

Lost in the supermarket: Quantifying the cost of partitioning memory sets in hybrid search

Sage E. P. Boettcher^{1,2} · Trafton Drew³ · Jeremy M. Wolfe^{4,5}

© Psychonomic Society, Inc. 2017

Abstract The items on a memorized grocery list are not relevant in every aisle; for example, it is useless to search for the cabbage in the cereal aisle. It might be beneficial if one could mentally partition the list so only the *relevant* subset was active, so that vegetables would be activated in the produce section. In four experiments, we explored observers' abilities to partition memory searches. For example, if observers held 16 items in memory, but only eight of the items were relevant, would response times resemble a search through eight or 16 items? In Experiments 1a and 1b, observers were not faster for the partition set; however, they suffered relatively small deficits when “lures” (items from the irrelevant subset) were presented, indicating that they were aware of the partition. In Experiment 2 the partitions were based on semantic distinctions, and again, observers were unable to restrict search to the relevant items. In Experiments 3a and 3b, observers attempted to remove items from the list one trial at a time but did not speed up over the course of a block, indicating that they also could not limit their memory searches. Finally, Experiments

4a, 4b, 4c, and 4d showed that observers were able to limit their memory searches when a subset was relevant for a run of trials. Overall, observers appear to be unable or unwilling to partition memory sets from trial to trial, yet they are capable of restricting search to a memory subset that remains relevant for several trials. This pattern is consistent with a cost to switching between currently relevant memory items.

Keywords Visual search · Memory

Real-world search tasks, like searching for items in a grocery store, are often defined by both a search through memory and a search through space. This combination of memory and visual search is known as a “hybrid search” (Schneider & Shiffrin, 1977). Wolfe (2012) showed that response times (RTs) in hybrid search increase linearly with the number of items in the visual display, and logarithmically with the number of items in memory. In other words, adding an additional item to a visual display will cost a fixed amount of time, irrespective of the number of items already in the visual display. However, in search through memory, adding the second item to a memory set will cost more time than adding the 100th item to that same memory set.

Cunningham and Wolfe (2014) proposed a three-stage model to explain hybrid search (Fig. 1). During the first stage, a particular item or subset of items in a display is selected via guided visual search (Wolfe, 1994, 2007). The visual selection is said to be “guided” because observers will preferentially select items with visual properties that could be those of items in the memory set. If observers are looking for strawberries, raspberries, and cranberries, they will tend to restrict visual search to items that are small and red. Within the set of plausible items, visual selection is proposed to be serial: Only one item may be selected at a time. This is reflected in the linear

Electronic supplementary material The online version of this article (doi:10.3758/s13421-017-0744-x) contains supplementary material, which is available to authorized users.

✉ Sage E. P. Boettcher
sage.boettcher@psy.ox.ac.uk

¹ Department of Experimental Psychology, Brain & Cognition Lab, University of Oxford, Oxford, UK

² Oxford Center for Human Brain Activity, University of Oxford, Oxford, UK

³ University of Utah, Provo, UT, USA

⁴ Brigham & Women's Hospital, Boston, MA, USA

⁵ Harvard Medical School, Boston, MA, USA

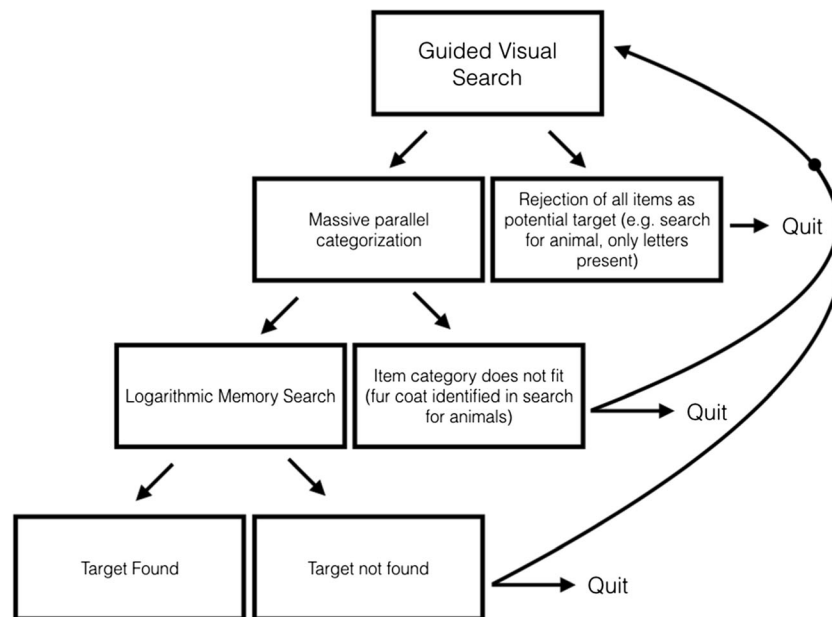


Fig. 1 Proposed model of hybrid search

increase in RTs with visual set size, though various parallel models also produce these linear functions (Townsend, 1990). Items that are selected by the first stage are then identified in the second step. This can be thought of as a massively parallel memory search, as the visual object is compared with the contents of long-term memory. If the identified object could be a possible member of the memory set of targets, then the third stage—a logarithmic search through the current memory set—is required. If not, then the third step is not needed, and the process can return to Step 1 for the selection of a new visual item.

Suppose that the memory set consists of a set of several animals: cat, chicken, moose, and badger. If the visual display includes letters, the guided search process will be able to avoid selecting those in Step 1 because letters lack the basic features of animals. If the image of a fur coat were selected in Step 1, Step 2 would identify it as a fur coat. This would be sufficient to establish that this item could not be in a memory set consisting of animals and, again, the process would return to Step 1 for another selection. If, however, the selected item were identified as a cow, a logarithmic search through the memory set would be needed to determine whether “cow” was one of the target items. This series of steps repeats until a target is found or until the search is abandoned. In this article, we consider the possibility that the memory set could be divided into subsections. For example, imagine that the observer was holding in memory the aforementioned list of animals, in addition to a list of clothing items. When encountering a fur coat, would participants be required to search through the union of both lists, or would they be able to limit the memory search to the clothing items? The purpose of this article is to address that gap in our knowledge.

The logarithmic memory search function can be modeled as a by-product of a drift diffusion process (Ratcliff, 1978) in which a separate diffuser is launched for each member of the memory set. Thus, for a memory set of N items, N diffusers accumulate evidence toward their respective decision boundaries. The more diffusers in action, the greater the chance that one will reach its bound erroneously, generating a false alarm or false positive error. To hold error rates roughly constant, the decision boundary would need to be made higher/more conservative as memory set size increases. This, in turn, increases RTs in a logarithmic fashion (Drew & Wolfe, 2013; Leite & Ratcliff, 2010; Ratcliff & Starns, 2013). This logarithmic function is characteristic of the average RT data in hybrid-search experiments. The order of items in lists and other such factors do have effects (Nosofsky, Cox, Cao, & Shiffrin, 2014), but in the present experiments the results were averaged over these factors. In addition, other routes might lead to logarithmic RT functions. For instance, if the memory set can be searched in a series of steps that eliminate a fraction of the items on each step, then the number of steps required will be a log function of the memory set size. This is akin to the children’s game of guessing a number: Is it bigger than 50?, Is it bigger than 25?, and so forth. Thus, if the items in memory can be coded in a manner that permits this sort of memory search, a log memory search function would be predicted.

Whatever the underlying mechanism, the logarithmic function is a robust finding. Previous work has shown that this logarithmic increase in RTs as a function of memory set size remains consistent over a range of conditions. It is seen in hybrid search for memory set sizes of up to 100 items (Drew & Wolfe, 2013; Wolfe, 2012). In a variant in which observers (Os) responded to the presence of a novel item, rather than to a

specific item held in memory, a logarithmic pattern persisted as memory set sizes increased to 500 (albeit with speed–accuracy trade-offs at the highest memory loads; Wolfe, Boettcher, Josephs, Cunningham, & Drew, 2015). If, instead of search for specific images, observers are searching for any member of several different categories (e.g., find any animals, plants, flags, or cars), RTs again increase linearly with the log of the number of categories in memory (Cunningham & Wolfe, 2014), as they do if the targets are words rather than pictures (Boettcher & Wolfe, 2015). Finally, when observers search for targets through time in an RSVP task, rather than through space, the threshold RSVP rate increases with the log of the memory set size (Drew & Wolfe, 2013).

Whatever its cause, logarithmic search through memory may serve a useful function. Wolfe (2012) offered the example of searching for any of your 1,000 friends in a picture of 100 people. If search through memory was, like search through the visual display, a linear function of set size with a slope of about 50 ms/item, then that photo search task would take about 45 min. With a logarithmic compression of the memory search, it takes seconds. Returning to the grocery store example, logarithmic compression of the shopping list in memory is just one factor that could contribute to making search for dinner into a reasonably efficient shopping experience. Our knowledge of the real world is also useful. From the visual search literature we know that search can be strongly guided by scene context (Neider & Zelinsky, 2006; Vö & Henderson, 2010; Wolfe, Vö, Evans, & Greene, 2011). In the grocery store, such knowledge would keep us from searching for apples on the ceiling or floor of the produce section. Even in arbitrary displays, the benefit of context is well-established in the visual search literature. Contextual-cueing experiments have shown that RTs are speeded when search displays repeat, even when observers have no explicit memory of the repetitions (e.g., Chun & Jiang, 1998). Additionally, in explicitly divided displays (e.g., drawn quadrants), partitions can facilitate difficult search and improve memory for target locations (Nakashima & Yokosawa, 2013; Solman & Kingstone, 2017).

Context, in various forms, also benefits memory recall. For instance, suppose you misplace your wallet in your home. One of your first reactions may be to “retrace your steps” when you return to the context in which you left the item (e.g., the bedroom) you may remember the act of leaving the item. Empirically, the benefits of context on recall have been established over the past several decades (Eich, 1985; Godden & Baddeley, 1975; Smith, 1979). Godden and Baddeley (1975) famously showed that when participants learned a list of words on either land or in water and were subsequently tested in the same context, they performed better than when they were tested in a different context. Hudson and Austin (1970) showed that when participants were cued with an item from the relevant category of memorized lists of categorized words, recall performance was enhanced as compared to the

uncued condition. These effects have been replicated repeatedly, indicating that during recall we are to some extent capable of using contextual links formed during the encoding phase. Despite these robust effects, the part-list cueing paradigm has long shown that performance in a recall task does not always improve when participants should only recall a subset of a memorized list. For example, after participants have been instructed to memorize a list of words, further instructing the participants to not recall parts of these lists (theoretically lowering the effective memory set size) does not benefit recall (Bäuml & Aslan, 2006; Roediger, 1973; Slamecka, 1968). This holds true even when participants memorize a categorized list of words and the instructions are based in semantic information (Roediger & Tulving, 1979).

Moreover, the effect of context on recognition memory is also unclear. Much of the early work showed no reliable effect of context on recognition memory (Godden & Baddeley, 1975, 1980; Smith, Glenberg, & Bjork, 1978). However, context-dependent recognition effects were subsequently demonstrated by a number of different groups (Dalton, 1993; Grant et al., 1998; Krafka & Penrod, 1985; Smith & Vela, 1992; for a review and meta-analysis, see Smith & Vela 2001). According to Macken (2002), these anomalies can be accounted for by differentiating between recollection and familiarity. Specifically, he showed that context yields an effect on recognition performance only when it is accompanied by conscious recollection rather than recognition based purely on familiarity.

However, it still remains unclear whether we should be able to boost the efficiency of a hybrid search by strategically limiting the memory search component. Consider a situation in which you have a grocery list that includes deli items and vegetable items. As you browse the produce aisle, are you searching through your entire mental shopping list or can the memory search be restricted to the vegetables on the list? Partitioning the entire memory set into produce and deli subsets would seem to allow more efficient search in each section of the store, but are we capable of limiting our memory search to the relevant items of a display?

In the following series of experiments, we investigated the ability to divide memory sets into subsets when performing hybrid search tasks. To preview our results, it is clear that observers can maintain separate memory subsets, but there is typically a significant cost of maintaining more than one set simultaneously. If the context specifies one subset, observers can search for the currently relevant items, with “lure” items from the currently irrelevant subset producing very few errors. However, although lures from a different target set did not lead to many errors, it appears that significant costs are associated with holding more than one target set in activated long-term memory (ALTM; see Cowan, 1995) at once. This cost can be eliminated when the observer completes many trials in a row with the same target set. That is, participants are

unlikely to inappropriately select an apple at the deli counter, but it seems that the presence of produce items on the grocery list in memory will slow search for salami at the deli counter.

Experiments 1a and 1b: Partitioning hybrid search within a block

Experiment 1 was intended to emulate the deli versus vegetable subset example described earlier, with some members of the memory set being relevant in one context and others in another context.

Method

Participants The 23 observers in Experiment 1a (mean age 33.4 years, 13 males, ten females) and 15 observers Experiment 1b (mean age 26.5 years, seven males, eight females) were recruited from the Harvard Decision Science Laboratory study pool as well as the general population in Cambridge, Massachusetts. In Experiment 1a, three observers did not complete the full experiment and were discarded from final analysis. In addition, two observers had extraordinarily high error rates in at least one of the conditions (above two standard deviations from the mean) and were therefore not included in the analysis. We aimed for 15 participants in each experiment, consistent with the numbers of observers needed to establish logarithmic $RT \times \text{Memory Set Size}$ functions in prior work. However, because we ran up to 12 participants at a time in Experiment 1a, we slightly overshot this goal. All participants gave informed consent and were compensated \$10/h for their participation. The Partners Healthcare Corporation Institutional Review Board approved all experimental procedures.

Stimuli and apparatus All experiments were written in MATLAB 7.9 (The MathWorks, Natick, MA) using Psychophysics Toolbox (Brainard, 1997; Pelli, 1997), version 3. Experiment 1a was run at two locations simultaneously. Sixteen of the observers were run at the Harvard Decision Science Laboratory on Dell Optiplex 9020 computers running Windows 7, 64-bit edition. Stimuli were presented on 19-in. LCD monitors set to a resolution of $1,440 \times 900$, with a refresh rate of 60 Hz. Seven observers from Experiment 1a and all of the observers in Experiment 1b were tested on 24-in. iMac computers running OS X 10.8.5 with the resolution and refresh rate set to match the conditions at the Harvard Decision Science Laboratory. The stimulus items were drawn from 2,300 photographs of objects used in the picture memory experiments of Brady, Konkle, Alvarez, & Oliva (2008), and they subtended approximately 3.2° . Six background scenes were additionally chosen from Google Image to serve as context for the memory subsets.

Analysis For all analyses, we used the ez package in R (Lawrence, 2013) for the analysis and StatCheck (Epskamp & Nuijten, 2016) to ensure that we reported the correct p -values. Figures were generated using ggplot2 (Wickham, 2009).

Procedure Experiments 1a and 1b differed slightly in their procedures. In both experiments there were three blocks and two main phases within each block: a memorization phase and a search phase. In each experiment there were three blocks, one for each condition (Fig. 2). The order of blocks was randomized between participants. In the “8Alone” condition, observers memorized a single group of eight objects during the memorization phase. In the “16Alone” condition, observers memorized a single group of 16 objects. In the critical “8Partition” condition, observers memorized two groups of eight objects each, for a total memory set of 16. The targets for each condition were chosen randomly from the stimulus set and could not repeat between conditions. Each item was associated throughout the experiment with a background scene chosen randomly from a set of six possibilities (beach, city, classroom, desert, forest, or mountain). This is how the “groups” of memory items were determined. That is, in the 8Alone and 16Alone conditions, all items were associated with the same background scene. However, during the 8Partition condition, eight of the items were associated with one background, and eight of the items were associated with another. Therefore, in total, each observer saw four unique background scenes. During the memorization phase, each item from the memory set was presented in a random order in isolation for 3 s at the center of the display on its assigned background. Note that the items in the memory sets of Experiments 1a and 1b were not semantically related to each other or the context in which they were presented. In these experiments, the subsets were arbitrary collections of objects rather than being categorically defined, as they might be in the produce aisle context in the grocery store.

After viewing each target, observers were required to pass two recognition tests with scores above 90% correct. If an observer failed to reach that accuracy, the memory set items were presented again and another recognition test was run. During the recognition test, observers saw a single object and were instructed to indicate whether it was “old” (i.e., part of their memory set) or “new” (a distractor). Observers saw a total of $2X$ objects, where X was the current memory set size. Of these objects, 50% were targets. Once observers had passed the recognition test, they were allowed to move on to the next portion of the experiment—the search phase. Note that the memorized target sets remained constant for the subsequent block of several hundred trials. Any order effects from the

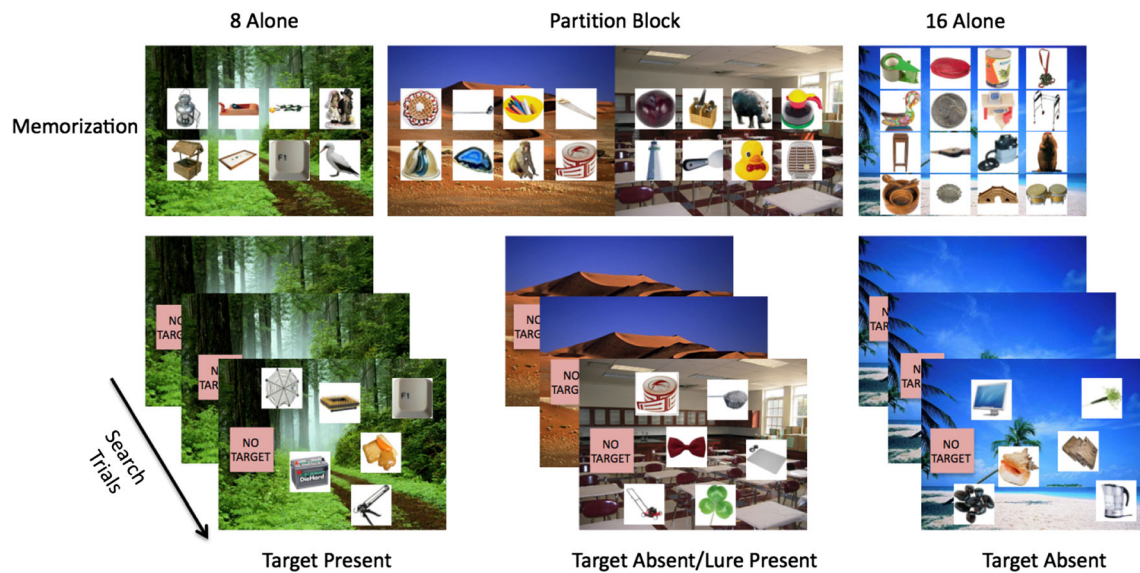


Fig. 2 Schematic design for Experiment 1. Each observer completed three distinct blocks of trials (8Alone, 8Partition, and 16Alone). During the search trials, the background image indicated the relevant target set to the observer. In the partition block, targets from the irrelevant set could appear as lures

initial memorization phase would not be likely to be a major factor driving responses over the course of the long block of trials (Nosofsky et al., 2014).

In the recognition memory test for the 8Partition condition, observers were required to show not only that they had learned the identity of each object, but that they had learned the association of each subset with its background. For this test, objects were presented in isolation in the center of a gray screen. Observers used the “1” key on the keyboard to indicated if they thought the object was from the first group, the “2” key if they believed the object was a part of the second group, and the “3” key if they believed the object was not a target. Again, performance of at least 90% correct was required in order to move to the next stage of the experiment. This test was repeated after the search trials, as well.

The critical difference between Experiments 1a and 1b was revealed in the search phase. In both experiments observers searched through visual displays of either six or 12 items. Targets were present on 50% of the trials. Observers were instructed to click on their target as quickly as possible, or, when they believed no target was present, they were instructed to click on a “no target” box positioned on the left side of the screen. As is shown in Fig. 2, all of the search items were overlaid on a background image that specified which set of targets was relevant for the current trial. For the 8Partition block, this could be either of the two scenes presented during the memorization portion. For the 8Alone and 16Alone blocks, only one background appeared during the search trials. Critically, “lures” from the irrelevant set were included in the 8Partition condition of Experiment 1a but excluded in Experiment 1b. A lure was an object, defined as a target in one subset, but presented on the background of the other subset (e.g., the apple on the deli counter). Clicking on a lure

constituted a false alarm error. That is, a target was correct only if it was presented on the relevant background, and not otherwise. Lures appeared on 50% of the search trials during the 8Partition block. Lure presence/absence was independent of target presence/absence.

Schneider and Shiffrin (1977) described their memory search tasks as involving either “consistent mapping” or “variable mapping.” In their original formulation, the memory set on consistent-mapping trials changed from trial to trial, but targets were always drawn from the same set of items, and a target on one trial did not appear as a distractor on another. On variable-mapping trials, members of the target set on one trial could appear as distractors on other trials. In Experiments 1a and 1b, the 8Alone and 16Alone conditions could be described as consistent-mapping conditions, since the memory set was the same on each trial and targets never appeared as distractors. In Experiment 1a, the 8Partition condition could be thought of as a version of a variable-mapping condition, since a target from one group could appear as a distractor/lure on trials in which the other group was relevant. In Experiment 1b, the 8Partition can be considered to be another form of consistent-mapping condition, because the targets were always targets. We will return to this topic in the discussion of Experiment 1b.

Results

Memorization phase The observers in Experiments 1a and 1b made very few errors (misses or false alarms) in the memorization test (Exp. 1a, 3.7%; Exp. 1b, 1.6%). In 8Partition blocks, observers performed a recognition test to determine whether they had learned the partition of items into two sets. Observers produced 6.1% errors in Experiment 1a and 7%

errors in Experiment 1b, indicating good memory for the partitions. After the search trials they performed this test again, producing only about a 1% error rate in each version of the experiment.

Experiment 1a: Comparing 8Partition, 8Alone, and 16Alone The primary goal of Experiment 1 was to determine whether search during the partition block would more closely resemble a search through eight objects alone or 16 objects alone. The main finding was that both partition conditions resembled the 16Alone condition. There was no evidence in these data for an ability to restrict memory search to a subset, though there was evidence that observers knew whether an item came from the relevant or the irrelevant subset for a trial's context.

Figure 3 shows a comparison of the three conditions in Experiment 1a. To fairly compare the three conditions, the trials from the 8Partition condition that contained lures are not included in Fig. 3 or in the associated analysis. This removed 50% of the trials from the analysis of the 8Partition block. The impact of those lures will be discussed later. Trials producing miss or false alarm errors were also excluded from this analysis.

A repeated measures analysis of variance (ANOVA) with Condition (16Alone, 8Alone, and 8Partition) and Set Size (6 and 12) as factors revealed significant main effects on RTs of condition [$F(2, 34) = 9.24, p < .001, \eta_G^2 = .12$] and set size [$F(1, 17) = 365.4, p < .001, \eta_G^2 = .53$]. Additionally, we found a significant interaction of these two factors [$F(2, 34) = 4.2, p = .02, \eta_G^2 = .004$]. When comparing the three conditions directly, the 8Alone condition, unsurprisingly, was significantly different from 16Alone [$t(17) = 2.68, p = .016$]. Additionally, 8Alone differed reliably from the 8Partition block [$t(17) = 4.25, p < .001$]. In contrast, 16Alone and 8Partition did not differ significantly [$t(17) = 1.24, p = .231$].

If the trials with lures were included in the analysis, the results were similar: There were main effects of condition [$F(2, 34) = 13.68, p < .001, \eta_G^2 = .16$] and visual set size [$F(1, 17) = 341.4, p < .001, \eta_G^2 = .52$], as well as an interaction [$F(2, 34) = 5.9, p = .006, \eta_G^2 = .007$]. The 8Partition condition still differed significantly from the 8Alone condition [$t(17) = 5.19, p < .001$]. In contrast, RTs for the 8Partition condition were now significantly longer than those in the 16Alone condition [$t(17) = 2.5, p = .02$].

Experiment 1a: Lures In Experiment 1a, in the 8Partition condition, objects could be presented from the target set that was not currently relevant. If observers could not keep the two memory sets separate, they should commit large numbers of false alarm errors; arguably, they should have false alarmed on 100% of target-absent trials that contained a lure. With false alarm rates of about 6% on these trials (Fig. 4b), it is clear that

this was not the case. However, Fig. 4a and b shows that lures did have a significant effect on RTs as well as on error rates. A two-way ANOVA on RTs, with Target Presence and Lure Presence as factors, revealed main effects of both [target presence, $F(1, 17) = 182.37, p < .0001, \eta_G^2 = .49$; lure presence, $F(1, 17) = 45.33, p < .0001, \eta_G^2 = .05$]. However, the interaction was not statistically significant [$F(1, 17) = 0.59, p = .45, \eta_G^2 = .001$] (Fig. 4a). A two-way ANOVA on the arcsine-transformed error rates with the same factors as above revealed a significant effect of lure presence [$F(1, 17) = 8.54, p = .009, \eta_G^2 = .05$], as well as a significant interaction [$F(1, 17) = 5.29, p = .03, \eta_G^2 = .05$]. Bonferroni-corrected multiple comparisons showed that lures caused more errors in target-absent trials [$t(17) = 2.85, p = .02$], but there was no significant difference in target-present trials [$t(17) = -0.17$].

Experiment 1b: The role of consistent versus variable mapping In Experiment 1a, the 8Partition condition was harder than the 8Alone condition. It could be argued that this was the result of 8Partition being a variable-mapping condition in which distractors (lures) on one trial could be targets on the next. The role of mapping was tested in Experiment 1b, which was exactly the same as 1a except that lures were excluded in the 8Partition case, making it a version of a consistent-mapping paradigm. The middle panel of Fig. 3 shows that this did not make much of a difference to the results. Importantly, the 8Partition condition did not get any easier (RTs are actually slightly higher in this version of the experiment, by ~75 ms). Unexpectedly, the 8Alone condition, whose procedure did not differ between the two experiments, became somewhat harder (RTs were ~280 ms slower), making the results of Experiment 1b less statistically reliable than those of 1a. A two-way repeated measures ANOVA on condition (16Alone, 8Alone, and 8Partition) and visual set size (6 and 12) was performed, and when necessary the results were corrected for nonsphericity using the Greenhouse–Geisser correction method, which resulted in noninteger degrees of freedom. We again found a main effect of visual set size [$F(1, 14) = 464.3, p < .001, \eta_G^2 = .25$], but the effect of condition was no longer significant [$F(1.03, 14.4) = 1.2, p = .29, \eta_G^2 = .03$]. The interaction term was also not significant [$F(1.2, 16.7) = 0.58, p = .48, \eta_G^2 < .001$]. Again, these results were most likely driven by the curious change of the 8Alone condition; planned comparisons revealed that the 8Alone and 16Alone conditions no longer differed significantly [$t(14) = 0.86$]. The 8Partition condition did not differ significantly from the 16Alone or the 8Alone condition (all t s $< 1.47, p$ s $> .16$).

Discussion

The results of Experiments 1a and 1b fall between the two extreme answers to the question of whether observers can

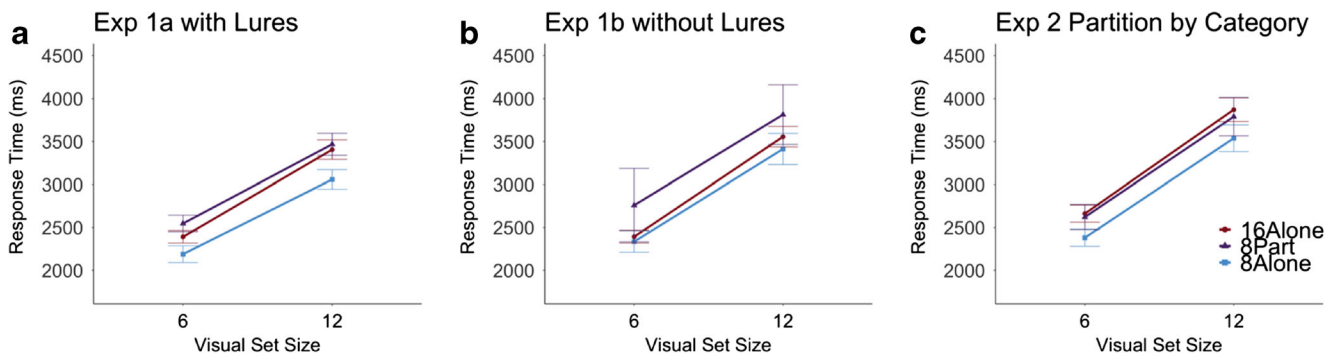


Fig. 3 (Left) During the 8Partition block, observers were significantly slower than in the 8Alone block, but their performance did not differ significantly from the 16Alone block. (Middle) The response time results from Experiment 1b show that even under conditions of consistent mapping, in which lures were removed, performance in the 8Partition block was actually slower than search through the entire set of 16 items, though

the difference was not significant. (Right) In Experiment 2, when the partitions were determined by semantic categories, search through the partition still more closely resembled search through the entire contents of activated long-term memory. Error bars here and throughout the article represent standard errors of the means, unless otherwise noted

maintain two separate memory sets during a hybrid-search task. If they had been able to maintain and use the two sets entirely separately, then the RTs for the 8Partition conditions should have been equivalent to the RTs for the 8Alone condition. Instead, the RTs were quite similar to those in the 16Alone condition, as if the two 8-item lists in the partition condition were simply combined into one 16-item list. However, had the partition not been maintained at all, observers should have chosen the lure items on 75% of the lure-present trials (100% when the target was absent, and 50% on target-present trials). This was clearly not the case. In general, observers were very good at the task (error rates < 7%).

Additionally, the data do not seem to be driven by the classic differences between consistent and variable mapping (Schneider & Shiffrin, 1977). That is, in Experiment 1a the targets from one trial could appear as distractors (or lures) on the next trial. It is known that search is much more efficient in consistent-mapping than in variable-mapping conditions. However, Experiment 1b did not

include lures, and therefore all of the conditions were consistently mapped. Consistent with Experiment 1a, Experiment 1b suggests that observers were unable to search through only a subset of the target items from one trial to the next. Although the difference between the 8Partition and 8Alone conditions was no longer significant, this appears to have been driven by increased RTs in the 8Alone condition, when we compare Experiments 1a and 1b directly, rather than faster RTs in the 8Partition condition, as one would expect from the variable- versus consistent-mapping account.

The data seem to support the view that observers were able to maintain the partition of 16 items into two subsets of eight to the extent that they did not make many lure errors. However, there was a cost in RTs and small cost in accuracy on lure-present trials. Presumably, observers were searching through the entire memory set of 16 items, but if they landed on a target in the wrong context, they were capable of accessing the context and moving on with the search.

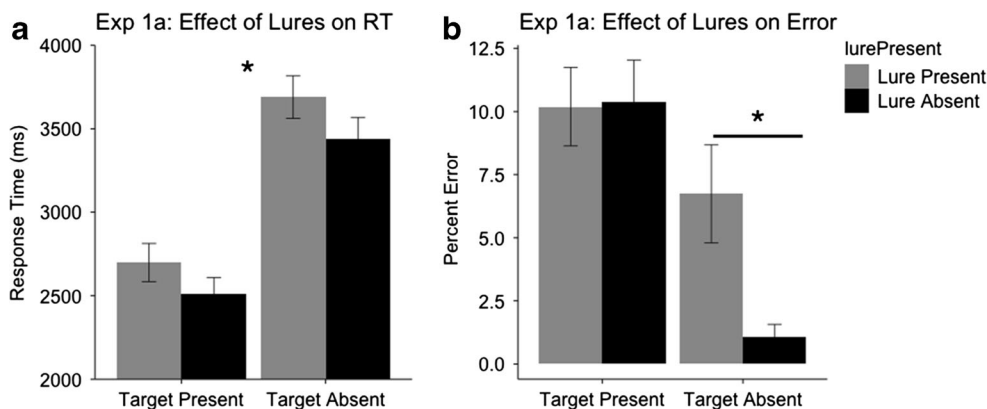


Fig. 4 Within the 8Partition condition in Experiment 1a, lures have a significant effect on response times **a** but no significant effect on accuracy **b**

Experiment 2: Partitioning memory by semantic category

Experiments 1a and 1b demonstrated that observers were either incapable or unwilling to search exclusively through a subset of memory. Although they made relatively few errors in response to targets appearing in the wrong context, search behavior seemed to be more closely aligned with a search through the entire set held in memory rather than a relevant subset. However, thus far the objects and their contexts had no semantic relationship to each other. Perhaps if the divisions in memory were based on well-learned semantic divisions, such as object categories, observers could limit their memory search more easily.

Method

Participants Twenty-one of the observers in Experiment 1a (mean age 29.6 years, 13 females, eight males) were recruited from the Harvard Decision Science Laboratory study pool. Five of these observers had extraordinarily high error rates in at least one of the conditions (above two standard deviations from the mean), and were therefore not included in the analysis. This left us with 16 participants. All participants gave informed consent and were compensated \$10/h for their participation. The Partners Healthcare Corporation Institutional Review Board approved all experimental procedures.

Stimuli and apparatus The observers were run at the Harvard Decision Science Laboratory on Dell Optiplex 9020s running Windows 7, 64-bit edition. Stimuli were presented on 19-in. LCD monitors set to a resolution of 1,440 × 900 with a refresh rate of 60 Hz. The stimulus items were drawn from several categories of objects: “Jewelry Store,” “Grocery Store,” “Candy Store,” and “Kitchen Items.” Each of the categories contained 200 objects pulled from the Internet as well as from the photographs of objects used in the picture memory experiments of Brady et al. (2008), and they subtended approximately 3.2°. Four background scenes were additionally chosen from Google Image to serve as context for the memory subsets. These background scenes matched the contexts of the objects.

Procedure As in Experiment 1, the observers in Experiment 2 completed three blocks of trials (8Alone, 8Partition, and 16Alone). In Experiment 2, however, the memory groups were all semantically related and could fall into the categories mentioned above: “Jewelry Store,” “Grocery Store,” “Candy Store,” or “Kitchen.” All distractors were pulled from the same category. During the 8Partition block, two categories were used and 100 items were chosen from each of the categories, in order to match the frequency at which any particular

item could be seen with the frequencies in the other, single-category blocks. The memorization portion was the same as in Experiment 1, except that the subset comparison test was no longer necessary, since distinguishing between the groups was trivial. All other procedures were identical to those of Experiment 1B. That is, no lures were included in this experiment.

Results

An ANOVA on RTs (Fig. 3, right) comparing the three conditions (8Alone, 8Partition, and 16Alone) and the two visual set sizes (6 and 12) showed significant effects of condition [$F(2, 30) = 4.4, p = .02, \eta_G^2 = .05$] and visual set size [$F(1, 15) = 360.8, p < .001, \eta_G^2 = .51$]. However, the interaction was not significant [$F(2, 30) = 0.3, \eta_G^2 < .001$]. Planned comparisons showed that the 16Alone and 8Alone conditions differed significantly [$t(15) = 3.78, p = .002$]. Again, the 8Partition condition differed significantly from the 8Alone condition [$t(15) = 2.25, p = .04$] but was not statistically distinguishable from the 16Alone condition [$t(15) = 0.35, p = .73$].

Discussion

Even when using divisions based on semantically congruent categories (e.g., jewelry in the jewelry store or produce in the grocery store), search was still indistinguishable from search of the entire memory set. Additionally, RTs in the 8Partition condition differed significantly from those in the 8Alone condition. Therefore, the categorical subsets were not sufficient to allow observers to limit their search exclusively to the currently relevant subset of items in memory. There were no lures in Experiment 2 because all of the distractors were pulled from the same categories.

Experiment 3a and 3b: Taking items off the list

At the outset of this article, we invoked the shopping list as a real-world example of hybrid search. However, there is a fundamental difference between our shopping list experiences and the experiments discussed thus far. Once you find the carton of milk, you can remove it from the list. Unlike the experimental situation, you are not going to randomly find yourself acquiring milk every few trials. Is this task, with an ever-shrinking memory set, more natural than the subset switching tasks in Experiments 1 and 2? In this task, as you cross items off a mental shopping list, the list becomes smaller. If the list relevant to the hybrid search also becomes smaller, then RTs should become faster as the number of remaining items shrinks. Experiments 3a and 3b search for (and fail to find) this effect.

Method

Participants For Experiments 3a and 3b, 13 observers apiece were recruited (Exp. 3a: mean age 29.6 years, seven females, six males; 3b: mean age 31.3 years, eight females, five males) from the general population in Cambridge, Massachusetts. One of the observers in Experiment 3a had high error rates (above two standard deviations from the mean), and was therefore not included in the analysis. All participants gave informed consent and were compensated \$10/h for their participation. The Partners Healthcare Corporation Institutional Review Board approved all experimental procedures.

Stimuli and apparatus Observers were run at the Visual Attention Laboratory of Brigham & Women's Hospital and Harvard Medical School on 24-in. iMac computers running OS X 10.8.5. The stimuli were chosen from the over 2,000 objects used in the picture memory experiments of Brady et al. (2008), and they subtended approximately 3.2° at a 60-cm viewing distance.

Procedure In Experiments 3a and 3b, observers completed eight blocks (one practice and seven experimental). In the practice block, observers memorized four objects and completed 16 search trials. In each experimental block, observers memorized 16 new objects and completed 64 trials. In all blocks, observers were informed that, once a target appeared in a search trial, it would never again be relevant to the task and could be “forgotten.” In Experiment 3a, each target appeared only once; in Experiment 3b, objects that had already appeared as targets could show up again as lures. That is, the milk might reappear but, since you had already placed one carton in your cart, the new milk lure should not be acquired. Observers searched for their targets in visual displays of eight and 16 items. Because the targets were relevant only once, we decreased the target prevalence to 25%. This increased the average number of data points per block per target. The purpose of this design was to judge whether observers were able to limit their memory searches to only the objects currently relevant. That is, as objects were being thrown out of memory, would RTs become quicker at the end of the block than at the beginning?

Results

Figure 5 shows RTs as a function of relevant memory set size for Experiments 3a and 3b. It is clear that RTs did not decrease as the relevant memory set size decreased. Observers' average slopes did not differ significantly from zero [Exp. 3a: visual set size 8, $t(11) = 1.1$, $p = .28$; visual set size 16, $t(11) = -0.21$, $p = .84$; Exp. 3b: visual set size 8, $t(14) = 0.44$, $p = .66$; visual set size 16, $t(14) = 0.26$, $p = .79$].

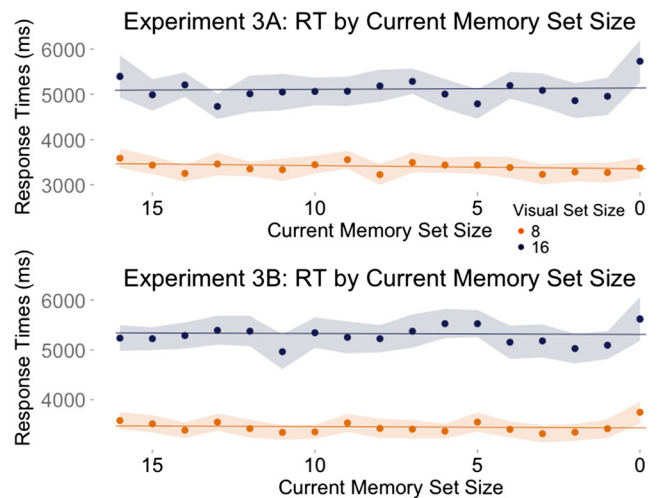


Fig. 5 Observer response times (RTs) on target-absent trials in Experiments 3a and 3b. The x-axis is reversed to reflect the progression through the block. Shaded regions represent one standard error above and below the mean RT at each memory set size. Observers do not seem to have speeded up as the target set decreases

If observers could not remove items one at a time, were they severely impaired when they must learn a new memory set for each block? If they were, we would expect RTs to increase from block to block, because the effective memory set size would increase from block to block. Figure 6 shows that this did not occur. Again we tested whether the average slope was different from zero. In Experiment 3A we do indeed find a significant effect; however, it was in the opposite direction from what one would expect if observers were accumulating items in their memory sets across blocks [Exp. 3a: visual set size 8, $t(11) = -3.2$, $p = .008$; visual set size 16, $t(11) = -3.49$, $p = .005$; Exp. 3b: visual set size 8, $t(11) = -0.36$, $p = .72$; visual set size 16, $t(11) = 0.42$, $p = .68$].

Figure 7 shows that, as in Experiment 1, observers were not greatly disturbed by lures in this experiment. Even if they could not reduce the effective set size, when they stumbled upon a lure item, they were able to refrain from marking it as a target. A two-way ANOVA with Target Presence and Lure Presence as factors showed no effect of lure presence on RTs but a significant effect on error rates within Experiment 3b [RTs, $F(1, 14) = 0.14$, $p = .71$, $\eta_G^2 < .001$; arcsine-transformed error rates, $F(1, 14) = 6.8$, $p = .02$, $\eta_G^2 = .05$].

Discussion

Apparently, observers could not reduce the memory set size on each trial in a way that would allow them to be more efficient in their search. This is in contrast to a previously reported study in which observers were asked to indicate the new object on each trial. In this case, memory set size and RT both increased on each trial, because observers needed to remember all of the old items in order to identify the new one. Observers proved capable of adding to their memory set sizes,

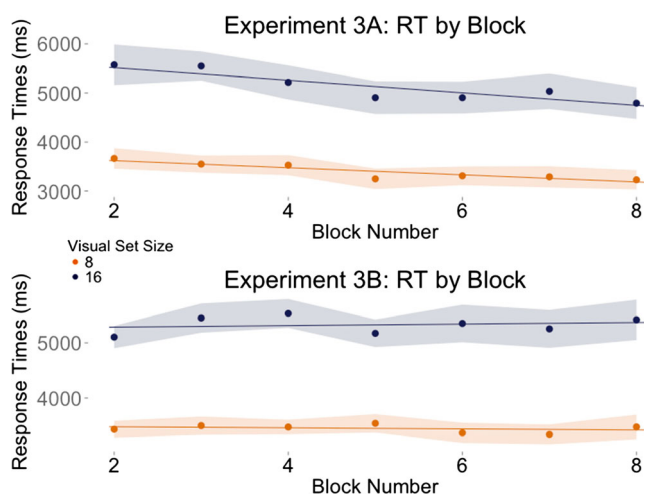


Fig. 6 Response time does not increase over the blocks. This indicates that observers were not carrying over extra items in memory between blocks

one at a time, for hundreds of trials (Wolfe et al., 2015). Decreasing a memory set does not appear to work the same way. Although the results indicate that observers cannot decrease the effective size of the memory set from trial to trial, it is interesting that, from block to block, they can successfully disregard the entire previous memory set while searching for items on the current list. We do not see a positive slope across blocks, as would be expected if observers were unable to put aside obsolete memory sets. In fact, in Experiment 3a we actually found the opposite effect—a negative slope as a function of block. This decrease is probably due to a standard practice effect, although we did not observe the same pattern in Experiment 3b. Regardless, it is clear that observers were not adding each new set of objects to an increasingly huge omnibus memory set.

In Experiment 3b an item that had served as a target once before could appear again as a lure. As in our earlier experiments, observers proved capable of ignoring these previous targets most of the time. When we compare lure trials to nonlure trials in Experiment 3b, observers paid a modest cost in terms of errors; however, there was no significant difference

in RTs. Once again, this shows that observers were capable of distinguishing lures from targets. If they had not been, they would have suffered much larger error rates.

Experiments 4a–4d: Explicit versus implicit partition switches

Thus far, observers have proven unable or unwilling to limit their memory searches on a trial-by-trial basis. It is possible that it is not worth it for an observer to switch between memory partitions from one trial to the next. A savings of a few hundred milliseconds, although massive to a cognitive scientist, may not be salient to the observer. Perhaps we could make it more salient if a subset of the memory set became relevant for an extended number of trials. Continuing with our grocery store example, one would not normally switch back and forth between the produce section and the deli. It would be more typical to search the deli section for deli items, and then to search the produce aisle for the fruits and vegetables. In addition, these experiments would generalize our results to different set sizes.

Method

Participants

In Experiment 4, 75 observers total were run (Exp. 4a: mean age 26.2 years, 13 females, nine males; Exp. 4b, mean age 20.3 years, 12 females, nine males; Exp. 4c, mean age 23.2 years, ten females, one nonbinary gender, six males; Exp. 4d, mean age 20.1 years, 12 females, three males). One observer was removed from the final analysis of each of the four experiments for an errors rate greater than two standard deviations away from the mean. In addition, some of the observers did not complete the entire experiment. For this reason, two observers from Experiment 4b, three observers from Experiment 4c, and one observer from Experiment 4d were excluded from the final analysis. All participants gave

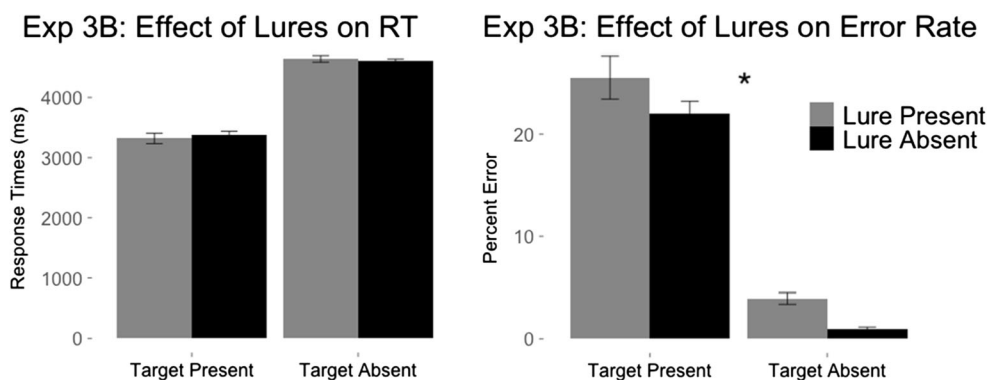


Fig. 7 The effects of lures are once again small: There is a significant difference in error rates within Experiment 3b between the lure and nonlure trials, but no effect on response times

informed consent and were compensated with course credit for their participation.

Stimuli and apparatus

The observers in Experiment 4 were run at the University of Utah on Dell Optiplex 990 computers running Windows 7, 64-bit edition. Observers were seated ~20 in. from an LCD monitor that subtended 18.5×10.5 in., with a $1,920 \times 1,080$ resolution and a 60-Hz refresh rate. Again, the stimuli were chosen from the over 2,000 objects used in the picture memory experiments of Brady et al. (2008), which subtended approximately 2.4° . Six background scenes were additionally chosen from Google Image to serve as context for the memory subsets.

Procedure

Experiments 4a and 4b Once again, observers completed a total of three blocks—8Alone, 16Alone, and 8Partition. The memorization portion was comparable to those in the previous experiments. The search portion was also similar, with a few critical differences, particularly in the partition block. For the first few trials, as in the previous experiments, the context varied randomly on each trial. At a particular point, search trials began to be divided into sets of 20 trials, with a single context held constant across those 20 trials. Thus, for 20 trials only one of the partitions would be relevant, followed by 20 trials for which the other partition would be relevant, and so forth. In Experiment 4a, observers were not explicitly told that they would see these runs of search trials. They would have to notice that they were in a run and adjust search, if possible. In Experiment 4b, observers received explicit information as a new run of trials began, alerting them which context they would see for the next 20 trials. Both experiments concluded with a set of trials on which the context again varied randomly on each trial. The observers completed 160 experimental trials in both experiments.

Experiments 4c and 4d Experiments 4c and 4d replicated Experiments 4a and 4b using different set sizes. In these experiments, observers memorized either 18 objects alone, two objects alone, or one set of two objects and one set of 16 objects.

Results

Experiment 4a: Implicit knowledge of subset relevance—8/8 partition In Experiment 4 there was finally some evidence for an ability to successfully, if not completely, partition the memory set. Figure 8 shows RTs for all four versions of this experiment. The top left panel in Fig. 8 shows the results for Experiment 4a. A repeated measures ANOVA with

condition and visual set size as levels showed significant effects of condition [$F(2, 40) = 6.00, p = .005, \eta_G^2 = .02$] and set size [$F(1, 20) = 125.7, p < .001, \eta_G^2 = .28$]. However, there was not a significant interaction of these two factors [$F(2, 40) = 0.99$]. The planned comparisons between the conditions revealed a significant difference between the 8Alone and 16Alone conditions [$t(20) = 2.17, p = .04$], as well as a difference between the 8Partition and 16Alone conditions [$t(20) = 3.79, p = .001$]. However, there was no significant difference between the 8Partition and 8Alone conditions [$t(20) = 1.39, p = .18$], indicating that during the 8Partition condition, observers were able to limit their memory search to only eight items.

Experiment 4b: Explicit knowledge of subset relevance—8/8 partition In Experiment 4b the switch in memory sets was made explicitly clear before a new run of trials began. RTs are shown at the top right of Fig. 8. A repeated measures ANOVA with Condition and Visual Set Size as factors showed significant effects of condition [$F(2, 34) = 10.60, p < .001, \eta_G^2 = .09$] and set size [$F(1, 17) = 315.5, p < .001, \eta_G^2 = .64$]. There was no significant interaction of these two factors [$F(2, 34) = 0.83$]. The planned comparisons between the conditions revealed a marginally significant difference between the 8Alone and 16Alone conditions [$t(17) = 2.09, p = .05$], as well as a significant difference between the 8Partition and 16Alone conditions [$t(17) = 5.93, p < .001$]. Curiously, the 8Partition condition was also marginally faster than the 8Alone condition [$t(17) = 2.11, p = .05$]. Apparently, observers partitioned the 16 items in the 8Partition conditions very successfully. It may be that the explicit instructions in this condition strongly focused the observers on the relevant memory set (see Exp. 4d below).

Experiment 4c: Implicit knowledge of subset relevance—2/16 partition In Experiments 4c and 4d, the partition condition split the memory set into two very unequal subsets. With 18 total items, it is unlikely that we could see the effect of reducing the memory set to 16. Thus, we will focus on the trials in which the relevant memory set dropped to two. Do those “2Partition” trials look like the 18Alone or the 2Alone trials? Average RTs are shown at bottom left of Fig. 8 for the four types of block.

An ANOVA with the four Conditions (18Alone, 16Partition, 2Partition, and 2Alone) and the Visual Set Size (6 and 12) as factors showed main effects of condition [$F(3, 36) = 27.34, p < .001, \eta_G^2 = .32$] and visual set size [$F(1, 12) = 143.85, p < .001, \eta_G^2 = .28$]. Additionally, these factors interacted significantly [$F(3, 36) = 5.86, p = .003, \eta_G^2 = .01$]. Planned comparisons showed a significant difference between 18Alone and 2Partition [$t(12) = 5.20, p < .001$]. However, the 2Partition trials were also significantly slower than the

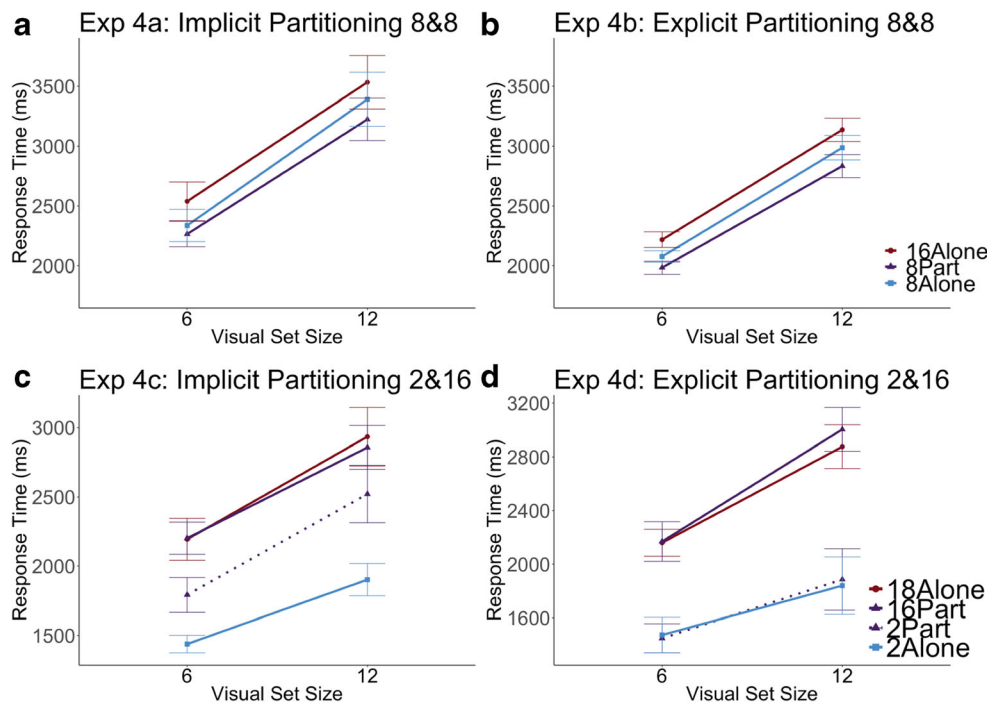


Fig. 8 (Top) In Experiments 4a and 4b, search in the 8Partition condition differed significantly from search through 16Alone arrays. In Experiment 4b, the 8Partition condition was also somewhat mysteriously faster than search through 8Alone arrays; this was not the case in Experiment 4a (Bottom left) In Experiment 4c, search in the 2Partition condition

differed significantly from search through the entire set (18Alone), as well as from search exclusively through the partition (2Alone). (Bottom right) In Experiment 4d, search in the 2Partition condition differed significantly from search through the entire set (18Alone), as well as from search exclusively through the partition (2Alone)

2Alone trials [$t(12) = 3.67, p = .003$]. Thus, as in Experiment 4a, even without explicit information, observers became more efficient when only the smaller partition was relevant. However, they could not achieve the 2Alone speed of search, seen when the memory set was always two items. The more dramatic difference between the 18Alone and 2Alone results makes it easier to see this effect in Experiment 4c than in Experiment 4a, but the pattern is essentially the same.

Experiment 4d: Explicit knowledge of subset relevance—2/16 partition The average RTs for the four types of block in Experiment 4d are shown at the lower right of Fig. 8. An ANOVA once again showed a significant effect of condition on RTs [$F(3, 36) = 65.4, p < .001, \eta_G^2 = .39$], as well as an effect of visual set size [$F(1, 12) = 58.40, p < .001, \eta_G^2 = .21$] and, once again, a significant interaction [$F(3, 36) = 10.00, p < .001, \eta_G^2 = .02$]. Multiple comparisons showed a significant difference between the 18Alone and 2Partition conditions [$t(12) = 8.36, p < .001$]. However, as is clear from the figure, there was no longer a significant difference between 2Partition and 2Alone trials [$t(12) = 0.05$]. Given enough information about the situation, observers are able to limit their memory search to the relevant items. This is comparable to the results of Experiment 4b, though without the mysterious advantage for the partition condition seen in that experiment.

Discussion

Throughout Experiment 4, we finally obtained evidence of observers' ability to restrict their memory search to only the relevant items. Without explicit instructions (Exps. 4a, 4c), observers could partition memory imperfectly. In Experiments 4b and 4d, with explicit information, partition was essentially perfect (or "better" than perfect, in Exp. 4b). The variation in set size did not change the basic pattern of results. The "perfect" partition is akin to what happens in hybrid-search experiments when a block ends and a new memory set is trained for the next block of trials. It is clear from the data in older hybrid experiments that observers are able to put aside the old list quickly and, apparently, entirely. Experiments 4b and 4d show that this can be done without forgetting the previous memory set.

General discussion

In hybrid search, we propose that observers search through a memory set that is resident in activated long-term memory (Cowan, 1995). We can reject the hypothesis that working memory is the home of the memory set, because, as we showed again in the present experiments, it is easy to perform hybrid-search tasks with memory sets much larger than the

approximately four items that working memory can accommodate. Moreover, work from our group has shown that observers' hybrid-search performance was largely unaffected by holding a working memory load in mind while performing a hybrid-search task (Drew, Boettcher, & Wolfe, 2016). The experiments presented here asked whether observers can restrict memory search to part of a memory set when the items in that part are the only ones relevant for the current visual search task. The results provide evidence against either of the extreme, "yes" or "no" answers to this question, and support a more nuanced position.

It appears that ALTM is not partitioned in hybrid search on a trial-by-trial basis—at least, not in a way that makes it possible to restrict search to one subset of ALTM or to shuttle subsets between ALTM and long-term memory on a trial-by-trial basis. In Experiment 1 we asked whether observers could search through one of two 8-item memory sets. The RTs in this 8Partition condition more closely resembled a search through the entire memory set (16Alone) than a search through the relevant subset (8Alone). Even when objects were semantically linked to the context they appeared in (Exp. 2), search through the partitioned set did not differ from search through the entire set. It seems likely that observers were searching through the entire contents of their memory set, and when they landed upon a potential target, they took an extra moment to decide whether the item was in fact a relevant target or a lure.

If a part of the memory set is going to be relevant again, it appears to remain in ALTM and to intrude into the current memory search even though this is not useful. In contrast, Experiment 3 and previous work showed that a memory set that is no longer relevant can be effectively removed from ALTM. At least in an experiment in which observers memorized a series of memory sets, they appeared able to remove the list from the previous block from ALTM and to substitute the new list for the current block. We found no evidence that memory sets from one block became aggregated in ALTM with memory sets from previous blocks. That is, observers were not forced to search a superset containing all of the previous targets. As noted, this finding is consistent with the ability of observers in other hybrid-search experiments to run multiple blocks with different hybrid-search memory sets in a single session (e.g., in Wolfe, 2012).

There is evidence that items that have been successfully removed from ALTM are not completely forgotten. Cunningham, Pailian, and Egeth (2014) showed that when observers are retrained on a subset of their initial target items, RTs decrease, indicating that they are able to disregard the items that are no longer relevant between training phases. In our Experiments 4a, 4b, and 4d, after a single training session, observers were able to shuttle subsets in and out of ALTM. This is in line with previous work looking at context-dependent recall. That is, when participants are asked to memorize a categorized list and during test are cued with a single

item (or category) and asked to recall all other items from said category, participants' performance is enhanced (Hudson & Austin, 1970). As of now, it is unclear whether such a benefit would be present if recall were probed on a trial-by-trial basis; however, our results suggest that there would be no such benefit. In Experiment 4c we found that observers could reduce their memory search, although there was still a cost of partitioning information relative to the 2Alone condition. When observers were successful in restricting memory search to a subset of either two or eight items, we speculate that the other items must have been waiting in the wings of long-term memory for the start of a block in which they would, once again, be relevant. The fact that partitioning was imperfect in Experiment 4c could mean that observers sometimes did not bother to move the irrelevant items out of the way (as was apparently the case in Exps. 1 and 2). It could also be that the metaphor of "moving" items from one place to another should not be taken too literally. The process may be more graded. Even when partitioning is imperfect or absent (Exps. 1–3), observers remain capable of keeping track of the subset membership of an item, thus preventing major interference from lures. However, under these conditions of faster change and/or less information, observers do not seem capable of banishing the currently irrelevant items from ALTM.

When observers fail to partition, one contributing factor may be the cost of switching between memory sets. If the act of switching between target sets costs some amount of time and effort, observers may implicitly decide that it is more efficient to avoid this switching cost by simply looking for any potential target item. This hypothetical memory set switching cost would be similar to the "task switching" costs seen in other types of experiments (see Monsell, 2003). Consistent with this idea, we observed small but reliable RT and accuracy costs in our experiments when lure items were present. Observers may only determine the target subset after they have identified a potential target from the superset in ALTM. Only when one is sure that a large number of trials will take place with a specific target set is it worth adjusting the contents of ALTM to hold just the subset. This was the situation in Experiments 4a, 4b, and 4d. In Experiments 1 and 2, there was enough uncertainty about the context of the next trial to maintain the entire set in ALTM and, thus, avoid paying the context-switching cost. In Experiment 4c, participants did see some benefit of the repeated context; however, this did not reach the levels of the 2Alone condition. This is in fact a bit curious, and as of now we cannot say with certainty why participants failed to completely optimize their memory search. One possibility is that the cost of switching is directly related to how many items must be "thrown out" of ALTM. In Experiment 4c, participants had to discard 16 items in order to successfully search exclusively through the partitioned set. Perhaps the cost of throwing out these

items was not worth the few hundred milliseconds saved during the search, particularly when it had not been made explicitly clear that they would not need those 16 items for several trials (i.e., Exp. 4c vs. 4d). In future work we hope to explore this idea by manipulating the timing between the onset of the context and the search array. These events occurred simultaneously in the present experiment, but one might imagine that if the two events were separated by a 1-s delay, it might be logical to partition memory prior to the onset of the search information.

The model of hybrid search originally proposed by Cunningham and Wolfe (2014) can be applied to these partition experiments. Cunningham and Wolfe proposed that hybrid search takes place in a series of three steps: a guided visual search, followed by massive parallel categorization of the selected item, and finally logarithmic memory search if it is determined that the item might be a member of the memory set. Changing the memory set for Step 3 seems to be expensive enough that it is not worth doing on a trial-by-trial basis. Experiment 2 illustrates this nicely. Suppose an item is identified as a piece of jewelry in Step 2. In the transition to Step 3, the observer must decide whether this piece of jewelry could be a member of the memory set. Even though the background tells the observer whether she is in a jewelry or, say, a candy store setting, it is simply easier to check the entire jewelry-plus-candy memory set. The observer can respond positively if the item is in the memory set and the background is correct. If the background is a candy story, then the item is a lure. The response is withheld at the cost of a slightly longer RT and/or a higher error rate when lures are present. In Experiment 4, the cost of adjusting the memory set needed to be made only once every 20 trials. That cost was worth paying. Under those conditions, observers shuffled the irrelevant objects out of ALTM and continued with their search as if they were looking only for the subset of items.

Returning to the initial example of the shopping list in the supermarket, we can ask whether the results presented here usefully describe the process of supermarket hybrid search. We systematically attempted to emulate some aspects of a trip to the grocery store (semantically related subsections, eliminating items one at a time, and long runs in a particular context) and found that long runs in a particular context are necessary for partitioning memory search. However, one suspects that the analogy is imperfect. Because the time scale of a shopping expedition is much greater than the time scale of the trials in this study, it may well be that items can be dropped from the master shopping list as they are deposited in the cart. Moreover, the experiments described here used arrays of items placed randomly on background contexts. We know that scene structure and contents exert a strong influence on search (Biederman, Teitelbaum, & Mezzanotte, 1983; Castelano & Heaven, 2011; Henderson & Ferreira, 2004). It would not be surprising to find that the scene context exerts an effect on

hybrid search. Answers to these questions await an experiment in which observers navigate through a scene (real or virtual) collecting multiple items from a list.

References

- Bäuml, K.-H., & Aslan, A. (2006). Part-list cuing can be transient and lasting: The role of encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*, 33–43. doi:10.1037/0278-7393.32.1.33
- Biederman, I., Teitelbaum, R. C., & Mezzanotte, R. J. (1983). Scene perception: A failure to find a benefit from prior expectancy or familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *9*, 411–429. doi:10.3758/0278-7393.9.3.411
- Boettcher, S. E. P., & Wolfe, J. M. (2015). Searching for the right word: Hybrid visual and memory search for words. *Attention, Perception, & Psychophysics*, *77*, 1132–1142. doi:10.3758/s13414-015-0858-9
- 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 Sciences*, *105*, 14325–14329. doi:10.1073/pnas.0803390105
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. doi:10.1163/156856897X00357
- Castelano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin & Review*, *18*, 890–896. doi:10.3758/s13423-011-0107-8
- Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*, *36*, 28–71. doi:10.1006/cogp.1998.0681
- Cowan, N. (1995). *Attention and memory: An integrated framework (Oxford Psychology Series)*. New York: Oxford University Press.
- Cunningham, C. A., Pailian, H., & Egeth, H. (2014). *Characterizing representations in activated long-term memory*. Article presented at the annual meeting of the Psychonomic Society, Long Beach, CA.
- Cunningham, C. A., & Wolfe, J. M. (2014). The role of object categories in hybrid visual and memory search. *Journal of Experimental Psychology: General*, *143*, 1585–1599. doi:10.1037/a0036313
- Dalton, P. (1993). The role of stimulus familiarity in context-dependent recognition. *Memory & Cognition*, *21*, 223–234. doi:10.3758/BF03202735
- Drew, T., Boettcher, S. E. P., & Wolfe, J. M. (2016). Searching while loaded: Visual working memory does not interfere with hybrid search efficiency but hybrid search uses working memory capacity. *Psychonomic Bulletin & Review*, *23*, 201–212. doi:10.3758/s13423-015-0874-8
- Drew, T., & Wolfe, J. M. (2013). Hybrid search in the temporal domain: Evidence for rapid, serial logarithmic search through memory. *Attention, Perception, & Psychophysics*, *76*, 296–303. doi:10.3758/s13414-013-0606-y
- Eich, E. (1985). Context, memory, and integrated item/context imagery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*, 764–770. doi:10.1037/0278-7393.11.1-4.764
- Epskamp, S., & Nuijten, M. B. (2016). *statcheck: Extract statistics from articles and recompute p values (Software)*. Retrieved from cran.r-project.org/package=statcheck.
- Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and under water. *British Journal of Psychology*, *66*, 325–331. doi:10.1111/j.2044-8295.1975.tb01468.x

- Godden, D. R., & Baddeley, A. D. (1980). When does context influence recognition memory? *British Journal of Psychology*, *71*, 99–104. doi:10.1111/j.2044-8295.1980.tb02735.x
- Grant, H. M., Bredahl, L. C., Clay, J., Ferrie, J., Groves, J. E., McDorman, T. A., & Dark, V. J. (1998). Context-dependent memory for meaningful material: Information for students. *Applied Cognitive Psychology*, *12*, 617–623. doi:10.1002/(SICI)1099-0720(199812)12:6<617::AID-ACP542>3.0.CO;2-5
- Henderson, J. M., & Ferreira, F. (2004). Scene perception for psycholinguists. In *The interface of language, vision, and action* (pp. 1–58). New York: Psychology Press.
- Hudson, R. L., & Austin, J. B. (1970). Effect of context and category name on the recall of categorized word lists. *Journal of Experimental Psychology*, *86*, 43–47. doi:10.1037/h0029990
- Krafka, C., & Penrod, S. (1985). Reinstatement of context in a field experiment on eyewitness identification. *Journal of Personality and Social Psychology*, *49*, 58–69.
- Lawrence, M. (2013). ez: Easy analysis and visualization of factorial experiments (R package version). Retrieved from <https://scholar.google.com/scholar?cluster=310992833082440893&hl=en&oi=scholar>
- Leite, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception, & Psychophysics*, *72*, 246–273. doi:10.3758/APP.72.1.246
- Macken, W. J. (2002). Environmental context and recognition: The role of recollection and familiarity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 153–161. doi:10.1037/0278-7393.28.1.153
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*, *7*, 134–140. doi:10.1016/S1364-6613(03)00028-7
- Nakashima, R., & Yokosawa, K. (2013). Visual search in divided areas: Dividers initially interfere with and later facilitate visual search. *Attention, Perception, & Psychophysics*, *75*, 299–307. doi:10.3758/s13414-012-0402-0
- Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*, 614–621. doi:10.1016/j.visres.2005.08.025
- Nosofsky, R. M., Cox, G. E., Cao, R., & Shiffrin, R. M. (2014). An exemplar-familiarity model predicts short-term and long-term probe recognition across diverse forms of memory search. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *40*, 1524–1539. doi:10.1037/xlm0000015
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. doi:10.1163/156856897X00366
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108. doi:10.1037/0033-295X.85.2.59
- Ratcliff, R., & Starns, J. J. (2013). Modeling confidence judgments, response times, and multiple choices in decision making: Recognition memory and motion discrimination. *Psychological Review*, *120*, 697–719. doi:10.1037/a0033152
- Roediger, H. L., III. (1973). Inhibition in recall from cueing with recall targets. *Journal of Verbal Learning and Verbal Behavior*, *12*, 644–657. doi:10.1016/S0022-5371(73)80044-1
- Roediger, H. L., III, & Tulving, E. (1979). Exclusion of learned material from recall as a postretrieval operation. *Journal of Verbal Learning and Verbal Behavior*, *18*, 601–615. doi:10.1016/S0022-5371(79)90334-7
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review*, *84*, 1–66. doi:10.1037/0033-295X.84.1.1
- Slamecka, N. J. (1968). An examination of trace storage in free recall. *Journal of Experimental Psychology*, *76*, 504–513. Retrieved from www.ncbi.nlm.nih.gov/pubmed/5650563
- Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning & Memory*, *5*, 460–471. doi:10.1037/0278-7393.5.5.460
- Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*, 342–353. doi:10.3758/BF03197465
- Smith, S. M., & Vela, E. (1992). Environmental context-dependent eyewitness recognition. *Applied Cognitive Psychology*, *6*, 125–139. doi:10.1002/acp.2350060204
- Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*, 203–220. doi:10.3758/BF03196157
- Solman, G. J. F., & Kingstone, A. (2017). Spatial partitions systematize visual search and enhance target memory. *Attention, Perception, & Psychophysics*, *79*, 449–458. doi:10.3758/s13414-016-1232-2
- Townsend, J. T. (1990). Serial versus parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished. *Psychological Science*, *1*, 46–54. doi:10.1111/j.1467-9280.1990.tb00067.x
- Vö, M. L.-H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision* *10*(3), 14:1–13. doi:10.1167/10.3.14
- Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Dordrecht: Springer. doi:10.1007/978-0-387-98141-3
- Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*, 202–238. doi:10.3758/BF03200774
- Wolfe, J. M. (2007). Integrated models of cognitive systems. In W. D. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York: Oxford University Press.
- Wolfe, J. M. (2012). Saved by a log: How do humans perform hybrid visual and memory search? *Psychological Science*, *23*, 698–703. doi:10.1177/0956797612443968
- Wolfe, J. M., Boettcher, S. E. P., Josephs, E. L., Cunningham, C. A., & Drew, T. (2015). You look familiar, but I don't care: Lure rejection in hybrid visual and memory search is not based on familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 1576–1587. doi:10.1037/xhp0000096
- Wolfe, J. M., Vö, M. L.-H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*, 77–84. doi:10.1016/j.tics.2010.12.001