
| Scooter | 0.75 | Cherub Statue | 0.67 | Hat | 0.68 |
| Shower Curtain | 0.72 | Chess Board | 0.69 | Headband | 0.69 |
| Shredder | 0.67 | Christmas Stocking | 0.58 | Headphones | 0.67 |
| Sink | 0.67 | Christmas Ornament | 0.61 | Helmet | 0.69 |
| Skateboard | 0.70 | Cigarette Pack | 0.75 | Hourglass | 0.72 |
| Sleeping Bag | 0.72 | Clock | 0.64 | Jacket | 0.74 |

Table 4 (continued)

| | | | | | |
|----------------|------|---------------------|------|----------------|------|
| Typewriter | 0.65 | Coat Rack | 0.62 | Jack-o-Lantern | 0.73 |
| Vacuum | 0.69 | Coffee Mug | 0.53 | Juice | 0.64 |
| Wall Lamp | 0.63 | Coin | 0.68 | Kayak | 0.65 |
| Box | 0.63 | Dog Collar | 0.72 | Key | 0.76 |
| Bagel | 0.72 | Compass | 0.78 | Keyboard | 0.73 |
| Keychain | 0.59 | Water Pitcher | 0.67 | Stapler | 0.67 |
| Knife | 0.71 | Pizza | 0.61 | Stool | 0.72 |
| Lamp | 0.77 | Poker Card | 0.63 | Suit | 0.65 |
| Lantern | 0.72 | Power Strip | 0.77 | Suitcase | 0.66 |
| Lawn Mower | 0.76 | Radio | 0.69 | Table | 0.64 |
| Leaf | 0.72 | Razor | 0.79 | Tape | 0.72 |
| Hawaiian Lei | 0.65 | Record Player | 0.68 | Telescope | 0.69 |
| License Plate | 0.66 | Ring | 0.73 | Tennis Racquet | 0.71 |
| Lipstick | 0.65 | 3-Ring Binder | 0.74 | Tent | 0.59 |
| Lock | 0.68 | Road Sign | 0.70 | Toilet Seat | 0.65 |
| Magazine | 0.63 | Rock | 0.59 | Tongs | 0.62 |
| Makeup Kit | 0.63 | Roller Skates | 0.79 | Toothpaste | 0.77 |
| Mask | 0.68 | Rosary | 0.64 | Toy Horse | 0.71 |
| Meat | 0.65 | Rug | 0.72 | Toy Rabbit | 0.72 |
| Microscope | 0.64 | Saddle | 0.67 | Toy Train | 0.70 |
| Microwave | 0.61 | Salt/Pepper Shakers | 0.57 | Tree | 0.72 |
| Motorcycle | 0.72 | Sandwich | 0.81 | Tricycle | 0.54 |
| Mp3 Player | 0.57 | Scale | 0.67 | Trophy | 0.66 |
| Muffin | 0.62 | Scissors | 0.69 | Trumpet | 0.72 |
| Mushroom | 0.69 | Scrunchie | 0.49 | Trunk | 0.68 |
| Nail Polish | 0.59 | Seashell | 0.66 | Turtle | 0.60 |
| Necklace | 0.69 | Shoe | 0.81 | Television | 0.70 |
| Necktie | 0.68 | Baby Sippy Cup | 0.71 | Umbrella | 0.53 |
| Nunchaku | 0.68 | Snow Globe | 0.61 | Vase | 0.67 |
| Handheld Fan | 0.63 | Socks | 0.63 | Watch | 0.67 |
| Pants | 0.70 | Soda Can | 0.68 | Water Gun | 0.81 |
| Poolside Chair | 0.60 | Sofa | 0.60 | Wig | 0.66 |
| Pen | 0.63 | Speaker | 0.71 | Wind Chime | 0.76 |
| Phone | 0.65 | Spoon | 0.67 | Wine Glass | 0.67 |
| Tobacco Pipe | 0.65 | Stamp | 0.40 | Yarn | 0.63 |

Appendix B. Visual search performance

Experiments 1a and 1b Search accuracy was examined in 3 (Working-Memory Load: High, Medium, Low) \times 2 (Number of Targets) repeated-measures ANOVA on search hit rates. The main effect of WM Load in Experiment 1a, $F(1.82, 109.06) = 58.13, p < .001, \eta_p^2 = .49, BF_{10} = 5.272e+18$, revealed higher hits rate on Low-Load trials ($M = .99, SE = .001$), followed by Medium-Load ($M = .97, SE = .004$) and High-Load ($M = .96, SE = .005$) trials. The main effect of Number of Targets, $F(1, 60) = 76.56, p < .001, \eta_p^2 = .56, BF_{10} = 3.614e+7$, is interpreted in the context of the reliable interaction, $F(2, 120) = 11.58, p < .001, \eta_p^2 = .16, BF_{10} = 247.5$: Search hits for one target ($M = .98, SE = .002$) were higher than two targets ($M = .97, SE = .004$), but only when searching under High or Medium WM Loads, both $ps < .05$. This was replicated in Experiment 1b, where a main effect of Number of Targets, $F(1, 63) = 45.06, p < .001, \eta_p^2 = .42, BF_{10} = 840,223$, and a reliable interaction was observed, $F(1.82, 114.39) = 7.86, p < .001, \eta_p^2 = .11, BF_{10} = 21.22$. The main

effect of WM Load in Experiment, $F(2, 126) = 48.05, p < .001, \eta_p^2 = .43, BF_{10} = 1.276e+15$, again revealed higher hits rate on Low Load trials ($M = .99, SE = .002$), followed by Medium Load ($M = .98, SE = .003$) and High Load ($M = .97, SE = .004$) trials.

Experiment 2 A paired-sample t -test was conducted on the search hit rate across Load conditions in Experiment 2. The effect of Load is reliable, $t(62) = 10.76, p < .001$, Cohen's $d = 1.36, BF_{10} = 7.79e+12$, indicating that more targets were detected in the Low- ($M = 97.1, SE = .01$), relative to High- ($M = 88.7, SE = .01$), Load condition. A similar effect was observed in false alarms, $t(62) = 2.82, p = .007$, Cohen's $d = 0.36, BF_{10} = 4.98$, with lower false alarms in the Low- ($M = .02, SE = .004$) than in the High-Load condition ($M = .04, SE = .01$).

Appendix C. Target recognition memory

Experiments 1a and 1b The proportion of *correct* recognition choices was examined in one-way repeated-measures

ANOVAs, comparing recognition for targets to nontargets encountered during Low-, Medium-, and High-Load searches. A reliable effect was observed in Experiment 1a, $F(3, 180) = 437.5, p < .001, \eta_p^2 = .88, BF_{10} = 1.475e+81$, in which recognition differed across all items, with target objects recognized the best ($M = .94, SE = .01$). More importantly, nontargets objects encountered during High-Load search ($M = .63, SE = .02$) were recognized better than nontargets encountered during Medium- ($M = .58, SE = .02$) and Low-Load search ($M = .48, SE = .01$). Similarly, a reliable effect was observed in Experiment 1b, $F(2.56, 161.24) = 660.88, p < .001, \eta_p^2 = .91, BF_{10} = 1.574e+101$, with target objects recognized the best ($M = .80, SE = .02$), while nontargets encountered during High-Load search ($M = .31, SE = .02$) were recognized more often than nontargets encountered during Medium- ($M = .26, SE = .01$) and Low-Load search ($M = .22, SE = .01$), both $ps < .001$. Nontargets from Medium-Load search were recognized better than nontargets from Low-Load search, $p = .035$.

Appendix D. Target recognition errors

Experiments 1a and 1b In Experiment 1a, each 4-AFC comparison included lures designated as relatively similar, moderately similar, and relatively dissimilar to the studied exemplar. Similarly, in Experiment 1b, lures were rank-ordered, from 1 (*highest similarity*) to 15 (*lowest similarity*). False alarms could thus be examined as a function of the perceptual overlap between studied exemplars and categorically related lures in both experiments. Separate one-way repeated-measures ANOVAs analyzed the impact of exemplar-lure similarity on the proportion of false alarms in Experiments 1a and 1b. Both experiments revealed reliable effects of lure similarity (Experiment 1a: $F(1.5, 90.02) = 25.79, p < .001, \eta_p^2 = .30, BF_{10} = 7.141e+9$; Experiment 1b: $F(9.92, 624.92) = 13.45, p < .001, \eta_p^2 = .18, BF_{10} = 4.035e+27$), indicating that when recognition of target objects failed, observers were more likely to commit false alarms to lures of high similarity with the target.

Appendix E

Table 5 Statistics from omnibus analyses on raw data values for proportion of false alarms as a function of search difficult and exemplar-lure similarity in Experiments 1a, 1b, and 2

| Experiment | Effect | Frequentist | | | | Bayesian | |
|-------------|----------------|--------------|-------|--------|------------|------------------|---------------------------|
| | | df | F | p | η_p^2 | BF ₁₀ | Evidence favoring effect? |
| 1a (4-AFC) | | | | | | | |
| | WM Load (WM) | 2,120 | -0.47 | 1.00 | 0.00 | 0.021 | No evidence |
| | Similarity (S) | 2,120 | 18.47 | < .001 | 0.24 | 5.183e+8 | Strong evidence |
| | WM x S | 4,240 | 1.65 | 0.152 | 0.03 | 0.55 | No evidence |
| 1b (16-AFC) | | | | | | | |
| | WM Load (WM) | 1,63 | 1.53 | .22 | 0.02 | 0.004 | No evidence |
| | Similarity (S) | 10.06,633.7 | 6.45 | < .001 | 0.09 | 1.167e+10 | Strong evidence |
| | WM x S | 18.4, 1157.3 | 1.14 | 0.30 | 0.01 | 0.005 | No evidence |
| 2 (16-AFC) | | | | | | | |
| | WM Load (WM) | 1,62 | 0.18 | 0.68 | 0.03 | 0.057 | No evidence |
| | Similarity (S) | 9.94,616.24 | 6.93 | < .001 | 0.10 | 2.10e+9 | Strong evidence |
| | WM x S | 10.07,624.4 | 1.66 | 0.058 | 0.03 | 0.167 | No evidence |

Note: BF₁₀ reflects the evidence favoring the inclusion of the effect relative to the null model

Table 6 Statistics from simple effect analyses on *raw* data values for the effect of exemplar-lure similarity across levels of search difficulty in Experiments 1a, 1b, and 2

| Experiment | Search | Frequentist | | | | Bayesian | | | Evidence for lure similarity effect? |
|-------------|----------------|-------------|----------|------------|------------|------------------|------------------|-----------------|--------------------------------------|
| | | <i>F</i> | <i>p</i> | η_p^2 | CI | BF ₁₀ | BF ₀₁ | | |
| 1a (4-AFC) | | | | | | | | | |
| | Low WM Load | 1.33 | .27 | .02 | .000, .091 | .311 | 3.22 | No evidence | |
| | Medium WM Load | 9.21 | < .001 | .13 | .034, .239 | 6250 | 1.600e-4 | Strong evidence | |
| | High WM Load | 6.84 | .002 | .102 | .026, .223 | 348 | .003 | Strong evidence | |
| 1b (16-AFC) | | | | | | | | | |
| | Low WM Load | 1.89 | .024 | .029 | .001, .041 | .172 | 5.82 | No evidence | |
| | Medium WM Load | 2.98 | <.001 | .045 | .009, .067 | 69.85 | .014 | Strong evidence | |
| | High WM Load | 3.09 | <.001 | .047 | .012, .072 | 137.4 | .007 | Strong evidence | |
| 2 (16-AFC) | | | | | | | | | |
| | Low WM Load | .2.99 | <.001 | .046 | .011, .060 | 77.29 | 0.013 | Strong evidence | |
| | High WM Load | 5.42 | <.001 | .08 | .037, .102 | 9.350e+ | 1.069e-8 | Strong evidence | |

Note: BF₁₀ reflects strength of evidence for the alternative, BF₀₁ reflects strength of evidence for the null
 CI reflect 95% confidence intervals for partial eta-square

Appendix F

Table 7 Transformed false-alarm data value means (± 1 SEM in parentheses) across conditions and exemplar-lure similarity in Experiments 1a, 1b, and 2

| Experiment | Lure similarity | Low Load | Moderate Load | High Load |
|-------------|-----------------|-------------|---------------|-------------|
| 1a (4-AFC) | | | | |
| | Similar | .625 (.015) | .670 (.016) | .667 (.018) |
| | Moderate | .614 (.011) | .616 (.018) | .598 (.019) |
| | Dissimilar | .596 (.013) | .527 (.021) | .553 (.018) |
| 1b (16-AFC) | | | | |
| | 1st rank | .269 (.010) | .296 (.011) | .288 (.012) |
| | 2nd rank | .272 (.014) | .240 (.014) | .267 (.014) |
| | 3rd rank | .246 (.011) | .255 (.012) | .250 (.014) |
| | 4th rank | .235 (.013) | .267 (.014) | .251 (.013) |
| | 5th rank | .238 (.014) | .235 (.012) | .243 (.014) |
| | 6th rank | .256 (.013) | .242 (.013) | .243 (.012) |
| | 7th rank | .249 (.011) | .241 (.013) | .244 (.013) |
| | 8th rank | .250 (.011) | .243 (.011) | .231 (.013) |
| | 9th rank | .261 (.010) | .238 (.013) | .231 (.014) |
| | 10th rank | .229 (.013) | .237 (.014) | .238 (.014) |
| | 11th rank | .240 (.014) | .229 (.011) | .242 (.013) |
| | 12th rank | .229 (.010) | .231 (.012) | .227 (.014) |
| | 13th rank | .227 (.012) | .219 (.013) | .196 (.014) |
| | 14th rank | .229 (.013) | .250 (.011) | .193 (.015) |
| | 15th rank | .219 (.013) | .212 (.011) | .215 (.016) |
| 2 (16-AFC) | | | | |
| | 1st rank | .279 (.014) | n/a | .304 (.014) |
| | 2nd rank | .285 (.014) | n/a | .249 (.016) |

Table 7 (continued)

| Experiment | Lure similarity | Low Load | Moderate Load | High Load |
|------------|-----------------|-------------|---------------|-------------|
| | 3rd rank | .242 (.016) | n/a | .283 (.015) |
| | 4th rank | .225 (.016) | n/a | .244 (.015) |
| | 5th rank | .237 (.017) | n/a | .261 (.014) |
| | 6th rank | .248 (.015) | n/a | .233 (.014) |
| | 7th rank | .245 (.013) | n/a | .223 (.015) |
| | 8th rank | .219 (.015) | n/a | .249 (.014) |
| | 9th rank | .238 (.015) | n/a | .199 (.017) |
| | 10th rank | .214 (.014) | n/a | .186 (.016) |
| | 11th rank | .199 (.015) | n/a | .226 (.015) |
| | 12th rank | .205 (.016) | n/a | .217 (.017) |
| | 13th rank | .214 (.013) | n/a | .211 (.015) |
| | 14th rank | .217 (.015) | n/a | .200 (.014) |
| | 15th rank | .221 (.014) | n/a | .170 (.016) |

Note: 1st rank reflects lure of highest perceptual similar to encoded exemplar

References

- Alexander, R. G., & Zelinsky, G. J. (2011). Visual similarity effects in categorical search. *Journal of Vision, 11*, 1–15.
- Alexander, R. G., & Zelinsky, G. J. (2012). Effects of part-based similarity on visual search: The Frankenbear experiment. *Vision Research, 54*, 20–30.
- Antonelli, K. B., & Williams, C. C. (2017). Task-relevant perceptual features can define categories in visual memory too. *Memory & Cognition, 45*, 1295–1305.
- Balaban, H. Assaf, D., Arad Meir, M., & Luria, R. (2019). Different features of real-world objects are represented in a dependent manner in long-term memory. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000716>
- Berman, M. G., Hout, M. C., Kardan, O., Hunter, M., Yourganov, G., Henderson, J. M., Hanayik, T., Karimi, H., & Jonides, J. (2014). The perception of naturalness correlates with low-level visual features of environmental scenes. *PLoS ONE, 9*: e114572. doi: <https://doi.org/10.1371/journal.pone.0114572>.
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Science, USA, 105*, 14325–14329.
- Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2013a). Real-world objects are not represented as bound units: Independent forgetting of different object details from visual memory. *Journal of Experimental Psychology: General, 142*, 791–808.
- Brady, T. F., Konkle, T., Gill, J., Oliva, A., & Alvarez, G. A. (2013b). Visual long-term memory has the same limit on fidelity as visual working memory. *Psychological Science, 24*(6), 981–990.
- Castelhano, M. S., & Henderson, J. M. (2005). Incidental visual memory for objects in scenes. *Visual Cognition, 12*(6), 1017–1040.
- Coburn, A., Kardan, O., Kotabe, H., Steinberg, J., Hout, M. C., Robbins, A., MacDonald, J., Hayn-Leichsenring, G., & Berman, M. (2019). Psychological responses to natural patterns in architecture. *Journal of Environmental Psychology, 62*, 133–145. doi: <https://doi.org/10.1016/j.jenvp.2019.02.007>.
- Cunningham, C. A., Yassa, M. A., & Egeth, H. E. (2015). Massive memory revisited: Limitations on storage capacity for object details in visual long-term memory. *Learning & Memory, 22*(11), 563–566.
- Draschkow, D., Wolfe, J. M., & Vö, M. L. H. (2014). Seek and you shall remember: Scene semantics interact with visual search to build better memories. *Journal of Vision, 14*(8):10, 1–18.
- Draschkow, D., Reinecke, S., Cunningham, C. A., & Vö, M. L. H. (2018). The lower bounds of massive memory: Investigating memory for object details after incidental encoding. *Quarterly Journal of Experimental Psychology*. doi: <https://doi.org/10.1177/1747021818783722>
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433–458.
- Godwin, H., 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: <https://doi.org/10.3758/s13423-013-0547-4>
- Goldinger, S. D. (1998). Echoes or echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.
- Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers, 26*(4), 381–386.
- Guerin, S. A., Robbins, C. A., Gilmore, A. W., & Schacter, D. L. (2012). Retrieval failure contributes to gist-based false recognition. *Journal of Memory and Language, 66*, 68–78. doi: <https://doi.org/10.1016/j.jml.2011.07.002>
- Guevara Pinto, J. D., & Papesch, M. H. (2019). Incidental memory following rapid object processing: The role of attention allocation strategies. *Journal of Experimental Psychology: Human Perception and Performance, 45*(9), 1174–1190.
- Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). Task interference in time-based, event-based, and dual intention prospective memory conditions. *Journal of Memory and Language, 53*, 430–444.
- Hollingworth, A. (2004). Constructing visual representations of natural scenes: The roles of short- and long-term visual memory. *Journal of Experimental Psychology: Human Perception & Performance, 30*, 519–537.
- Hollingworth, A. (2006). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition, 14*, 781–807.
- Hollingworth, A., & Henderson, J. M. (2002). Accurate visual memory for previously attended objects in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 28*(1), 113–136.

- Homa, D., Blair, M., McClure, S. M., Medema, J., & Stone, G. (2018). Learning concepts when instances never repeat. *Memory & Cognition*, 47(3), 395–411.
- Horst, J. S., & Hout, M. C. (2015). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behavior Research Methods*, 48, 1393–1409. doi: <https://doi.org/10.3758/s13428-015-0647-3>.
- Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception, & Psychophysics*, 72, 1267–1282.
- 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(1), 90–112.
- 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, 128–149. doi: <https://doi.org/10.3758/s13414-014-0764-6>.
- Hout, M. C., & Goldinger, S. D. (2016). SpAM is convenient, but also satisfying: Reply to Verheyen et al. (2016). *Journal of Experimental Psychology: General*, 3, 383–387. doi: <https://doi.org/10.1037/xge000017>.
- Hout, M. C., Papesch, M. H., & Goldinger, S. D. (2012). Multidimensional scaling. Wiley Interdisciplinary Reviews (WIREs): Cognitive Science, 4, 93–103. doi: <https://doi.org/10.1002/wcs.1203>
- Hout, M. C., Goldinger, S. D., & Ferguson, R. W. (2013). The versatility of SpAM: A fast, efficient spatial method of data collection for multidimensional scaling. *Journal of Experimental Psychology: General*, 142, 256–281. doi: <https://doi.org/10.1037/a0028860>.
- Hout, M. C., Goldinger, S. D., & Brady, K. J. (2014). MM-MDS: A multidimensional scaling database with similarity ratings for 240 object categories from the Massive Memory picture database. *PLoS one*, 9(11), e112644
- 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: <https://doi.org/10.3758/s13414-015-1010-6>.
- JASP Team (2018). JASP (Version 0.8.5)
- Josephs, E. L., Draschcow, D., Wolfe, J. M., & Vö, M. L. H. (2016). Gist in time: Scene semantics and structure enhance recall of searched objects. *Acta Psychologica*, 169, 100–108.
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010a). Conceptual distinctiveness supports detailed visual long-term memory for real-world objects. *Journal of Experimental Psychology: General*, 139(3), 558–578.
- Konkle, T., Brady, T. F., Alvarez, G. A., & Oliva, A. (2010b). Scene memory is more detailed than you think: The role of categories in visual long-term memory. *Psychological Science*, 21(11), 1551–1556.
- Lourenço, J. S., Hill, J. H., & Maylor, E. A. (2015). Too easy? The influence of task demands conveyed tacitly on prospective memory. *Frontiers in Human Neuroscience*, 9, 1–6.
- Madrid, J., & Hout, M. C. (2019). Examining passive and active strategies on search behavior during hybrid visual memory search: Evidence from eye tracking. *Cognitive Research: Principles and Applications*, 4(1), 39.
- Marsh, R. L., Hicks, J. L., & Cook, G. I. (2005). On the relationship between effort toward an ongoing task and cue detection in event-based prospective memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(1), 68–75.
- Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). Task interference from event-based intentions can be material specific. *Memory & Cognition*, 34(8), 1636–1643.
- Menneer, T., Barrett, D. J. K., Phillips, L., Donnelly, N., & Cave, K. R. (2007). Costs in searching for two targets: Dividing search across target types could improve airport security screening. *Applied Cognitive Psychology*, 21, 915–932.
- Menneer, T., Cave, K. R., & Donnelly, N. (2009). The cost of search for multiple targets: Effects of practice and target similarity. *Journal of Experimental Psychology: Applied*, 15, 125–139.
- Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10(1), 104–114
- Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, 43(1), 25–53.
- Papesch, M. H., & Goldinger, S. D. (2010). A multidimensional scaling analysis of own- and cross-race face spaces. *Cognition*, 116(2), 283–288.
- Pedely, L., Cohen-Levine, S., & Shevell, S. K. (1985). Developmental changes in face processing: Results from multidimensional scaling. *Journal of Experimental Child Psychology*, 39, 421–236.
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5), 356–374.
- 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.
- Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, 6, 156–163.
- Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237, 1317–1323
- Smilek, D., Enns, J. T., Eastwood, J. D. & Merikle, P. M. (2006) Relax! Cognitive strategy influences visual search. *Visual Cognition*, 14, 543–64.
- Standing, L. (1973). Learning 10,000 pictures. *Quarterly Journal of Experimental Psychology*, 25, 207–222.
- Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2,500 visual stimuli. *Psychonomic Science*, 19, 73–74.
- Theeuwes, J. (1994a). Stimulus-driven capture and attention set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human perception and performance*, 20(4), 799–806.
- Theeuwes, J. (1994b). Endogenous and exogenous control of visual selection. *Perception*, 23(4), 429–440.
- Thomas, M. D., & Williams, C. C. (2014). The target effect: Visual memory for unnamed search targets. *Quarterly Journal of Experimental Psychology*, 67(11), 2090–2104.
- Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, 17, 401–419.
- Utochkin, I. S., & Brady, T. F. (2019). Independent storage of different features of real-world objects in long-term memory. *Journal of Experimental Psychology: General*. <https://doi.org/10.1037/xge0000664>
- Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology*, 43(2), 161–204.
- Vogt, S., & Magnussen, S. (2007). Long-term memory for 400 pictures on a common theme. *Experimental Psychology*, 54, 298–303.
- Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., Selker, R., Gronau, Q. F., Šmíra, M., Epskamp, S., Matzke, D., Rouder, J. N., & Morey, R. D. (2018). Bayesian inference for psychology: Part I: Theoretical advantages and practical ramifications. *Psychonomic Bulletin & Review*, 25, 35–57.
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Science*, 15(4), 160–168.
- Williams, C. C. (2010). Incidental and intentional visual memory: What memories are and are not affected by encoding task? *Visual Cognition*, 18, 1348–1367.

- Williams, C. C., Henderson, J. M., & Zacks, R. T. (2005). Incidental visual memory for targets and nontargets in visual search. *Perception & Psychophysics*, *67*, 816–827.
- Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 1–8.
- 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*(3), 419–433.
- Yang, H., & Zelinsky, G. J. (2009). Visual search is guided to categorically-defined targets. *Vision Research*, *49*(16), 2095–2103.
- Yonelinas, A. P. (2002). The nature of recollection and familiarity: a review of 30 years of research. *Journal of Memory and Language*, *46*, 441–517.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.