



How does searching for faces among similar-looking distractors affect distractor memory?

Geoffrey L. McKinley¹ · Daniel J. Peterson¹ · Michael C. Hout^{2,3}

Accepted: 9 February 2023 / Published online: 27 February 2023
© The Psychonomic Society, Inc. 2023

Abstract

Prior research has shown that searching for multiple targets in a visual search task enhances distractor memory in a subsequent recognition test. Three non-mutually exclusive accounts have been offered to explain this phenomenon. The mental comparison hypothesis states that searching for multiple targets requires participants to make more mental comparisons between the targets and the distractors, which enhances distractor memory. The attention allocation hypothesis states that participants allocate more attention to distractors because a multiple-target search cue leads them to expect a more difficult search. Finally, the partial match hypothesis states that searching for multiple targets increases the amount of featural overlap between targets and distractors, which necessitates greater attention in order to reject each distractor. In two experiments, we examined these hypotheses by manipulating visual working memory (VWM) load and target-distractor similarity of AI-generated faces in a visual search (i.e., RSVP) task. Distractor similarity was manipulated using a multidimensional scaling model constructed from facial landmarks and other metadata of each face. In both experiments, distractors from multiple-target searches were recognized better than distractors from single-target searches. Experiment 2 additionally revealed that increased target-distractor similarity during search improved distractor recognition memory, consistent with the partial match hypothesis.

Keywords Incidental memory · Similarity · RSVP · Visual search

Introduction

Imagine during a third-grade class field trip at a local aquarium that one student goes missing, requiring the teacher to perform a visual search for the lost child. While scanning the sea of children, the teacher must (to some degree) process the face of each student (i.e., the kids who are not missing) in order to determine that they are not the target of her search. One can appreciate that the difficulty of this task increases as the number of lost children increases. For example, a teacher searching for a single student can quickly scan and verify whether or not each viewed child matches the one

that she is looking for. In contrast, a teacher looking for three students who have wandered off has a much more onerous task, as the teacher needs to decide whether each kid before her matches one of three unique targets.

The exact nature of the decision-making processes involved in situations like this is not precisely understood. We consider three non-mutually exclusive accounts to help explain why searching for multiple targets is more challenging. One possibility is that the teacher would need to mentally compare each face before her to each of the missing students. In this case, searching for multiple students would be more effortful (and more time-consuming) than searching for a single student because more mental comparisons are required. We henceforth refer to this as the *mental comparison hypothesis*. Another possibility is the teacher would simply allocate more attention to each child when searching for multiple students because the searcher is metacognitively aware of the difficulty of this task. That is, when searching for multiple students, a teacher may anticipate a more difficult search task, causing them to allocate greater attention to each face. We will refer to this as the *attentional allocation*

✉ Geoffrey L. McKinley
Geoffmck59@gmail.com

¹ Department of Psychology, Skidmore College, 815 N. Broadway, Saratoga Springs, NY 12866, USA

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

³ National Science Foundation, Alexandria, VA, USA

hypothesis. Finally, it is important to remember that searching for multiple students will most likely also increase the number of features (e.g., hair color, nose shape) that are relevant. That is, unless the teacher is looking for an identical set of triplets, the missing students will vary in appearance. This variation will increase the extent to which the students who are not missing will elicit a partial match, because they are more likely to share features with one (or more) of the missing students. Thus, searching for multiple students may be more effortful (and time-consuming) because the increase in the number of overlapping features makes it more difficult to discern missing students from students who are not missing. We will refer to this as the *partial match hypothesis*.

Although these three hypotheses are not necessarily mutually exclusive, they imply different mechanisms, and as we describe below, make distinct predictions for downstream behaviors (e.g., memory). There is, however, one shared prediction: they all assert that searching for multiple faces is more effortful than only searching for one face. This increased effort could be due to the quantity of comparisons that are required, or it could be due to the intensity and extent of each comparison. Regardless, it stands to reason that the increased effort spent during search may come with an unintended benefit: improved incidental memory for the faces of the kids who are *not* missing. Below, we review relevant empirical data to detail how each of the three accounts explains both performance decrements (slower, poorer visual search) and enhancements (improved distractor memory).

Focusing first on the mental comparison hypothesis, research suggests that searching for multiple targets is more onerous than searching for a single target, as evidenced by decrements in search performance (Godwin, Menneer, Cave, & Donnelly, 2010a; Godwin, Menneer, Cave, Helman, et al., 2010b; Guevara Pinto et al., 2020; Guevara Pinto & Papesch, 2019; Hout & Goldinger, 2012; Menneer et al., 2012; Mestry et al., 2017; Walenchok et al., 2016). Although all of the three hypotheses make this prediction, the mental comparison hypothesis suggests that a greater VWM load requires one to make more mental comparisons between each target and distractor. In one series of experiments, Hout and Goldinger (2012) manipulated visual working memory (VWM) load by having participants search a display of objects for one, two, or three potential targets among a set of distractors. Afterwards, participants were given a surprise recognition test for the distractors they encountered earlier. The authors found that a greater VWM load slowed search times for the targets and increased visual fixations to the distractors. Importantly, though search performance suffered, incidental memory for the distractors was enhanced; as VWM load during search increased, subsequent recognition for the distractors similarly increased (see also Hout & Goldinger, 2010). To explain these data, the authors suggested that in a single-target search, each

distractor must be compared with only one target. However, in a multiple-target search, one must compare each distractor to each target, resulting in more mental comparisons. This explanation easily accounts for the effect of load on search because additional comparisons require greater effort. This explanation also explains the benefit in memory performance as a byproduct of this increased attentional effort. One may be concerned, however, that it is not the number of comparisons per se but rather the additional time required to perform the search that drives these effects. Hout and Goldinger (2010) addressed this potential confound by using a rapid serial visual presentation (RSVP) task to sequentially present each item (both targets and distractors) for a consistent duration. In such a paradigm, after the entire array has been presented, participants indicate whether one of the unique targets appeared or not. Their results showed that the load effect on distractor memory persisted (see also Guevara Pinto et al., 2020; Guevara Pinto & Papesch, 2019), indicating that this mnemonic benefit on distractor recognition cannot be attributed to differences in encoding time alone.

More recently, Guevara Pinto and Papesch (2019) proposed the *attentional allocation hypothesis*, a more strategic account arguing that encoding *intensity* is greater during multiple-target search because such search cues signal a more difficult task. As a result, participants allocate more attentional resources to the distractors than they would when searching for a single target. Indeed, these authors found that participants performed worse on a secondary peripheral item-detection task on search trials with multiple targets. This result suggests that participants allocated more attentional resources to the items in the search stream because participants *expected* a more difficult search trial when presented with multiple targets. Importantly, this explanation predicts better distractor encoding only in search trials in which participants can reasonably predict the difficulty of the upcoming search based on the search cue.

An alternative (though again, not mutually exclusive) account, the *partial match hypothesis*, focuses on the fact that as the number of search targets increases, so does the number of features that are shared between the targets and distractors (Lavelle et al., 2021; see also Guevara Pinto & Papesch, 2019). That is, a greater VWM load increases the functional similarity between the targets and distractors in visual search (Duncan & Humphreys, 1989). Such distractors are more likely to capture attention due to their target-evoking features, creating a more challenging and effortful search task. As a byproduct of this difficulty, these distractors are encoded to a greater extent, and thus are more likely to be remembered in a subsequent recognition test.¹

¹ It is worth noting that this partial match could stem from a lack of precision in the target template (Bravo & Farid, 2009), which might result in more consideration of a larger set of distractors. However, it is also possible that distractors may capture attention simply because

There is much research that underscores the importance of target-distractor similarity in visual search (e.g., Alexander & Zelinsky, 2011, 2012; Duncan & Humphreys, 1989; Mestry et al., 2017). For example, Mestry et al. (2017) found that as the similarity between the target faces and the distractor faces increased in a visual search task (similarity was achieved through face morphing), so did visual fixations on the distractors (see also Alexander & Zelinsky, 2011, Exp. 2). This suggests that distractors that are similar to the target capture more attention than less similar distractors. There is also some evidence showing that as the similarity of targets in a multiple-target search decreases, participants spend more time searching, and scan more areas of the display before making a response (Hout & Goldinger, 2015, Exp. 2C). This suggests that distractors can also capture more attention when the search template is less precise. However, the focus of much of the research that has manipulated target-distractor similarity (or target-target similarity) has been on visual search performance, and as such, has not examined distractor memory.

There is also some evidence to suggest that distractors that are similar to the target are better remembered than unrelated distractors (Thomas & Williams, 2014; Williams, 2010a; Williams, 2010b; Williams & Henderson, 2005). However, it should be noted that much of this research used an active visual search task in which all of the objects were presented simultaneously (i.e., not using a RSVP task that controls for presentation time). Therefore, participants were able to control where they guided their attention, and did not necessarily have equal exposure to all items across conditions. However, as mentioned above, memory is affected by load even in RSVP tasks (Guevara Pinto et al., 2020; Hout & Goldinger, 2010).

Only one study has found superior recognition memory for similar distractors while controlling for presentation time (Williams, 2010a). In that study, participants were given a verbal search cue (e.g., “white car”) prior to an RSVP search trial. Distractors that were either semantically related to the target (e.g., “yellow car”) or shared the same color as the target (e.g., “white umbrella”) were better recognized in a later memory test compared to unrelated distractors (e.g., “yellow umbrella”). However, in that study, participants were given a verbal search cue for stimuli that were easily nameable. As such, it is possible that the phonological characteristics of the target could have influenced search performance (and subsequent memory) to a greater extent than usual (Walenchok et al., 2016).

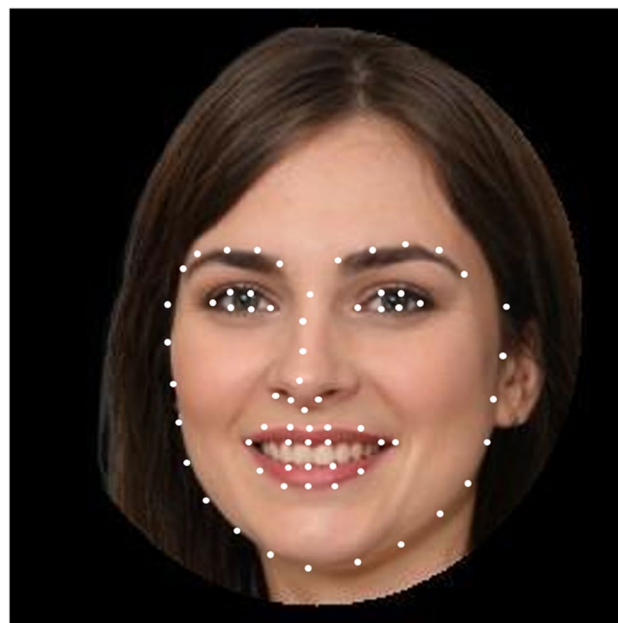


Fig. 1 An AI-generated face with their facial landmarks overlaid on their face in white dots

In the current study, we used unfamiliar AI-generated faces to populate our search arrays. Because faces are inherently less verbalizable (e.g., see Meissner et al., 2007), we are more confident in the assumption that participants were relying more on visual features rather than verbal labels in their search of a target. Additionally, because unfamiliar faces are primarily visual, it is unlikely that semantic knowledge could be used here in the same way as with well-known objects. As a result, using unfamiliar faces allows us to more easily manipulate the visual similarity of the target(s) to the distractors in a way that is separate from their semantic similarity. Similarity was manipulated based on an MDS model constructed from facial landmarks (see Fig. 1) and other metadata of each face. MDS is an analytical tool that models the similarity of a set of stimuli. Specifically, it does this by providing a spatial arrangement of stimuli in which the distance between items indexes how dissimilar they are from each other (see Hout et al., 2013, 2016).

The goal of the current study was to test how well the aforementioned hypotheses explain distractor memory in a visual search task with variable memory loads. In two experiments, participants were given an RSVP search task during which they were instructed to search for one (low VWM load) or two (high VWM load) potential target faces amongst distractor faces that were either more or less similar to the target(s). As previously mentioned, the similarity of the distractors to the target(s) was defined via a MDS space and manipulated within-subjects. Later, participants were given a surprise recognition memory test for the distractor (i.e., non-target) faces they encountered previously. We

Footnote 1 (continued)

they are more similar. Nevertheless, both versions of this explanation would require that participants pay more attention to each distractor in order to reject it.

recruited samples from two different participant populations (a traditional undergraduate student sample and a more heterogeneous online sample) to assess the generalizability of our outcomes.

As our primary outcome of interest is recognition performance for the distractor faces, we consider the memory predictions borne from each of the three previously-discussed accounts. According to the mental comparison hypothesis, the boost to distractor memory is caused simply by the difference in effort that is expended when mentally comparing each distractor to multiple targets, as opposed to only one target. According to the attention allocation hypothesis, the distractor memory benefit is due to the fact that participants approach the search with increased effort because they expect a more difficult task. According to the partial match hypothesis, a greater VWM load increases the featural overlap between targets and distractors, causing the distractors to be encoded to a greater extent and therefore be better remembered. That is, all three accounts predict better distractor memory from the two-target trials compared to the one-target trials. Importantly, the accounts diverge when trying to predict the effect of distractor similarity. The mental comparison account focuses solely on the *number* of comparisons made rather than the relative difficulty of those comparisons, and so suggests that distractor similarity is inconsequential. Similarly, the attention allocation hypothesis predicts no effect of distractor similarity because there is no explicit cue that the trials with similar distractors will be more challenging, so there is no reason to differentially allocate one's attention across conditions. More precisely, participants have no reason to expect that the current trial will have similar or dissimilar distractors. However, if greater featural overlap helps drive the benefits in distractor memory, then similar distractors should be more difficult to reject, be encoded to a greater extent, and therefore be correctly recognized more often than distractors in the dissimilar condition. That is, the partial match hypothesis uniquely predicts that we should find an effect of distractor-target similarity on distractor recognition.

Experiment 1

Method

Participants Using the effect size of the load effects from Hout and Goldinger (2010, Exps. 3a and 3b) as our starting point, we conducted a power analysis in R (R Core Team, 2021) that indicated that 180 participants were required in order to detect an interaction with an effect size that is roughly equal to $\eta_p^2 = 0.044$ with approximately 0.8 power, consistent with our pre-registration. Participants were

recruited from Skidmore College in exchange for course credit and Prolific in exchange for \$5.00. All of the participants were between 18 and 60 years old. A total of 187 people participated; 97 were recruited through Skidmore College, and 90 were recruited from Prolific. Data from seven participants were excluded because they reported that they were not paying attention. Data from an additional nine participants were excluded because their search performance was worse than 95% of the participants. All of these exclusions were pre-registered: <https://osf.io/q5gy7>. This led to a final sample of 171 participants (122 females, 42 males, four non-binary/other, three not reported) with an age range of 19–60 years ($M = 25.09$, $SD = 9.73$). 115 participants self-reported that they were Caucasian, eight were African American, 21 were Asian, one was American Indian or Alaska Native, 12 were multi-racial, and one was Native Hawaiian or Other Pacific Islander. Seven participants reported that none of the categories applied, and six participants did not provide their race.

Design We employed a 2 (Similarity: similar vs. dissimilar distractors) \times 2 (Load: 1 vs. 2 targets) \times 2 (Target presence: present vs. absent) within-subjects design.

Materials Presented images came from a database of 10,000 artificially generated faces from Generated. Photos (<https://generated.photos/>). Each face included metadata of 67 pairs of facial landmarks (see Fig. 1), as well as the AI predictions of whether the person was a male or a female, had makeup, showed a given emotional expression, was of a given ethnicity, had facial hair, had a given hair color, had a given hair length, had a given eye color, had a smile, and whether their forehead, mouth, or an eye were occluded by their hair; all features were coded in terms of probability. The metadata also included a prediction of the age of a given face. Because the variables were on different scales, each variable was z-standardized. We then set a threshold for sex and each ethnicity variable. Images in which the predicted sex was male with an unstandardized confidence that was greater than 0.64 were assigned to be a “male” face. Images in which the predicted sex was male with an unstandardized confidence that was less than 0.34 was assigned to be a “female” face. To avoid a racially homogenous set of faces, we then divided the images into “White” and “Latino” groups.² Images that were predicted to be Latino with an unstandardized confidence greater than 0.70 were assigned to be Latino. Images that were predicted to be White with an

² We note that these labels are somewhat arbitrary and that these images are of faces that vary on a continuum for both the Sex and Ethnicity variable. However, we use these labels to be consistent with the labels that were used in the original dataset.

Table 1 Actual cutoffs for each item type of males and females for Experiments 1 and 2 based on the ranked order of distance from a given seed face. The seed face was assigned a 1, and the farthest

image (i.e., least similar image) was assigned N, where N refers to the number of images within a given sex subgroup

	Experiment 1		Experiment 2	
	Males	Females	Males	Females
Targets	2 – 290	2 – 279	2 – 290	2 – 290
Similar distractors	291 – 578	280 – 557	340 – 490	340 – 490
Dissimilar distractors	867 – 1,155	835 – 1,112	1,398 – 1,548	1,651 – 1,800
New items	579 – 866	558 – 834	819 – 1,069	946 – 1,196

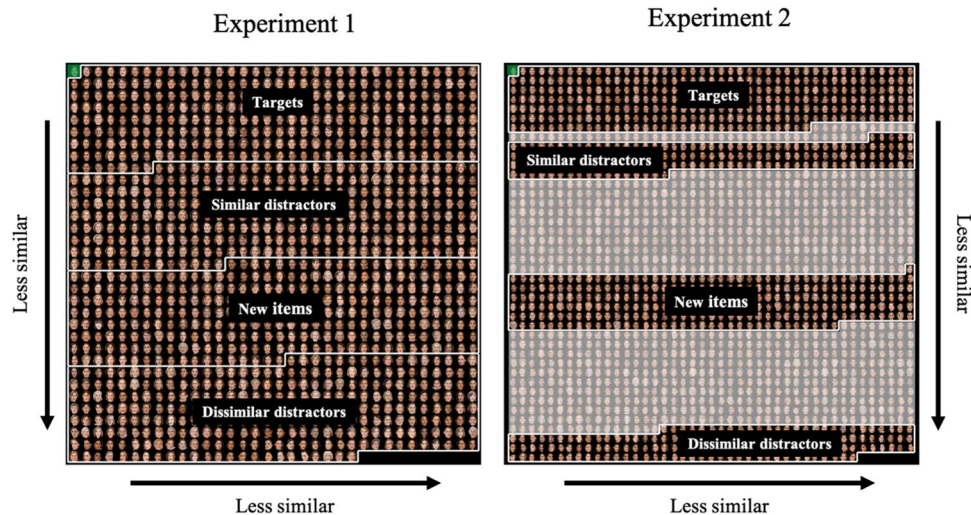


Fig. 2 Example of list construction for Experiment 1 (left panel) and Experiment 2 (right panel). The “seed face” (pictured in green) is in the first position. Images are ordered based on their distance from

the seed face, from left to right and top to bottom. Grayed-out images indicate sections that were eliminated from selection to increase dissimilarity of the faces across conditions (in Experiment 2)

unstandardized confidence of at least 0.69 were assigned to be White. Any images that fell inside of any of these thresholds were not included. In other words, images that were not classified as White or Latino and Male or Female with sufficient confidence were not included in this experiment. In applying these criteria, we were left with 483 Latino males, 343 Latina females, 672 White males, and 769 White females (2,267 in total). We then applied multidimensional scaling, using the metadata of each face, to create two ten-dimensional face spaces, one for males and one for females.

We computed the inter-item distances between all of the images within a given subgroup using their coordinates in the face space. Based on these inter-item distances, 100 lists were created for each sex. For each list, a random “seed image” was selected and the remaining images were ordered based on distance from that image. That is, each image was assigned a number from 1 to N corresponding to how close (i.e., similar) it was to the seed face. The seed face was given a value of 1, and the farthest (i.e., most dissimilar) image was given a value

of N. Based on this vector of dissimilarities, targets were sampled from the subset of items in the range of 2 to roughly $N/4$, similar distractors were sampled from the subset defined roughly by $(N/4)+1$ to $N/2$, the subset for the new items for the recognition test was defined roughly by $(N/2)+1$ to $(3N/4)$, and the subset for dissimilar distractors was defined roughly by $(3N/4)$ to N. The images of the targets, the similar distractors, the dissimilar distractors, and the new items on the recognition test were randomly selected from these subsets of images with the constraint that an equal number of images from each ethnicity would be sampled for each item type. That is, targets, similar distractors, dissimilar distractors, and new items each had an equal number of images from each ethnicity. The precise cutoffs for each subset are shown in Table 1 and Fig. 2. Figure 3 shows example dissimilar and similar search trials for each experiment. For each list, we selected a total of 120 targets, 12 distractors for each combination of VWM load and target-distractor similarity condition, and 48 images were selected to be new

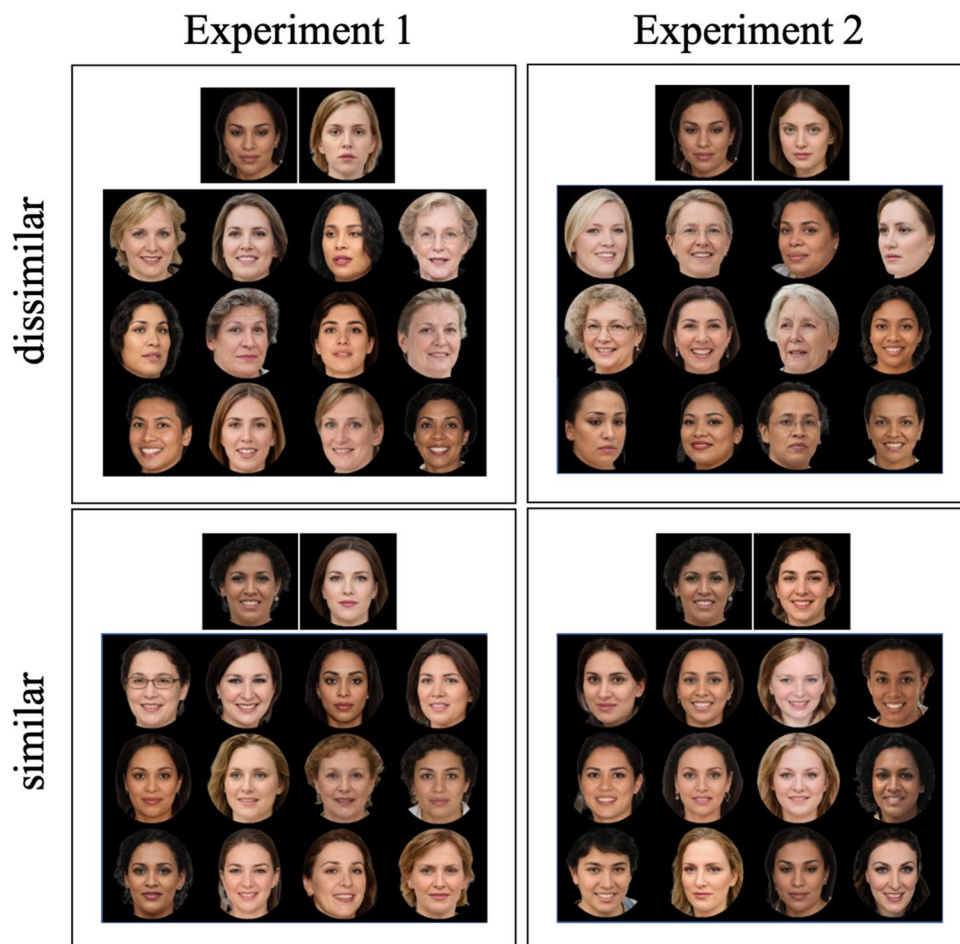


Fig. 3 Example trials of high visual working memory (VWM) load for a given target. Each panel depicts two targets at the top, with 12 distractors below. The top left panel depicts a dissimilar trial in Experiment 1. The top right panel depicts a dissimilar trial in Experiment 2.

The bottom left panel depicts a similar trial in Experiment 1. The bottom right panel depicts a similar trial in Experiment 2. Each of these trials are from different participants

items during the recognition test. The orders of both the search trials and the recognition trials were completely randomized.

Each search trial used a novel target. Distractors were repeated across search trials within each VWM load and similarity combination. There was a total of ten trials for each combination of VWM load and target-distractor similarity conditions. Each combination of conditions used a unique set of distractors. Within each combination of conditions, half of the targets were “White,” and the other half of the targets were “Latino.” Within each set of distractors, half of the faces were “White,” and the other half were “Latino.” For some participants, all of the faces were “Male,” and for other participants, all of the faces were “Female.”

Procedure Participants were told that their task was to search for one (low VWM load) or two (high VWM load)

target face(s) among a set of sequentially presented faces.³ Regardless of load, participants were presented with either zero (target absent trials) or one (target present trials) target during the ensuing array. After target presentation (self-paced), 12 additional face images (either 11 distractors and a target face, or 12 distractor faces) were rapidly presented at a rate of 650 ms per image. Participants then indicated via keystroke whether the target was present or absent in the previously presented array. To help familiarize them with the task, participants began with several practice trials during which feedback was given on the accuracy of their present/absent decisions. Following this, they were given 80 critical trials with no feedback.

³ Our decision to use two targets in the high VWM load condition was based on pilot work suggesting that more than two targets may be prohibitively difficult.

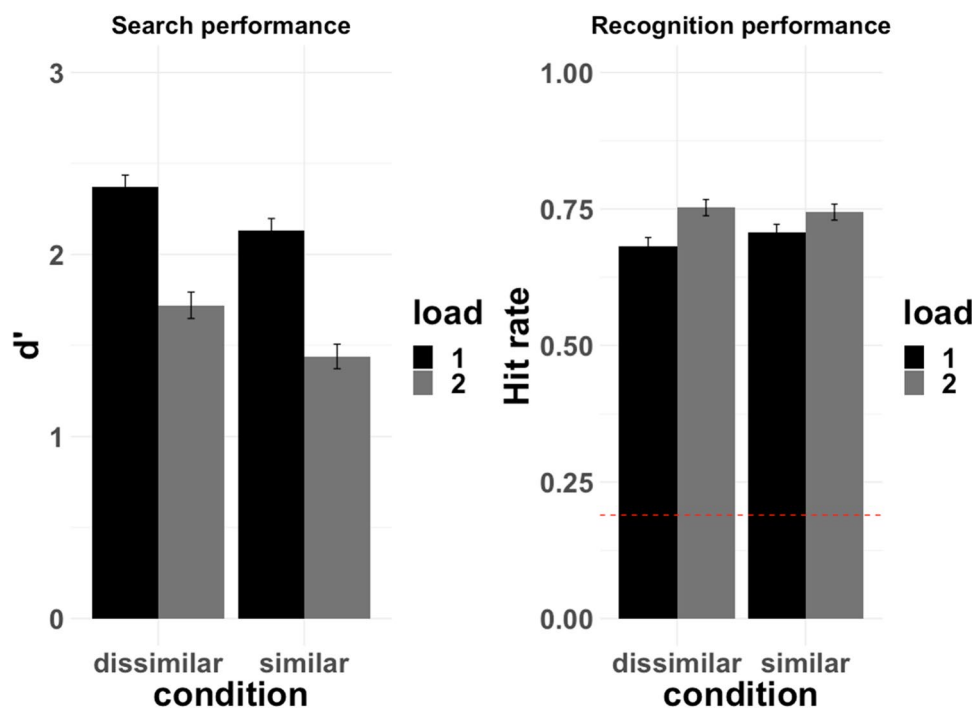


Fig. 4 Experiment 1: Search performance and recognition performance as a function of similarity and load. The red dashed line in the right panel indicates the false alarm rate. Error bars represent the standard errors of the means. None of these effects were moderated by Site

After all of the search trials were completed, participants were given a surprise recognition test for the distractor faces (i.e., not the search targets). For each trial, participants were asked whether they saw a given face in the previous phase. Participants were asked to respond on a 4-point scale, where 1 indicated that they had “Not Seen” the face, 2 indicated that they had “Unlikely Seen” the face, 3 indicated that they had “Likely Seen” the face, and 4 indicated that they had “Seen” the face. Half of the images were old (i.e., distractors from the search phase), and half of the images were new (i.e., faces which had not been presented during the search phase).

Results

An initial analysis suggested that none of the outcomes reported here were moderated by site ($ps > .17$), so the following analyses present the online and undergraduate samples collapsed together.

Search performance

To evaluate whether VWM load and similarity affected target detection performance, we ran a 2×2

repeated-measures ANOVA on d' . This is a measure of discriminability, which indexes the extent to which a participant discerns between trials in which a target is present in the search stream from trials in which no target is present in the search stream (Green & Swets, 1966; Macmillan & Creelman, 2005). Before computing d' , all hit rates and false alarm rates were corrected as recommended by Snodgrass and Corwin (1988). This analysis revealed a main effect of Load, $F(1, 170) = 172.70$, $MSE = 0.45$, $p < .001$, $\eta_p^2 = 0.50$, $BF_{10} = 8.56 \times 10^{33}$, indicating worse performance in the higher load condition ($M = 1.58$, $SD = 0.93$) compared to the lower load condition ($M = 2.25$, $SD = 0.86$). The main effect of Similarity was also significant, $F(1, 170) = 35.62$, $MSE = 0.33$, $p < .001$, $\eta_p^2 = 0.17$, $BF_{10} = 3301.55$, indicating worse target detection when target-distractor similarity was high ($M = 1.79$, $SD = 0.94$) compared to low ($M = 2.05$, $SD = 0.95$). The Similarity \times Load interaction was not significant, $F(1, 170) = 0.19$, $MSE = 0.37$, $p = .67$, $\eta_p^2 = 0.00$, $BF_{10} = 0.13$. A pre-registered planned contrast was conducted testing whether the similar and dissimilar conditions differed at a load of one. This contrast was significant, $t(170) = 3.82$, $p < .001$, $d = 0.28$, $BF_{10} = 82.66$. We also compared these two conditions at a load of two, which was also significant, $t(170) = 4.38$, $p < .001$, $d = 0.31$, $BF_{10} = 659.09$. These data are depicted in Fig. 4.

Recognition performance

To examine the effects of similarity and VWM load on subsequent distractor recognition performance, we ran a 2×2 repeated-measures ANOVA using the hit rates as our dependent measure.⁴ Ratings ≥ 3 were considered “yes” responses, and ratings ≤ 2 were considered “no” responses. Consistent with the search performance data, the main effect of Load was significant, $F(1, 170) = 26.28$, $MSE = 0.02$, $p < .001$, $\eta_p^2 = 0.13$, $BF_{10} = 16585.67$, indicating better recognition for distractors from the higher load condition ($M = 0.75$, $SD = 0.19$) than the lower load condition ($M = 0.69$, $SD = 0.20$). In contrast to the search performance data, the main effect of Similarity was not significant, $F(1, 170) = 0.58$, $MSE = 0.02$, $p = .45$, $\eta_p^2 = 0.00$, $BF_{10} = 0.11$, indicating that similar distractors ($M = 0.73$, $SD = 0.19$) and dissimilar distractors ($M = 0.72$, $SD = 0.20$) were recognized at comparable rates. The Similarity \times Load interaction was also not significant, $F(1, 170) = 2.64$, $MSE = 0.02$, $p = .11$, $\eta_p^2 = 0.02$, $BF_{10} = 0.37$. A planned contrast testing whether the similar and dissimilar condition differed at a load of one was not significant, $t(170) = -1.68$, $p = .09$, $d = -0.13$, $BF_{10} = 0.34$. These two conditions also did not differ from each other at a load two, $t(170) = 0.54$, $p = .59$, $d = 0.04$, $BF_{10} = 0.10$. These data are shown in Fig. 4.

Discussion

We found that searching for multiple targets both hurt search performance and enhanced distractor recognition, replicating earlier findings (cf. Hout & Goldinger, 2010, 2012) and extending them to face stimuli. In contrast to earlier findings (Williams, 2010a), however, similar distractors were not better recognized than dissimilar distractors. That is, the predictions borne out of the *partial match hypothesis* were not supported. Instead, these data suggest that the only relevant variables were some combination of the number of search targets (*mental comparison hypothesis*) and participants’ strategic choice to allocate attention during search (*attention allocation hypothesis*).

Before reaching such a conclusion, however, we considered an alternative explanation: perhaps the manner in which we operationalized similarity was simply not effective. Such a possibility is undermined by the fact that our design did yield similarity effects elsewhere; similarity hurt search performance in theoretically predictable ways. However, perhaps the difference in similarity between the two conditions in the current experiment was simply not large enough. It

should be noted the magnitude of the differences in high versus low similarity conditions here was likely smaller than in prior research. For example, in Williams (2010a), a target (e.g., white car) either shared the same category or color as the distractors (e.g., yellow car, white umbrella), whereas in the present experiment, all of the distractors were faces that could not be distinguished with simple verbal labels. Though that level of differentiation is likely impossible with face stimuli, in Experiment 2, we nevertheless attempted to increase the differences in similarity between the two similarity conditions again using the MDS space.

Experiment 2

In Experiment 1, despite finding an effect of target-distractor similarity on target detection, this did not appear to have any downstream effects on distractor recognition. The goal of Experiment 2 was to further examine the partial match hypothesis by widening the gap between the similar and dissimilar distractors; that is, to make the similar distractors more similar to the targets, and make the dissimilar targets less similar to the targets. As previously detailed, targets, distractors, and new items were sampled based on the ranked distance to a given seed face. In Experiment 2, based on a given seed face, the dissimilar distractors were sampled from a subset that was on average 1,034.5 images away from the subset that similar distractors were sampled from, compared to an average of 283.5 images in Experiment 1 (see Table 1 and Fig. 2 for details). To do this, we first needed to increase the number of images to sample from. This was accomplished by relaxing the thresholds for what counted as a male or a female, as well as what counted as Latino or White.

One additional change in Experiment 2 came in the form of participant populations. Because none of the findings in Experiment 1 varied as a function of site (Skidmore undergraduates vs. Prolific), we decided to exclusively recruit participants from Prolific.⁵

Method

Participants A total of 194 participants were recruited through Prolific. Data from ten participants were dropped because their search performance was worse than 95% of the sample (see our pre-registration for a list of possible exclusions: <https://osf.io/jfa6z>). This yielded a total of 184 participants (89 females, 87 males, five Non-binary/Other, three Not reported) with an age range of 19–60 years ($M =$

⁴ Unlike the search task, new items in the recognition test were not unique to any of the conditions. Therefore, it would not be possible to compute d' .

⁵ In our pre-registration, we wrote that we would recruit participants from Amazon Turk. We ultimately decided to recruit via Prolific but neglected to change this in the pre-registration.

36.26, $SD = 10.48$). 138 participants self-reported that they were Caucasian, 16 were African American, 12 were Asian, and seven were multi-racial. Five participants reported that none of the categories applied, and six participants did not provide their race.

Design The design of Experiment 2 was identical to Experiment 1 (though the differences between similar and dissimilar distractors was magnified as outlined below).

Materials Images in which the predicted sex was male with an unstandardized confidence that was at least 0.55 were assigned to be a “male” face. Images in which the predicted sex was male with an unstandardized confidence that was ≤ 0.46 were assigned to be a “female” face. Within each group, we then divided the images into “White” and “Latino” groups. Images that were predicted to be Latino with an unstandardized confidence 0.56 or greater were assigned to be Latino. Images that were predicted to be White with an unstandardized confidence of at least 0.63 were assigned to be White. For each list, a random “seed” image was selected and images were ordered based on their distance from the seed image. Images were then randomly sampled in the same manner as Experiment 1, from different subsets of the images. Based on the ranked distance of images to the seed face, targets were sampled from the subset defined as 2–290. The subset for similar distractors was defined as 340–490.

The subset for the dissimilar distractors was defined as (N-151) to N. The subset for the new items was defined such that it was of equal distance to the similar and dissimilar distractors (see Figs. 2 and 3 as well as Table 1 for details).

Procedure The procedure in Experiment 2 was identical to the procedure of Experiment 1.

Results

Search performance

As in Experiment 1, we ran a 2×2 repeated-measures ANOVA on d' scores of target identification after correcting the hit and false alarm rates. This analysis revealed a main effect of Load, $F(1, 183) = 275.20$, $MSE = 0.35$, $p < .001$, $\eta_p^2 = 0.60$, $BF_{10} = 4.39 \times 10^{38}$. Searching for two targets yielded poorer performance ($M = 1.61$, $SD = 0.90$) than searching for one target ($M = 2.33$, $SD = 0.90$). The main effect of Similarity was also significant, $F(1, 183) = 122.20$, $MSE = 0.40$, $p < .001$, $\eta_p^2 = 0.40$, $BF_{10} = 2.12 \times 10^{17}$. Searching for a target among similar distractors yields worse search performance ($M = 1.71$, $SD = 0.95$) than among dissimilar distractors ($M = 2.23$, $SD = 0.93$). The Similarity \times Load interaction was not significant, $F(1, 183) = 1.78$, $MSE = 0.30$, $p = .18$, $\eta_p^2 = 0.01$, $BF_{10} = 0.24$ (see Fig. 5). A planned contrast comparing the similar and

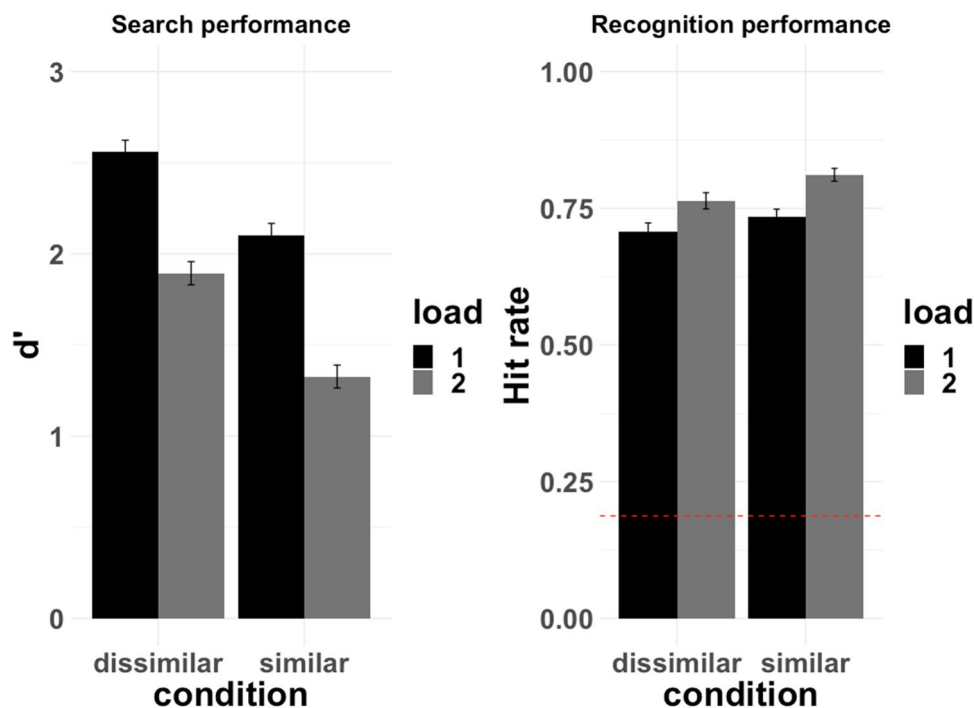


Fig. 5 Experiment 2: Search performance and recognition performance as a function of similarity and load. The red dashed line in the right panel indicates the false alarm rate. Error bars represent the standard errors of the means

dissimilar condition with a load of one was significant, $t(183) = 7.39, p < .001, d = 0.52, BF_{10} = 1.43 \times 10^9$. These two conditions were also significantly different at a load of two, $t(183) = 9.34, p < .001, d = 0.66, BF_{10} = 1.56 \times 10^{14}$.

Recognition performance

As done with Experiment 1, ratings ≥ 3 were considered “yes” responses, and ratings ≤ 2 were considered “no” responses. A 2×2 repeated-measures ANOVA on distractor recognition revealed a significant main effect of Load, $F(1, 183) = 51.48, MSE = 0.02, p < .001, \eta_p^2 = 0.22, BF_{10} = 4.96 \times 10^7$. Participants recognized distractors from multiple-target searches ($M = 0.79, SD = 0.18$) better than from single-target searches ($M = 0.72, SD = 0.20$). Importantly, there was also a significant main effect of Similarity, $F(1, 183) = 9.89, MSE = 0.03, p = .002, \eta_p^2 = 0.05, BF_{10} = 42.17$, indicating better memory for similar distractors ($M = 0.77, SD = 0.18$) compared to dissimilar distractors ($M = 0.74, SD = 0.21$). The Similarity \times Load interaction was not significant, $F(1, 183) = 1.38, MSE = 0.01, p = .24, \eta_p^2 = 0.01, BF_{10} = 0.18$ (see Fig. 5). A planned contrast comparing the similar and dissimilar condition at a load of one was not significant, $t(183) = -1.82, p = .07, d = -0.14, BF_{10} = 0.41$. However, these two conditions were significantly different at a load of two, $t(183) = -3.29, p = .001, d = -0.26, BF_{10} = 14.62$.

Discussion

In this experiment, we replicated the VWM load effects on both search performance and distractor recognition from Experiment 1. We also replicated the effect of target-distractor similarity on search performance. In contrast to Experiment 1, we observed a reliable effect of target-distractor similarity on recognition performance. These data suggest that the difference in similarity in Experiment 1, though sufficient to impact search performance, was not large enough to affect memory for the distractors. Consistent with the partial match hypothesis, this suggests that during search, similar distractors received more attention than dissimilar distractors, presumably because they were more difficult to reject. In turn, this additional distractor encoding led to better memory.

General discussion

In the current study, we manipulated both VWM load and target-distractor similarity in an RSVP task, using unfamiliar face stimuli to test three different accounts of why distractor memory is improved in more challenging searches. In Experiment 1, we found an effect of VWM load on both target detection in the initial search as well as

subsequent distractor recognition; this is the first time such effects have been demonstrated with face stimuli, which is an important extension of previous findings. We did not, however, find an effect of target-distractor similarity on distractor recognition. Critically though, similarity did hurt target detection in the search task. In Experiment 2, we increased the difference in similarity between the high versus the low similarity conditions, and found an effect of target-distractor similarity on search performance and distractor memory.

Taken together, the present experiments help us make sense of why distractor memory is enhanced in RSVP tasks when people are tasked with looking for multiple targets. In both experiments, we found that VWM load hurt search performance and helped distractor memory. According to the mental comparison hypothesis, this is due to the fact that in a multiple-target search, each target needs to be mentally compared to each distractor, and in doing so, this enhances memory for each distractor. The attention allocation hypothesis states that a multiple-target search cue serves to set an expectation of difficulty in participants, who subsequently allocate more attention to distractors on these trials, yielding better distractor memory. The partial match hypothesis asserts that on multiple-target search trials, there is more featural overlap between the targets and the distractors. As a result, a distractor is encoded to a greater extent in order for it to be rejected.

Importantly, the partial match hypothesis uniquely predicts that target-distractor similarity will impact distractor memory. In Experiment 1, although we found an effect of similarity on search performance, this effect was not observed in the distractor recognition data. Thus, we did not find any evidence for the partial match hypothesis. However, we speculated that this might not be an indictment of the partial match hypothesis per se, but rather a consequence of our similarity manipulation being too subtle. In Experiment 2, we magnified the difference between the low and high similarity conditions and demonstrated the predicted similarity effects on distractor memory.

We should emphasize that most of the relevant extant research did not use face stimuli, and that none of the prior research examined distractor memory using faces. For example, Williams (2010a) showed better memory for distractors that share the same category or color as the target. However, in that study, participants could potentially rely on verbal labels in searching for the target (e.g., yellow car) among distractors (e.g., white umbrella). Indeed, there is evidence that search performance is influenced by the phonological characteristics of the verbal label of the target (Walenchok et al., 2016). By using face stimuli as we do in the current study, participants could not easily rely on verbal labels during their search. Even with this adjustment, we still found the predicted effect of similarity, suggesting a robust phenomenon.

Such an outcome is not necessarily a given because unlike the common object stimulus sets used in prior research (e.g., cars, umbrellas, etc.), faces do not lend themselves as easily to verbal labels, rendering the similarity (or lack thereof) between two faces difficult to articulate. Using face stimuli had the added benefit of allowing us to manipulate the *visual* similarity of the distractors to the targets in a more continuous fashion rather than manipulating discrete features of the items (e.g., color or category). That is, because these faces were unfamiliar, it is unlikely that participants could have used category membership in a way that may render their visual similarity irrelevant (or less impactful).

The current research shows that findings from previous research generalize to facial stimuli. This too was not a foregone conclusion, as performance on other tasks utilizing objects does not always generalize to unfamiliar faces (see Mestry et al., 2017, for a discussion). This could be because of resource limitations arising from the complexity of unfamiliar faces (Eng et al., 2005; Mestry et al., 2017), the uniqueness of faces, or because faces are processed differently than other stimuli (e.g., Tanaka & Farah, 1993). Despite all of these potential differences, we nevertheless replicated a number of key previous findings.

Importantly, the current study provided evidence for the partial match hypothesis. Future directions could include testing the mental comparisons hypothesis by manipulating the number of comparisons. However, we should note that this is difficult because it is not clear how to manipulate this variable without confounding a number of other variables. For example, by increasing the VWM load, and therefore the number of comparisons required, one is likely increasing the featural overlap between targets and distractors. The attention allocation hypothesis would likely be easier to test in future work. If participants are allocating more of their attention to distractors on trials that they expect to be difficult, then simply indicating the difficulty of an upcoming trial should enhance memory for those distractors.

When searching for multiple faces among a set of distractors, we must process each face to some extent in order to determine whether or not it is the face that we are looking for. Although searching for multiple faces (as opposed to one) will likely decrease our ability to find any of the targets, we will likely show better memory for the faces that we had rejected. The current study suggests that this is due to the fact that searching for multiple faces will often increase the amount of featural overlap between the target faces and the non-target faces. Returning to the “missing children” example discussed above, a teacher who is looking for a missing child among similar looking children will likely develop better memory for the non-missing children’s faces, compared to a situation in which the non-missing children are less similar. In addition, when searching for multiple missing children, it is likely that more of the faces of the non-missing

children will look like one or more of the missing children. In this situation, a teacher may encode the faces of the non-missing children to a greater extent, and develop a stronger memory of seeing their faces. If another child goes missing, the stronger memory may make it easier to identify and reject each non-missing child because the information about each child is processed faster. Alternatively, the stronger memory could result in a less conservative threshold for how much information the teacher needs in order to reject each non-missing child (Hout & Goldinger, 2012). In both cases, if another child goes missing, the stronger memory for the non-missing children would likely facilitate search performance.

Author Note Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Acknowledgements This work was supported by the James S. McDonnell Foundation [Grant Number 220020429].

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.
- Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision*, *9*(1), 1–9.
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*(3), 433–458.
- Eng, H. Y., Chen, D., & Jiang, Y. (2005). Visual working memory for simple and complex stimuli. *Psychonomic Bulletin & Review*, *12*(6), 1127–1133.
- Godwin, H. J., Menneer, T., Cave, K. R., & Donnelly, N. (2010a). Dual-target search for high and low prevalence X-ray threat targets. *Visual Cognition*, *18*(10), 1439–1463.
- Godwin, H. J., Menneer, T., Cave, K. R., Helman, S., Way, R. L., & Donnelly, N. (2010b). The impact of relative prevalence on dual-target search for threat items from X-ray screening. *Acta Psychologica*, *134*(9), 79–84.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.
- Guevara Pinto, J. D., & Papesh, 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.
- Guevara Pinto, J. D., Papesh, M. H., & Hout, M. C. (2020). The detail is in the difficulty: Challenging search facilitates incidental object encoding. *Memory & Cognition*, *48*(7), 1214–1233.
- 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 learnings speeds visual search by lower 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.

- Hout, M. C., Papesh, M. H., & Goldinger, S. D. (2013). Multidimensional scaling. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(1), 93–103.
- Hout, M. C., Godwin, H. J., Fitzsimmons, G., Robbins, A., Menneer, T., & Goldinger, S. D. (2016). Using multidimensional scaling to quantify similarity in visual search and beyond. *Attention, Perception, & Psychophysics*, 78, 3–20. <https://doi.org/10.3758/s13414-015-1010-6>
- Lavelle, M., Alonso, D., Luria, R., & Drew, T. (2021). Visual working memory load plays limited, to no role in encoding distractor objects during visual search. *Visual Cognition*, 29(5), 288–309.
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: a user's guide* (2nd ed.). Erlbaum.
- Meissner, C. A., Sporer, S. L., & Schooler, J. W. (2007). Person descriptions as eyewitness evidence. In R. Lindsay, D. Ross, J. Read, & M. Tolia (Eds.), *Handbook of Eyewitness Psychology: Memory for People* (pp. 3–34). Lawrence Erlbaum & Associates.
- Menneer, T., Stroud, M. J., Cave, K. R., Li, X., Godwin, H. J., Liversedge, S. P., & Donnelly, N. (2012). Search for two categories of target produces fewer fixations to target-color items. *Journal of Experimental Psychology: Applied*, 18(4), 404–418.
- Mestry, N., Menneer, T., Cave, K. R., Godwin, H. J., & Donnelly, N. (2017). Dual-target cost in visual search for unfamiliar faces. *Journal of Experimental Psychology: Human Perception & Performance*, 43(8), 1504–1519.
- R Core Team. (2021). *R: A language and environment for statistical computing. (version: 4.1.2)*. R Foundation for Statistical Computing <https://R-project.org/>
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117(1), 34–50.
- Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology*, 46(2), 225–245.
- Thomas, M. D., & Williams, C. C. (2014). The target effect: Visual memory for unnamed search targets. *The Quarterly Journal of Experimental Psychology*, 67(11), 2090–2104.
- Walenchok, S. C., Hout, M. C., & Goldinger, S. D. (2016). Implicit object naming in visual search: Evidence from phonological competition. *Attention, Perception, & Psychophysics*, 78(8), 2633–2654.
- Williams, C. C. (2010a). Incidental and intentional memory: What memories are and are not affected by encoding task? *Visual Cognition*, 18, 1348–1367.
- Williams, C. C. (2010b). Not all visual memories are created equal. *Visual Cognition*, 18, 1348–1367.
- Williams, C. C., & Henderson, J. M. (2005). Incidental visual memory targets and distractors in visual search. *Perception & Psychophysics*, 67(5), 816–827.

Open practices statement All of the data are available at <https://osf.io/fx5r6/>.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.