Attention, Perception, & Psychophysics
https://doi.org/10.3758/s13414-025-03077-8

Mixing it up: Intermixed and blocked visual search tasks produce similar results

Jeremy M. Wolfe[1,2] · Injae Hong[1,2] · Ava A. Mitra[1] · Eduard Objio Jr.[3] · Hula Khalifa[3] · Yousra Ali[4]

Accepted: 26 March 2025

Abstract

We have decades of visual search data from experiments where observers look for targets among distractors. Typically, observers are tested in blocks of several hundred trials, and conclusions about underlying mechanisms are inferred from Reaction Time × Set Size functions and errors. However, in the real world, searchers almost never search for the same target or the same type of target hundreds of times in a row. You search for cereal, then milk, then a bowl. Do the rules derived from blocks of trials apply when search tasks are mixed? Here, we compare mixed and blocked conditions in five experiments. In Experiment 1, four different feature searches are tested. In Experiments 2 and 3, the target was the same in four tasks that were defined by different distractor sets. In Experiment 4, different targets are searched for amongst distractors that remained constant across trials. Finally, in Experiment 5, we allowed participants to choose which of four tasks to perform on each trial. In each experiment, there was no qualitative change in search behavior as a function of the mixed/blocked manipulation. The results support the generality of rules of search learned from blocked trials. However, these results do pose a challenge to simple adaptive models of search termination.

Keywords Eye movements and visual attention · Reaction time methods · Visual search

Introduction

In a typical visual search task, observers look for a target of some sort in a display or scene containing distractors (Wolfe, 2023). As a technical term, visual search seems to have entered the literature almost a century ago (e.g., "Search: A Function Intermediate Between Perception and Thinking" Kingsley, 1926, 1932) though, of course, people always knew that they needed to search for things. Precursors of our modern interests can be found, for instance, in the works of Aristotle and Lucretius (Hatfield, 1998). Visual search became more active as a research topic after World War II with the work of Mackworth (1948) leading the way. Both basic and applied search tasks attracted research interest in the 1950 s and 1960 s (Green & Anderson, 1956; Koopman, 1956; Neisser, 1963; Neisser et al., 1963; Thomas & Lansdown, 1963; Tuddenham & Calvert, 1961). In the 1970s and 1980s, the topic became a standard topic in vision/attention research (Egeth, 1977; Egeth et al., 1972; Shiffrin & Gardner, 1972; Sperling et al., 1971; Sternberg & Scarborough, 1971; Townsend, 1971). Publication of Treisman's *A Feature-Integration Theory of Attention* (Treisman & Gelada,

This research was supported by NSF-2146617 and NIH-EY017001.

Significance: Most laboratory visual search experiments involve performing many trials of the same task in a block. Most real-world visual search tasks do not. They involve switching targets, distractors, or both with almost every search. Do the rules, uncovered using blocked search, apply to mixed search? In a series of five experiments, we find that mixing four different tasks does not qualitatively change search behavior. Even on target-absent trials, mixed search behavior looks like blocked search behavior. It is encouraging to think that decades of work on the rules of search might apply beyond the artificial world of the lab.

✉ Jeremy M. Wolfe
jwolfe@bwh.harvard.edu

[1] Visual Attention Lab, Brigham and Women's Hospital, Boston, MA 02135, USA

[2] Harvard Medical School, Boston, MA, USA

[3] University of Massachusetts, Amherst, MA, USA

[4] Josiah Quincy Upper School, Boston, MA 02116, USA

Published online: 08 May 2025

1980) may be the event that put the topic into every textbook covering vision and/or cognition.

The thousands of subsequent papers on visual search are diverse in form and content, but they typically begin by invoking some real-world search task, such as finding your car in the parking lot or locating the pickles in your refrigerator. In their modal form, these empirical papers then go on to describe experiments where the search tasks are performed in blocks of dozens, hundreds, or even thousands of the same type of trials (Wolfe et al., 2010). However, there is a mismatch here between the motivating examples and the experimental method. No one does a block of 200 pickle searches in their kitchen, one after the other. You search for the pickles, then the bread, then a plate, and so on. The central question of this paper is whether the rules that govern mixed searches in the real world are similar to those uncovered by studying blocks of visual searches in the lab.

Of course, there are some real-world search tasks that are performed in blocks of many similar searches. Popular examples in the literature include screening tasks in radiology (Berbaum et al., 1990; Gandomkar & Mello-Thoms, 2019; Gur et al., 2003; Krupinski, 1996; Kundel et al., 1978; Taylor, 2007) and airport security (Gale et al., 2000; Godwin et al., 2010; Mendes et al., 2015; Muhl-Richardson et al., 2020; Wolfe et al., 2013). Indeed, choices about which real-world visual tasks to study have been driven both by the importance of the tasks and their similarity to standard laboratory experimental designs. As a result, there is much more work on screening mammography, where the radiologist does more or less the same task over and over, than on emergency room radiology, where the task changes from case to case (e.g., Berge et al., 2020). Certainly, there is nothing wrong with studying problems like mammography, but one would like to be able to assert that the rules that have been established in the lab apply more generally to important tasks that do not have this blocked structure in the world. Driving is a good example. It involves a continual mix of search tasks: searching for landmarks, finding the right icon in the in-car display, monitoring for danger. These tasks repeat. The driver must search for road signs over and over, but those searches are intermixed with other search tasks.

There are reasons to think that behavior might differ between blocked and mixed searches. In recent years, there has been increasing emphasis on the role of "history" in the control of selective attention (Anderson et al., 2021; Awh et al., 2012; Becker et al., 2023; Wolfe & Horowitz, 2017; but see Ramgir & Lamy, 2021). "History" is a term covering a variety of phenomena. To quote the very useful review by Anderson et al. (2021), selection history is "prior experience, broadly construed, that exerts a direct influence on the control of attention." This includes the priming of the current trial by the previous trial (or a few previous trials; Huang et al., 2004; Kristjansson & Driver, 2008; Kruijne & Meeter, 2015; Maljkovic & Nakayama, 1994), as well as the consistency of target identity. Are you always searching for the same thing or does the target change from trial to trial (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977)? Clearly, a history of searching for the same thing over hundreds of trials is different than a history of mixing different search tasks, so one might expect different search results if search tasks are mixed rather than blocked. Note that "the same target" might be exactly the same, like exactly the same red dot or exactly this image of a cat. Alternatively, the task might stay the same while the exact target could vary, such as when observers search for an animal or any vowel in a set of letters. In either case, we can ask if blocked trials produce different results from a mixture of several search tasks.

For some search tasks, performance seems unlikely to change in any qualitative manner, if the tasks are blocked or mixed. For instance, if observers are looking for a red item among homogeneous distractors, performance is likely to be highly efficient for target-present and target-absent trials. For an inefficient task, where each well-known item probably still needs to be identified in series (e.g., a search for a *T* among *L*s), it seems probable that the process of finding the target might be unaffected by whether the tasks are blocked or mixed. The absent trials might be a different story, as discussed below. There are a range of intermediate cases where blocking trials might be more important. Consider the case of "weak" preattentive features. There is a limited set of basic features that are capable of guiding attention (Wolfe & Horowitz, 2017). These features are not equal to one another. Some, like color and size, guide attention vigorously, and salient examples will "pop-out" of a search display of homogeneous distractors. Others, including such features as lighting direction (Sun & Perona, 1998) or Vernier offset (Fahle, 1991b), do appear to guide attention, but the effects are weaker. One could imagine that a clear demonstration of the efficacy of such features could require that observers be focused on that feature for a block of trials.

A related issue concerns search termination. If an observer is searching for the presence of only one target in a display, it is easy enough to decide when to end the search on positive trials. If the target is found, the search is over. The same is not true for absent trials or for searches with an unknown number of targets. An unsuccessful search must end, but it is not clear how this is done. The most obvious thought is that the search ends when one has looked at "everything" and determined that none of those things are the target. However, even for searches where participants probably do examine item after item in series (e.g., a *T* among *L*s), it can be shown that they are not doing a serial exhaustive search on absent trials (Horowitz, 2006; Horowitz & Wolfe, 1998). If the search is for an object in a real-world scene, the definition of "everything" becomes very unclear (Neider & Zelinsky, 2008; Wolfe et al., 2011a, 2011b). Obviously, participants will not scrutinize every leaf on a tree when looking for a coffee mug at a picnic in the park (Biederman et al., 1973; Hwang et al., 2011; Wolfe et al., 2011a, 2011b) and, just as obviously, it will be possible to end that search, even if the coffee mug is absent.

There have been multiple efforts to model the decision process for absent trials. One common approach is to propose some version of a "quitting threshold" that is adaptively set over the course of multiple trials (Chun & Wolfe, 1996; Moran et al., 2013; Schwarz & Miller, 2016; Wolfe, 2012; Zenger & Fahle, 1997). The general idea is that a current absent response makes the next absent response a bit faster, and a miss error makes the next response slower. The relative size of these changes creates a staircase process (Rose et al., 1970) that converges on a quitting threshold that produces some specific level of errors. Some models have quitting threshold for the decisions made about each attended item as well as for the trial as a whole (Peltier & Becker, 2016; Wolfe et al., 2022a, 2022b). Such models generally assume feedback, which is not always available in the real world. At best, real-world feedback tends to be partial. A searcher almost always knows if they found something, so the target that they found usually constitutes accurate positive feedback, though this would be less true for ambiguous targets (Is that really a tumor in that x-ray?). Becker et al. (2022) showed that quitting times could be adjusted by changing the time required to find target without providing feedback on absent trial. Feedback from a negative, target-absent response is less clear-cut, however. It is easier to be sure that you found something than to be sure that there was nothing to be found.

Characteristically, these models and the experiments testing quitting problems involve blocks of similar trials. That raises the question of how people quit unsuccessful searches when they encounter the task outside of a block of trials. If someone was asked to search for a cow in the parking lot, it is obvious that they could do so, and, if the cow was not present, this search could be terminated successfully even if the observer had not previously performed a block of repeated cow-in-parking-lot searches. Presumably, the observer would have a set of priors about such targets and such scenes and could apply these priors in order to set a quitting threshold (see, for example, Yang et al., 2022). Mazor and Fleming (2022) looked into this for some simple tasks. They show, for example, that in a search for a red dot among blue, observers need no training to quit a target-absent trial consisting of all blue items. They also show that quitting times speed up with experience for more difficult searches. This raises the possibility that searchers might start with some quitting prior that is then modified by an adaptive process like those cited above.

It is difficult to study quitting for novel cow-in-parking-lot searches because the search is only novel once. Mazor and Fleming (2022) used the first four trials in a relatively large sample of observers to assess search termination without experience. In the work described here, we will look for differences between situations where four search tasks are randomly intermixed. We will compare this mixed condition to a blocked condition where each of the four tasks is run in a block of only that type of search. We will report on four versions of this mixed versus blocked design. In Experiment 1, four different feature searches are tested. This is akin to the driving example given earlier, where several tasks recur, but not in blocks of many repeated trials. In Experiments 2 and 3, Participants search for the same target in different situations, something like searching for the same cat; first, in the bathroom (easy), then in the garden (harder), and so forth. In Experiment 4, different targets are searched for amongst a fixed set of distractors. This can be seen as analogous to searching for multiple distinct items, one after the other in the same kitchen or in the same abdominal CT images in radiology. Finally, in Experiment 5, we compare situations where the observer chooses the task for the next trial versus when the task is predetermined. While these experiments do not exhaust the questions about possible differences between mixed and blocked searches, the results will show no qualitative change in search behavior as a function of this mixed/blocked manipulation. The results reassure us about the generalizability of the rules of search learned from blocked trials, though these results can be seen to challenge simple adaptive models of search termination.

Experiment 1: Four feature search tasks

Experiment 1 was preregistered on the Open Science Framework (OSF; https://osf.io/26jsr). In Experiment 1, participants searched for a target defined by a single feature. The four feature searches are shown in Fig. 1.

When a search target is defined by a single, preattentively processed feature like color, attention is drawn to the target efficiently and the number of distractors has little or no influence on response time (Buetti et al., 2016; Treisman & Gelade, 1980). Task A from Experiment 1 is a search of this sort (Fig. 1A). Participants search for a red target amongst homogenous green distractors. This will reliably produce fast response times (RTs) regardless of the set size with RT × Set Size slopes near zero. There are several features that produce this kind of unambiguous evidence of "pop-out" search behavior, such as sufficiently salient motion, size, or orientation targets (Wolfe & Horowitz, 2004).

Many other features appear to guide attention, but the evidence is weaker. Tasks B, C, and D of Experiment 1 test such features. In Task B, the target is a cube with a

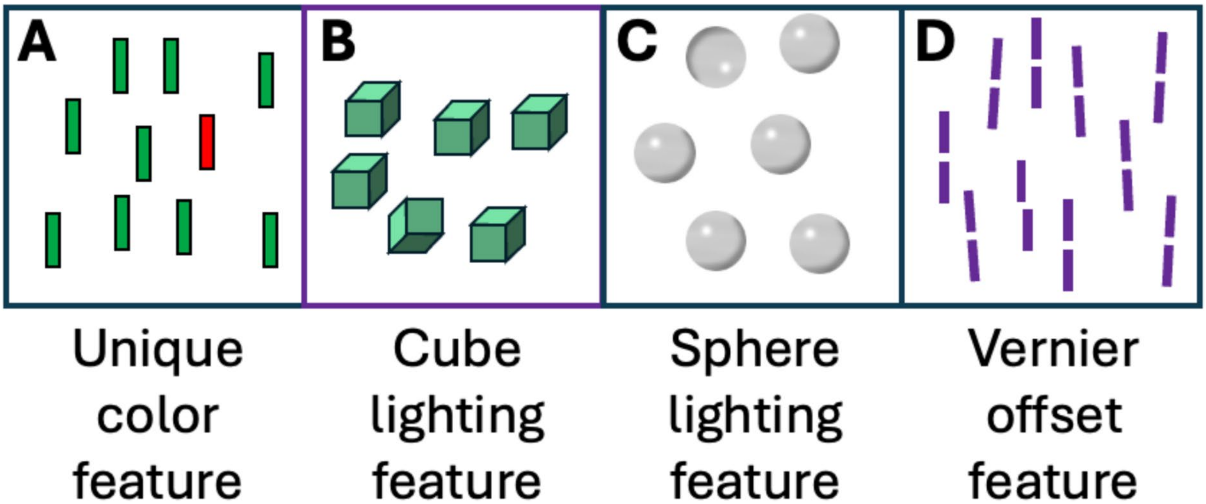


Fig. 1 The four search tasks of Experiment 1. See text for details. Items are not drawn to scale. (Color figure online)

different orientation and lighting direction from that of the distractors (Enns & Rensink, 1990). In Task C, the target is the sphere that appears to be lit from below (Sun & Perona, 1998). Finally, in Task D, the target is the pair of lines that show a Vernier offset among distractors where the lines in the pair are colinear (Fahle, 1991a). Notice that the distractors vary in overall orientation so as to disrupt an orientation cue that might otherwise be used to detect the Vernier target (Findlay, 1973). Each of these features can produce efficient search but, as preattentive features, they are more "fragile" than a feature like color. Accordingly, if mixing different tasks did make a qualitative difference in search, these four seem like good candidate tasks to try. Of course, a vast array of other tasks might have been chosen. Nevertheless, in choosing only four tasks, we are balancing the desire to mix different search tasks with the need to have enough data on each task to obtain reliable results.

Participants

Standard search experiments looking for effects of search efficiency/search slopes traditionally involve 10–12 observers. Since we were not sure about effect sizes and were using fewer trials per condition than we might normally use, we opted for a larger number of observers, testing 39 in total. Fourteen participants self-identified as male, 25 as female. Average age was 29.8 years($SD =$ 8.8, range: 20–53). Participants were recruited online through Prolific (https://www.prolific.com). All attested to normal color vision and no oculo-motor disorders. They gave informed consent as approved by the Mass General Brigham IRB Protocol #2007P000646 and were paid at a rate of $12/hour.

Methods

The experiment was performed online, thus, stimulus size was not precisely defined in degrees of visual angle. Participants were not allowed to perform the experiment on a phone or tablet. Stimuli were placed in random locations in a jiggled 5×5 grid. The grid was placed within square field that was 0.5 of the maximum height of the screen. Each stimulus item fit into an invisible box that was 0.11 of the screen height. If the viewing distance was about 60 cm, a standard computer screen would show a stimulus field of about 15° on a side and each item would be place in a box of about $3 \times 3°$. The background was white and the colors were as shown in Fig. 1, though precise statements of color cannot be made for online testing.

Participants were tested both in the "blocked" and "mixed" conditions in order. Within the blocked condition, participants ran a block of 100 trials for each task. The task order was randomized across observers. After a break, they ran the mixed condition with a further 400 trials with all tasks randomly intermixed. The blocked condition was always run first and mixed condition second because we did not want inferior performance on the mixed condition to be attributable to a learning effect. In retrospect, this was probably not the ideal choice (see Results). Trials were divided evenly between three set sizes: 8, 16, and 24 items and evenly between target-present and target-absent trials. Stimuli were presented until the observer responded by pressing either "p" or "q" to indicate target present or absent, respectively. Accuracy feedback was provided after each trial.

Results

We removed RT outliers, here defined as RTs less than 200 ms and greater than 10 s. Filtering RTs is a somewhat subjective art. We seek to remove RTs in the upper tail of the RT distribution that might result from motor errors or interruptions. In none of the experiments reported here is the qualitative shape of the data markedly changed if all RTs are included. The filters that we used in Experiment 1 removed 1.3% of the data. However, one participant lost 41% of their trials to these filters and accounted for 80%(!) of all removed RTs. That observer was excluded from analysis, leaving 99.7% of RTs for the remaining participants. Participants were also removed if d' was lower than 0.5 for any task. This is a change from the preregistration that proposed excluding participants with more that 20% errors in any one condition. We had not anticipated that the Vernier condition would be as difficult as it was. The $d' < 0.5$ criterion allowed us to be more flexible in keeping observers while removing those with behavior that suggested that they were not really doing the task as asked. Two subjects were removed by this accuracy criterion, leaving 36 participants in the final analysis.

Figure 2 shows RT as a function of set size for each of the four tasks. Note the very different *y*-axes. The tasks differ considerably in difficulty. The most obvious effect is that the mixed condition is actually faster than the blocked condition for the three harder tasks. This is almost undoubtedly a learning effect, resulting from our decision to do the blocked condition first. Certainly, it provides no support to the idea that mixing tasks would disrupt search.

Supplementary Table S1 shows the results of three-way analyses of variance (ANOVAs) with factors Set Size × Mixed vs. Blocked × Hit vs. True Negative. For the color feature condition, only the main effect of set size is significant. For the three harder searches, all ANOVA terms except the triple interaction are significant. The critical analyses for present purposes are the main effect of Mixed vs. Block and the interaction of Mixed vs. Blocked with Set Size. The interaction serves as a test for a change in slope. As noted, these are significant, but the direction shows the mixed condition to be faster and more efficient than the blocked condition. This is evidence *against* the hypothesis that mixing different tasks together would disrupt search, making it slower and/or less efficient.

It could be hypothesized that Fig. 2 shows one side of a speed–accuracy trade-off. Perhaps, mixing tasks causes participants to become fast and sloppy. That would predict that the error rates would be greater in the mixed conditions. However, Fig. 3 argues that this is not the case. The blocked conditions tended to have the higher error rates. Again, it seems likely that this is attributable to a learning effect, resulting from the blocked condition having been run before the mixed condition. For statistical analysis, errors were arcsin-transformed. This is done to make the error data more normally distributed for statistical purposes (Hogg & Craig, 1995). In the present case, the results of analysis are qualitatively the same if the untransformed data are used. Table S2 shows the results of a three-way ANOVA with the factors Set Size × Mixed vs. Blocked × Miss vs. False Alarm (FA). The Mixed vs. Block factor produced a significant effect only in interaction with Miss vs FA in the Vernier condition. In general, the error data are unremarkable and do not show any evidence for an increase in errors in the mixed condition.

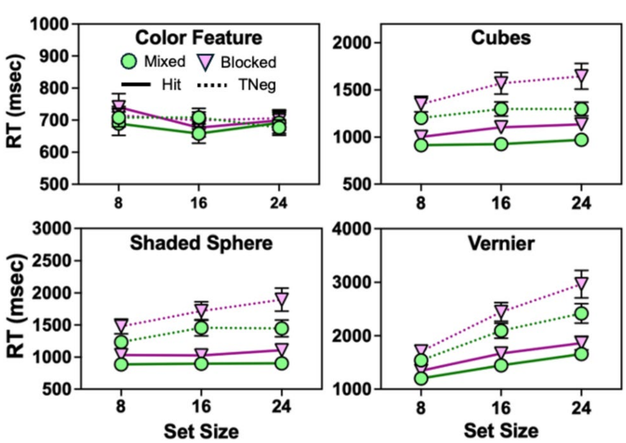


Fig. 2 RT × Set Size functions for the four tasks of Experiment 1. Error bars, where visible, are ±1 s.e.m. Note the different *y*-axes. Solid lines show hit (correct present) trials. Dashed lines show true negative (TNeg), target-absent trials. Green circles show the mixed condition. Purple triangles show the blocked condition. (Color figure online)

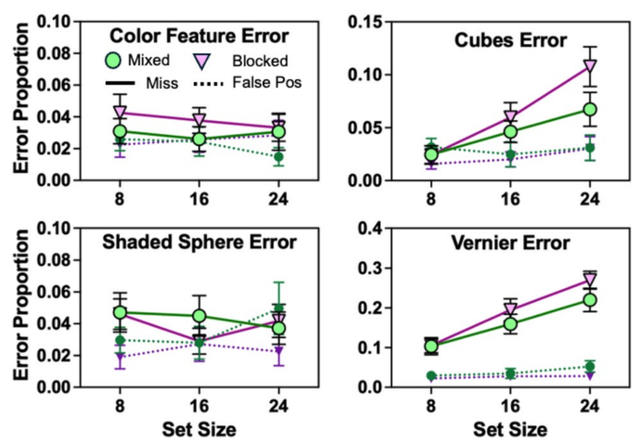
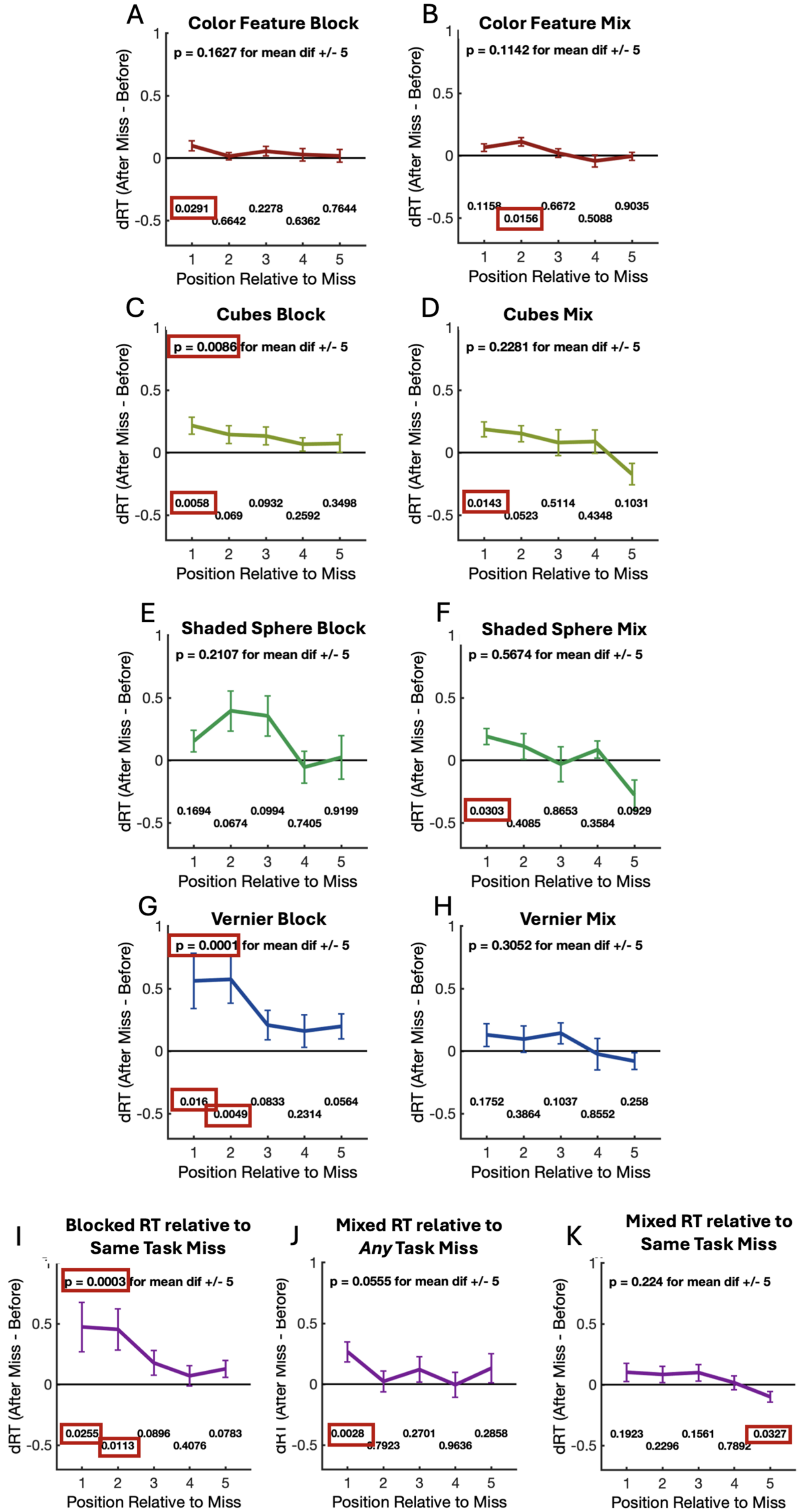


Fig. 3 Error rate as a function of set size for the four tasks of Experiment 1. Error bars, where visible, are ±1 s.e.m. Note the different *y*-axes. Solid lines show miss errors trials. Dashed lines show false-positive/false-alarm errors. Green circles show the mixed condition. Purple triangles show the blocked condition. (Color figure online)


A
Color Feature Block
p = 0.1627 for mean dif +/- 5
0.0291
0.6642
0.2278
0.6362
0.7644
B
Color Feature Mix
p = 0.1142 for mean dif +/- 5
0.1158
0.0156
0.6672
0.5088
0.9035
C
Cubes Block
p = 0.0086 for mean dif +/- 5
0.0058
0.069
0.0932
0.2592
0.3498
D
Cubes Mix
p = 0.2281 for mean dif +/- 5
0.0143
0.0523
0.5114
0.4348
0.1031
E
Shaded Sphere Block
p = 0.2107 for mean dif +/- 5
0.1694
0.0674
0.0994
0.7405
0.9199
F
Shaded Sphere Mix
p = 0.5674 for mean dif +/- 5
0.0303
0.4085
0.8653
0.3584
0.0929
G
Vernier Block
p = 0.0001 for mean dif +/- 5
0.016
0.0049
0.0833
0.2314
0.0564
H
Vernier Mix
p = 0.3052 for mean dif +/- 5
0.1752
0.3864
0.1037
0.8552
0.258
I
Blocked RT relative to Same Task Miss
p = 0.0003 for mean dif +/- 5
0.0255
0.0113
0.0896
0.4076
0.0783
J
Mixed RT relative to *Any* Task Miss
p = 0.0555 for mean dif +/- 5
0.0028
0.7923
0.2701
0.9636
0.2858
K
Mixed RT relative to Same Task Miss
p = 0.224 for mean dif +/- 5
0.1923
0.2296
0.1561
0.7892
0.0327
dRT (After Miss - Before)
Position Relative to Miss
1
0.5
0
-0.5
1 2 3 4 5

◂**Fig. 4** Difference in RTs before and after a miss error for each task in mixed or blocked conditions (**A–H**). Panel **I** shows blocked data pooled over Tasks. Panel **J** shows differences in RT in the mixed condition when the TNeg trials are chosen independent of task. Panel **K** shows differences in RT when miss and TNeg trials come from the same task. Error bars are ±1 s.e.m; *y*-axis units are seconds; *t*-test results with $p < 0.05$ a marked with red boxes. (Color figure online)

How do participants know when to end an absent trial? As discussed in the Introduction, a number of modelers have proposed that a quitting time is set adaptively by having the observers monitor their accuracy, increasing RT after an error and decreasing it after a correct response. The mixed condition could produce difficulties for such a process. If one misses a Vernier target, should the quitting threshold be raised for all tasks or just for the Vernier task? The latter seems like the more adaptive strategy. To gain some insight into this question, we computed average target-absent RTs relative to miss errors. We then compared these RTs for equivalent positions before and after the miss. The results of this analysis are shown in Fig. 4. Datapoints show *RT(after miss) – RT(before miss)* for each of five positions relative to the miss error, for each participant in each task. The prediction would be that, if the quitting time RTs are adaptive, this difference should be positive—at least for the first couple of positions. Numbers on the graphs at each position represent the *p* values for one-sample *t* tests, evaluating the hypothesis that that the differences between RT after a miss and the equivalent RT before the miss are above zero on average (not corrected for multiple comparison). The *p* value at the top of each graph is the *t* test for the average difference over all five positions.

For the mixed condition, the analysis can be done in two ways. Consider a miss in the Vernier task. We could look for the first true negative Vernier task trials before and after the miss. Those trials will be quite distant from the miss (~ 8 trials for Position 1) compared with the distances for Position 1 in the blocked condition (~ 2 trials). Results of that analysis are shown in Fig. 4B, D, F, & H, with pooled results shown in Fig. 4K. Blocked results are in Fig. 4A, C, E, F, G and pooled in Fig. 4I. Alternatively, following that Vernier miss in the mixed condition, one could look for the first TNeg before and after the miss, regardless if the task that produced those TNegs. That preserves the distance from the miss to the TNegs but not the match of miss and TNeg tasks. Those pooled results are shown in Fig. 4J.

The data show a somewhat noisy effect in the expected direction for blocked conditions. The result is particularly clear in in the pooled results (Fig. 4I). The tasks show either statistically significant or, at least, numerical slowing of the first true negative trial after a miss. Two of the tasks show an average slowing over five true negative trials after a miss. In the mixed condition, the effect is clearly weaker. If the mixed condition's analysis is done only for trials where the TNeg is of the same task type as the miss, the effects are small and insignificant (except for an odd result in the wrong direction at Position 5). If match between miss and TNeg tasks is not required (Fig. 4J), there is evidence for slowing of the first TNeg after the miss.

There are many aspects of this analysis that are not ideal. Different observers have different error rates. Each observer gets the same contribution to the data shown in Fig. 4. Different tasks produce very different error rates, making the easier tasks less well-powered. In the pooled data, the harder tasks make a larger contribution to the result. Still, the overall conclusion would be that the evidence for an adaptively set quitting rule is reasonably clear for the blocked. It is weaker in the mixed condition, even though there is no evidence for a qualitative difference in target-absent behavior between the mixed and blocked conditions (i.e., no evidence for markedly higher TNeg RTs or error rates in the mixed condition). It could be that the bulk of any adaptive process took place during the blocked conditions and was retained for each task in the subsequent mixed condition since blocked was done first. This hypothesis will be tested in later experiments in this paper.

Discussion

The main conclusion from Experiment 1 is that mixing four tasks of different difficulty together did not dramatically change search behavior, compared with the situation where each task was run in a block of similar trials. If anything, search behavior was better—somewhat faster and somewhat more accurate—in the mixed condition than in the blocked condition. However, this could be a by-product of the choice to run the blocked condition first. We deal with this possibility in Experiments 2 and 2a. It is particularly interesting that performance on the absent trials did not suffer in the mixed condition. This would be consistent with the idea that participants can maintain separate quitting rules for each task and can switch to the right one at little or no cost on each trial. Determining just how well participants can do this would require different experiments, designed to get around the limitations of the present design. Alternatively, it could be that the quitting rule is computed anew on each trial, based on the participant's rapid assessment of the stimulus.

There are an essentially infinite variety of mixed conditions that could be tested so it would not be wise to assert that mixing search tasks never disrupts search. In the following experiments, we examine the effect of mixing tasks with several different sets of tasks in an effort to assess the generalizability of the results of Experiment 1.

Experiment 2: One target; multiple backgrounds

Experiment 2 is preregistered on OSF (https://osf.io/cmjrb). In Experiment 2, the target was a green *O* in all tasks. As shown in Fig. 5, the four tasks differed in their distractors. This is akin to looking for the same cat in different settings or to looking for the same pathology in different patients. In this case, we deliberately varied the tasks across a wide range of difficulty. In the "easy color" task, the target was uniquely green among blue distractors. In the conjunction task, Participants searched for a green *O* among *O*s that were not green and green items that were not *O*s (Egeth et al., 1984; Wolfe et al., 1989). In the hard color task, participants looked for the green *O* among yellow-green and blue-green distractors. This task is difficult because the target-distractor distance in color space is reduced (Duncan & Humphreys, 1989; Nagy & Sanchez, 1990) and, more importantly, because the distractor colors flank the target in color space. The target is said not to be "linearly separable" from the distractors (Bauer et al., 1996a, 1996b; but see Vighneshvel & Arun, 2013). Finally, in the hard shape condition, participants look for the green *O* among vertical and horizontal green ovals of slightly different sizes. This is hard because the distractors are heterogeneous (Xu et al., 2021) and because the target is defined by the absence of an attribute (orientation) and search for absence is harder than search for presence (Treisman & Souther, 1985).

We tested 47 participants for this experiment, expecting to lose a fair number to the vagaries of online testing. Eighteen participants self-identified as male, 29 as female. Average age was 28.4 years ($SD = 8$, range: 18–51). Participants were recruited online through Prolific. All attested to normal color vision and no oculo-motor disorders. They gave informed consent as approved by the Mass General Brigham IRB protocol #2007P000646 and were paid at a rate of $12/hour.

As in Experiment 1, participants were tested on the blocked condition before the mixed condition. As will be discussed below, having recognized that this was a problem, we ran a separate version of the experiment in which participants were tested only in the mixed condition. The results for this control are very similar to the mixed condition of the current experiment. Participants were tested for 100 trials of each of the four tasks in mixed and blocked conditions. Set sizes were 8, 16, and 24. Targets were present on 50% of trials. In all other respects, the experiment followed the methods of Experiment 1.

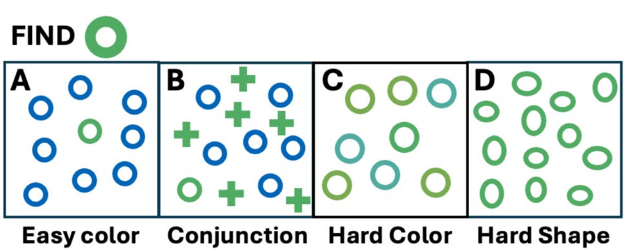


Fig. 5 The four tasks in Experiment 2. (Color figure online)

Results

Data and subject exclusion

As before, we removed RTs < 200 ms and greater than 10 s. We then removed participants who had more than 20% of their RTs removed in any one task by those criteria or who produced $d' < 0.5$ in any task. A lamentably large number of participants needed to be removed. In most cases, they seem to have decided not to do one or more of the tasks. For example, they might simply push the response key as fast as possible whenever ovals appeared. After exclusions, 23 participants remained. This is lower than our preregistered goal. We ran another version (Experiment 3) as a partial corrective for this deviation. It is worth noting that if all participants are included (while still removing the "bad" RTs), the overall pattern of the RT data does not change in any qualitative manner, though RTs increase in both mean and variance.

Figure 6 shows the RT × Set Size functions for Experiment 2. Note that there is a very wide range of difficulty in these search tasks from target-present slopes of ~0 ms/item for the color feature search to slopes of about 65 ms/item for circles among ovals. Mixing these very different tasks

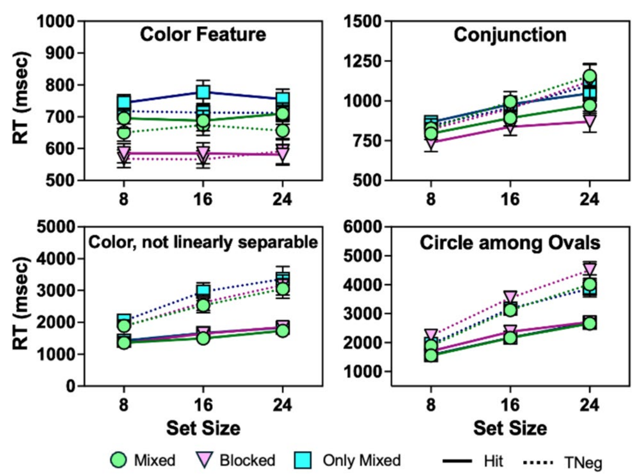


Fig. 6 RT × Set Size functions for Experiment 2 and for the mixed-only control version of the experiment. Note the very different *y*-axes. Error bars, where visible, are ±1 s.e.m. Circles = mixed condition, triangles = blocked, squares = mixed-only control, solid lines = correct target-present (hit) trials, dashed line = correct target-absent (true negative) trials. (Color figure online)

in a random sequence does not make much difference to the results. There are no significant differences in slope of RT $\times$ Set Size functions between blocked and mixed conditions, all $t(22) < 1.5$, all p values > 0.17. A three-way ANOVA with factors set size, hit versus TNeg, and mix versus block shows a significant main effect of the mix versus block factor for the simple color search, $F(0.9, 20.7) = 31.92$, $p < 0.0001$. It can be seen that overall RTs are faster in the blocked condition. There is a more modest main effect of the mix versus block factor for the *O* among ovals task, $F(0.5, 10.0) = 6.791$, $p = 0.0423$, but note that Fig. 6 shows, again, that the mixed condition was faster. The full ANOVA table is found in Table S3.

Like slopes, error rates vary widely in this task with average miss error rates of about 3% in the color feature task compared to error rates of over 30% in the circle among ovals task (and that is after removing participants who did worse). A three-way ANOVA on arcsin transformed errors has only one significant result involving the Mixed vs. Blocked manipulation. For the circle among ovals task, there is an interaction of the Mixed vs. Block and Miss vs. FA factors, $F(0.9, 20.1) = 14.56$, $p = 0.0013$. Miss errors are somewhat higher in the mixed condition. A two-way ANOVA (Block/Mix, Set Size) shows that false-alarm errors are significantly higher in the blocked condition, $F(1, 22) = 12.57$, $p = 0.0018$. Beyond that, mixing tasks appears to have no particular effect on errors. The full ANOVA table for the error analysis is found in Table S4.

Figure 7 shows the results of a comparison of RTs before and after miss errors. The figure shows the results pooled over tasks with the mixed condition analyzed in the two ways described for Experiment 1 and shown in Fig. 4J–K. In Experiment 2, there is very little evidence for the adaptive process proposed to adjust quitting times on absent trials. There was evidence for an elevation in post-miss RTs for the first TNeg after a miss in the conjunction task for both

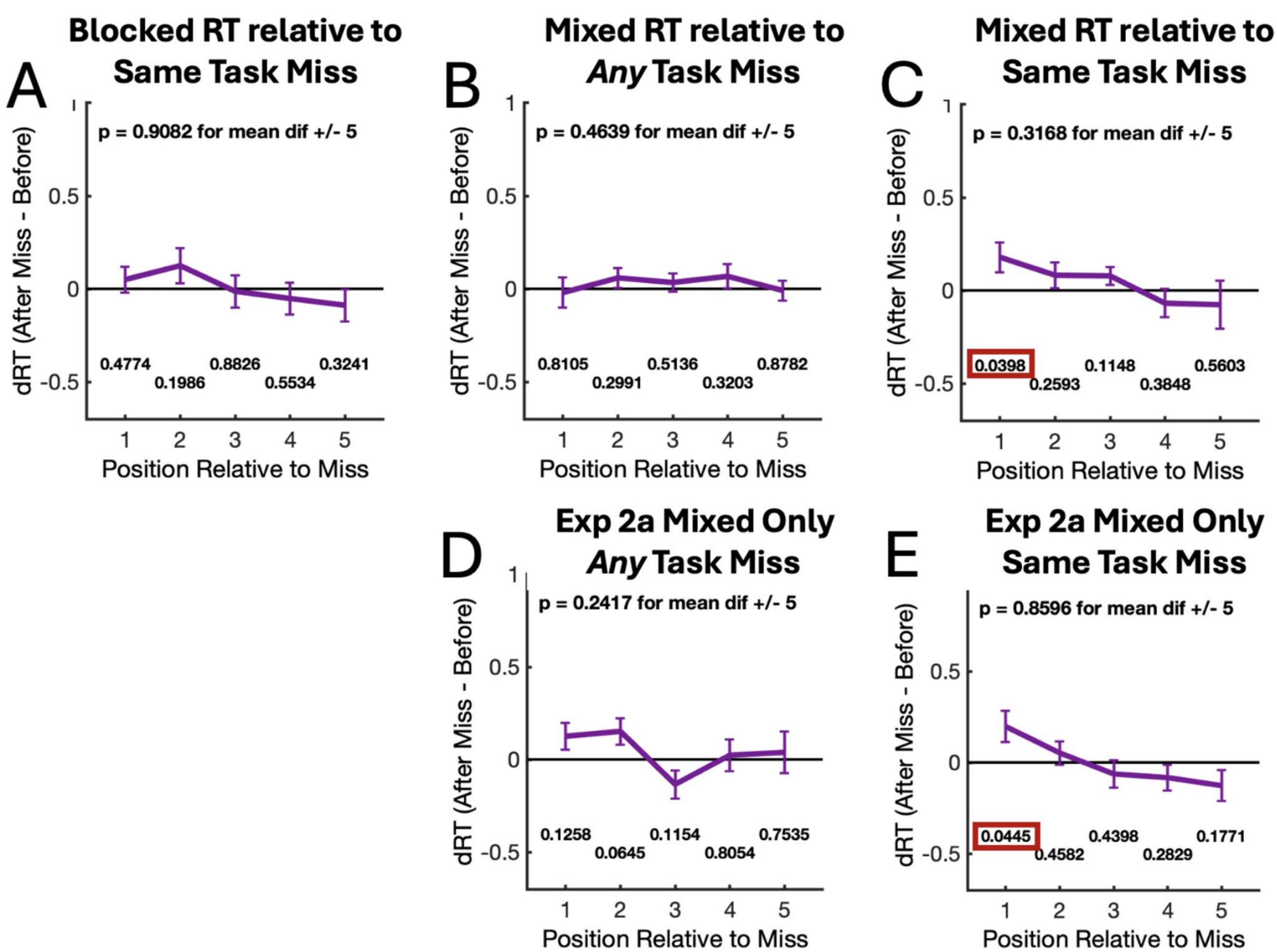


Fig. 7 Difference between RTs before and after a miss error, pooled across tasks for Experiment 2 (**A–C**) and the mixed-only control (Experiment 2a, Panels **D & E**). Mixed conditions are calculated with the requirement that the miss and TNeg tasks match (**C & E**) and without that requirement (**B & D**). Red boxes highlight conditions where the mean value deviates significantly from zero. See text for details. Error bars are ± 1 s.e.m. (Color figure online)

mixed and blocked conditions and for the first position in the blocked condition of the color feature task (all p values < 0.05). In the pooled data, the only significant result was a weak effect for the first TNeg after a miss in mixed condition when miss and TNeg were from the same task ($p = 0.04$). No other effects were significant.

Experiment 2a: A mixed-only control

Since running the blocked condition first introduces a variety of possibly confounding factors into the comparison of mixed and blocked conditions, we ran a control experiment where participants only ran the mixed condition. Of course, these participants must be different from the participants in the main Experiment 2. However, it is still possible to compare the mixed condition of Experiment 2 to this mixed only condition. All conditions were identical to the mixed condition of Experiment 2. We tested 40 online participants for this control experiment (21 women, $M_{age} = 25.6$ years, $SD = 5.5$, range: 18–47).

The rules for exclusion of trials and observers were the same as in the main portion of Experiment 2. Again, many observers were removed from analysis either because they had too many excluded RTs or because d' was below 0.5 in one or more tasks (generally in the two harder tasks). After exclusion, 21 participants remained. The blue squares in Fig. 6 show the average RT data for those 21 participants. For the simple color feature task, the mixed-only control condition is slower than the mixed condition of the main experiment, $F(1, 126) = 6.997$, $p = 0.0092$. No other comparisons of the mixed and mixed-only conditions were significant. Indeed, it is obvious from the figure that the RTs for the blocked and both mixed conditions are very similar.

Figure 7 shows the results of the analysis of RTs before and after Miss errors for both Experiment 2 and for this mixed-only control. For the mixed-only control, there is modest evidence that the first TNeg RT after a miss is slowed. This was true for each of the four tasks, but only the conjunction task showed a significant effect ($p = 0.015$). In the pooled data, the effect is significant for Position 1 if miss and TNeg tasks are constrained to be the same ($p = 0.044$). Recall that the relevant TNeg trials in Position 1 are about 10 trials removed from the miss. The Position 5 TNegs average almost 50 trials away from the miss. These results provide some modest evidence that there are separate quitting rules for each task (more effect in Fig. 7C & E than B & D). More compelling evidence would require a more extensive study to overcome noise and power issues in this design.

As an added, exploratory test for the effect of one trial on the next, we measured the RTs for all trials where the task repeated and compared those RTs to the trials where the task switched. When we restricted analysis to cases of two adjacent target-present trials, we found no difference between repeat and switch trials, $t(20) = 1.14$, $p = 0.27$. The result was similar if repeats and switches were assessed regardless of with the trials were present or absent, $t(20) = 0.85$, $p = 0.41$. This does not mean that repetition priming is not a real phenomenon; merely that we failed to find it in this case.

Discussion

From these results, the conclusion appears to be that mixing four different searches for the same target has very little impact on performance compared with testing each task in a block of its own. The only effects of mixing are increases in RT for the simple color feature task. The elevated false-alarm rates in the blocked condition suggests that this could be some form of speed–accuracy trade-off. In any case, there was no qualitative change in the search. Color feature search remained efficient in all conditions. The results of the mixed-only control experiment suggest that the decision to run the blocked conditions first in Experiments 1 and 2 did not make any systematic difference to these results. The analysis of the RTs relative to miss errors provides some evidence in favor of an ability to adjust multiple quitting thresholds in a mixed condition. The failure to see this effect for the color feature task may arise from the small number of errors for that task.

Even for online testing, the need to remove half of the observers from analysis is unfortunate. Accordingly, Experiment 3 is a close replication of Experiment 2, but with all participants run in the lab.

Experiment 3: One target; multiple backgrounds, replication

Experiment 3 is preregistered on OSF (https://osf.io/m9z57). Experiment 3 was a replication of Experiment 2, with two notable changes. First, the hard color search task using targets that were not linearly separable from the distractors was replaced with a less challenging search for an *O* target among *C* distractors (see Fig. 8). Searching for an *O* among *C*s is a relatively inefficient search. Search for a *C* among *O*s would be much easier—a classic search asymmetry (see Experiment 4 of Treisman & Souther, 1985). The second change was to run the experiment in the lab rather than online. The experiment was otherwise identical to Experiment 2. We tested 24 participants for this control experiment. Ten participants self-identified as male, 14 as female. Average age was 22.2 years ($SD = 7.55$, range: 18–50). Of these, four participants were recruited through the Mass General Brigham Rally system and paid at a rate of $15/hour. Twenty participants were recruited through Boston University's SONA Psych 101 pool. They received undergraduate class credit for their participation. All participants had 20/25 or better vision with correction, passed the Ishihara Color Blindness Test, and had no history of any eye or muscular disease. They all provided informed consent as approved by the Mass General Brigham IRB protocol #2007P000646.

As in the other experiments, we removed RT outliers, here defined as RTs less than 200 ms and greater than 10 s. These filters removed only 0.75% of the data. We again planned to exclude any observer with greater than 20% of trials removed in any task and any observer whose d' was lower than 0.5 for any task. In this case, none of the 24 participants were removed from analysis, addressing the issue of high rejection rates in Experiment 2.

The RT × Set Size functions, shown in Fig. 9, are similar to those in the previous experiments. However, in this case, there is an interesting effect of the mixed/block manipulation. In Fig. 9A, it is clear that the mixed condition is slower than the blocked condition for the simple color feature search. This is borne out by a three-way ANOVA (Set Size, Mixed vs. Blocked, and Hits vs. TNegs) that shows a significant main effect of the mixed/block factor, $F(0.43, 9.89) = 7.020$, $p = 0.0413$. Full results of the ANOVAs are found in Table S5. Figure 9B shows the same difference between mixed and blocked for the conjunction task, $F(0.61, 14.13)$

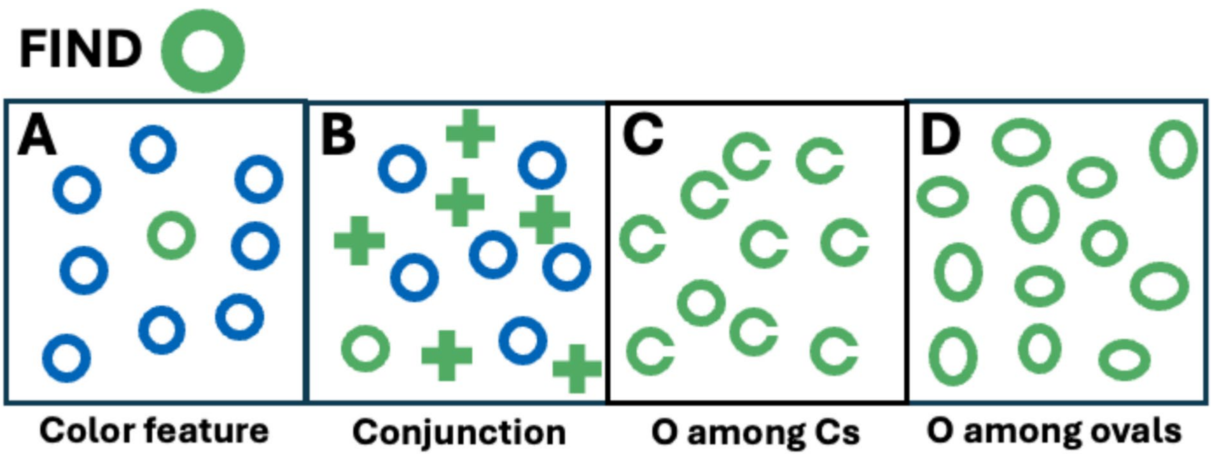


Fig. 8 **S**timulus conditions for Experiment 3. (Color figure online)

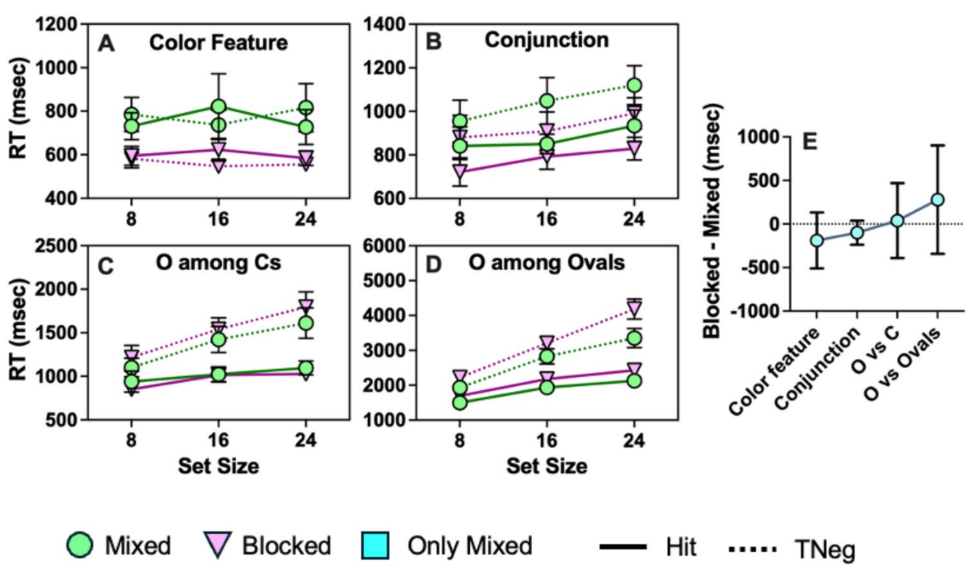


Fig. 9 **A-D** RT × Set Size functions for Experiment 3. Note the very different y-axes. Error bars, where visible, are ± 1 s.e.m. Circles = mixed condition, triangles = blocked condition, solid lines = correct target-present (hit) trials, dashed line = correct target-absent (true negative) trials. Panel **E** Average difference in RT (blocked − mixed) for all trials in each condition of Experiment 3. Error bars ± 1 s.e.m. (Color figure online)

= 12.55, $p = 0.0073$. But this effect reverses for the two harder tasks. The mixed advantage is not significant for the *O* among *C*s task, $F(0.45, 10.42) = 0.2560$, $p = 0.4404$, but it is for the *O* among ovals task, $F(0.49, 11.33) = 8.809$, $p = 0.0244$. It is as if the effect of mixing tasks made the harder tasks faster and the easier tasks slower. This is illustrated in Fig. 9E, where the RT for each task is averaged across set size and target presence/absence. The figure plots blocked – mixed average RTs. If mixed and blocked conditions did not differ, the difference would be zero. A two-way ANOVA on the average RTs with factors task and mixed/block shows a significant interaction of task and mixed/blocked factors. As the task becomes harder, the difference goes from negative to positive, $F(2.06, 47.25) = 8.627$, $p = 0.0006$. Note that a similar, although less clear-cut, version of this pattern can be seen in the data for Experiment 2. The blocked condition seems somewhat faster for the easiest task while mixed is faster (albeit not significantly) for the harder tasks.

One possible account is that this represents a speed–accuracy trade-off. If so, one might expect to see the reverse trend in the error data: Fewer errors in the mixed condition for the easier tasks and more for the harder tasks since they show faster mixed than blocked RTs. However, no such pattern is seen. In Fig. 10A, there are more miss errors in the blocked condition (main effect of mixed vs. blocked, from a three-way ANOVA on arcsin transformed errors), $F(0.9517, 21.89) = 8.28$, $p = 0.0095$. However, that reverses for the conjunction task, $F(0.94, 21.71) = 4.574$, $p = 0.0459$, and is not statistically significant for either of the two harder tasks. (Full ANOVA table is found in Table S6).

Figure 10E shows the difference in all errors between blocked and mixed conditions for each task. As can be seen, unlike Fig. 9E, there is no systematic change as a function of task difficulty.

Turning to the question of the setting of a quitting threshold, Fig. 11 shows the difference in RTs before and after a Miss error, pooled over all four tasks. In this experiment, there is reasonable evidence for elevated RTs after a miss in the blocked condition, but no significant effects in the mixed condition, regardless of how the results were pooled.

Discussion

Probably by virtue of being run in the lab rather than online, no participants in Experiment 3 needed to be excluded because of the quality of their data. This is in contrast to Experiment 2, where about 50% of participants were excluded. If we compare Figs. 6 and 9, we can conclude that this difference in data quality did not have a major impact on the results. The main conclusions remain the same; there are no qualitative differences between the mixed and blocked conditions. In particular, the efficiency of the searches is comparable in the two conditions. Experiment 3 does seem to show that participants became, in a sense, more "moderate" in the mixed condition. Relative to the blocked condition, participants became a little slower on the easier tasks

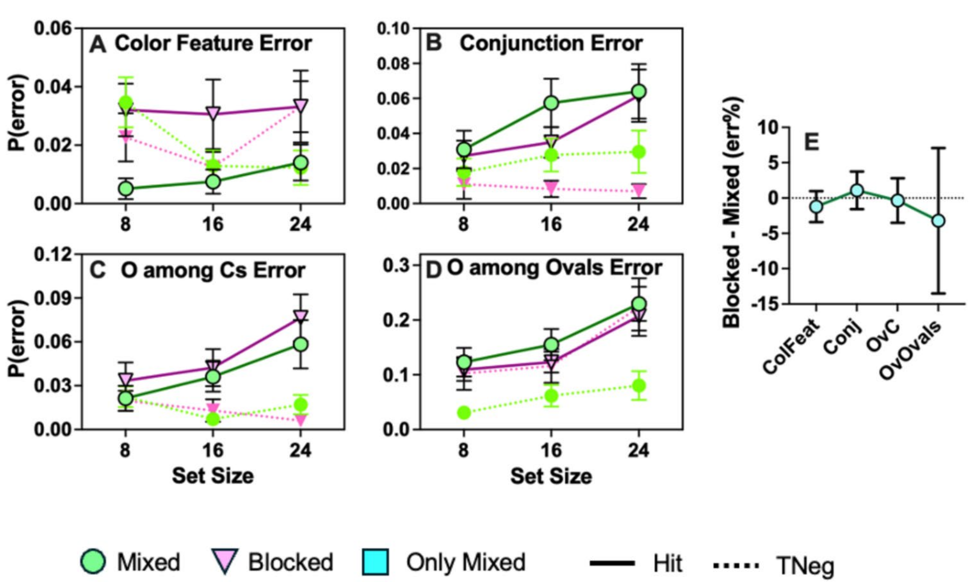


Fig. 10 **A–D** Error rate as a function of set size for the four tasks in Experiment 3. Error bars, where visible, are ±1 s.e.m. Note the different *y*-axes. Solid lines show miss errors trials. Dashed lines show false-positive/false-alarm errors. Green circles show the mixed condition. Purple triangles show the blocked condition. **E** Difference in total error percentage for each task. (Color figure online)

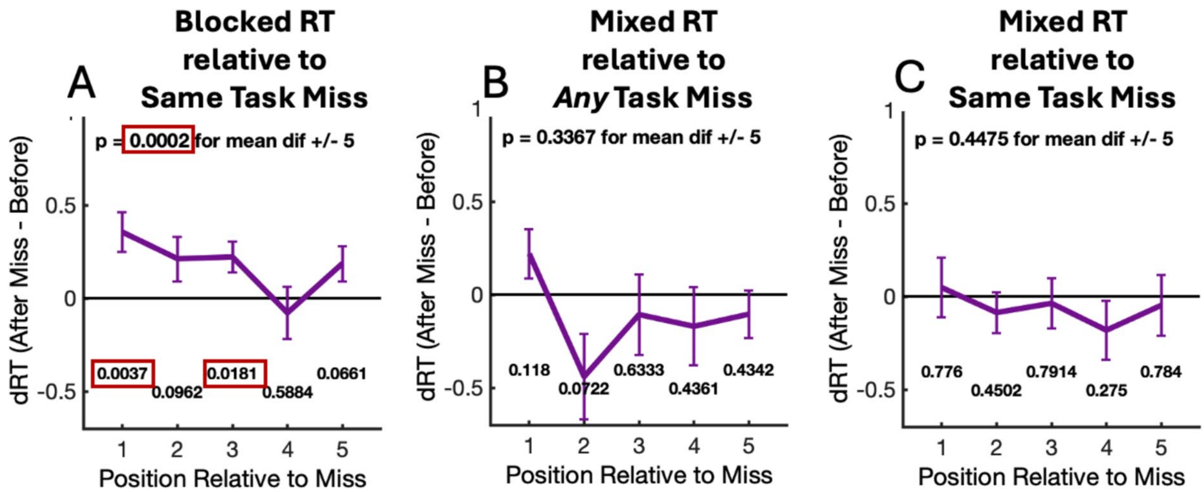


Fig. 11 Difference in RT for matched positions before and after a miss error for blocked and mixed condition. Results are pooled over all four tasks. Panel **A** shows the blocked condition. Panels **B** shows the results for all misses regardless of whether the miss and TNeg trials involve the same task. Panel **C** shows the results where the TNeg trial task matches the miss trial task. (Color figure online)

and a little faster on the harder tasks but without any significant impact on the error rates.

Experiment 3 is analogous to the search for an object that can be found in different places (e.g., that cat). The target remained constant while the distractors changed between tasks. Experiment 4 tests the condition where the distractors remain fixed and the target changes. This situation mirrors a search through the same kitchen cabinet for several different ingredients.

Experiment 4: One background, several targets

Method

Experiment 4 is preregistered on OSF (https://osf.io/yv73p). Figure 12 shows the tasks for Experiment 4. In this case, the image of a single search display serves to illustrate all five tasks, underlining the fact that the items in the background

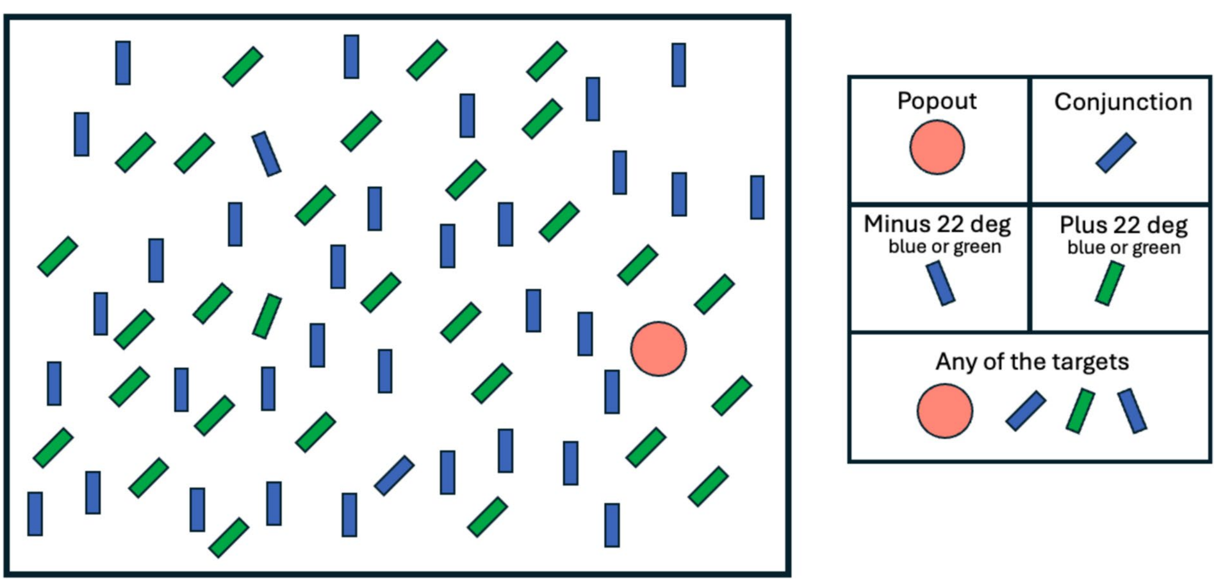


Fig. 12 Stimuli for Experiment 4. The background elements were the same for all tasks. In the case, where the target could be any of the targets, there was no specific cue (no-cue condition). For convenience, the figure shows examples of each of the four possible targets in a single panel. In the experiment, only one of the target types could appear on any trial. (Color figure online)

remained the same across tasks. The distractors were a random mixture of 50% blue vertical lines and 50% green lines tilted 45° clockwise from vertical. The specific arrangement of distractors changed from trial to trial. There were five tasks. An example of each of the targets is shown in the search display.

In the pop-out condition, the target was a large reddish circle that differed in color, size, orientation, and shape from the distractor items. The target for the conjunction task was a 45° tilted blue line. For the minus 22° condition, the target was tilted to the left by 22°, making it categorically unique and easy to find (Wolfe et al., 1992). For the plus 22° condition, the target was tilted to the right of vertical by 22°. Its orientation was flanked by the 0 and 45° distractors, making it hard to find. In both 22° tasks, the target could be randomly blue or green. In the no-cue task, the target could be any of the targets from the other four tasks, and participants were asked to search for the “oddball” item. For each task, the target or targets were cued for 1,000 ms before the trial. Set sizes were 8, 16, and 24. There were a total of 400 trials in the mixed and blocked conditions, with 80 trials for each task. Unlike in the previous experiments, in this case, half of the participants did the blocked condition first while the other half did the mixed condition first.

Twenty-six participants were tested in the lab and recruited through the Boston University SONA system (20 women, six men, Avg age: 21 years, range: 18–42). Of the 26 participants, six received payment at a rate of $15/hour. Twenty were recruited through the Boston University Psych 101 pool and received class credit for their participation. All gave consent, had at least 20/25 vision with correction, normal color vision, and no history of any eye or muscular disease.

Results and discussion

Two participants were removed from analysis because of poor performance ($d' < 0.5$) on the hard, plus 22° task. RTs were removed from analysis if they were less than 200 ms or greater than 10 s. The upper limit is higher than in previous experiments because observers were slower overall. This left 99.75% of the data.

Looking at Fig. 13, we see that there are effects of the Mixed vs. Blocked conditions for the pop-out, conjunction, and no-cue tasks. Specifically, the mixed condition is somewhat slower than the blocked condition. The results of three-way ANOVAs (Set Size × Present/Absent × Mixed/Blocked show significant main effects of mixed/block for those three tasks (all p values < 0.03, full details in Table S7). Separate two-way ANOVAs (Set Size × Mixed/Blocked) show that the mixed/blocked effect is not significant for present trials for pop-out, $F(1, 23) = 0.8449$, $p = 0.3676$. It is significant for conjunction, $F(1, 23) = 18.53$, $p = 0.0003$, and for no-cue, $F(1, 23) = 9.177$, $p = 0.0060$. For the absent trials, the mixed/blocked effect is significant for pop-out, $F(1, 23) = 33.03$, $p < 0.0001$, and conjunction, $F(1, 23) = 13.69$, $p = 0.0012$, but not for no-cue, $F(1, 23) = 2.786$, $p = 0.1087$, though the interaction with set size is weakly significant in that case, $F(2, 46) = 3.494$, $p = 0.0387$.

Turning to the error data, Fig. 14 shows that there are no very orderly differences between blocked and mixed conditions. In the plus 22° condition, there is a significant

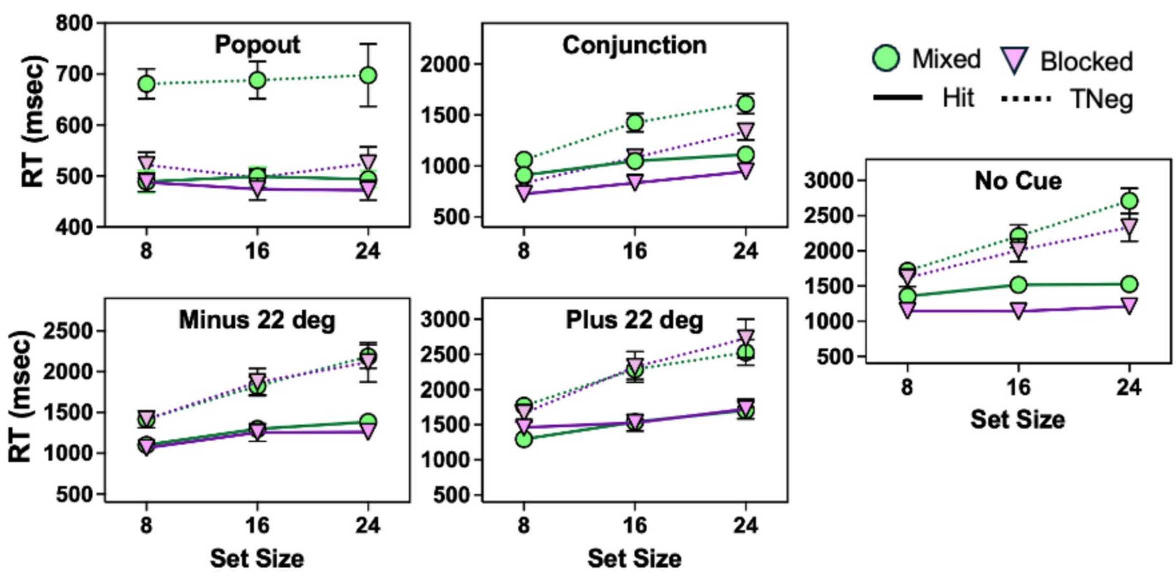


Fig. 13 RT × Set Size functions for the five tasks of Experiment 4. Mixed conditions are shown with green circles; blocked, with purple triangles. Correct present (hit) trials are shown with solid lines; true negative (TNeg), with dashed lines. Error bars, where visible are ±1 s.e.m. Note the very different y-axes for these tasks of different difficulty. (Color figure online)

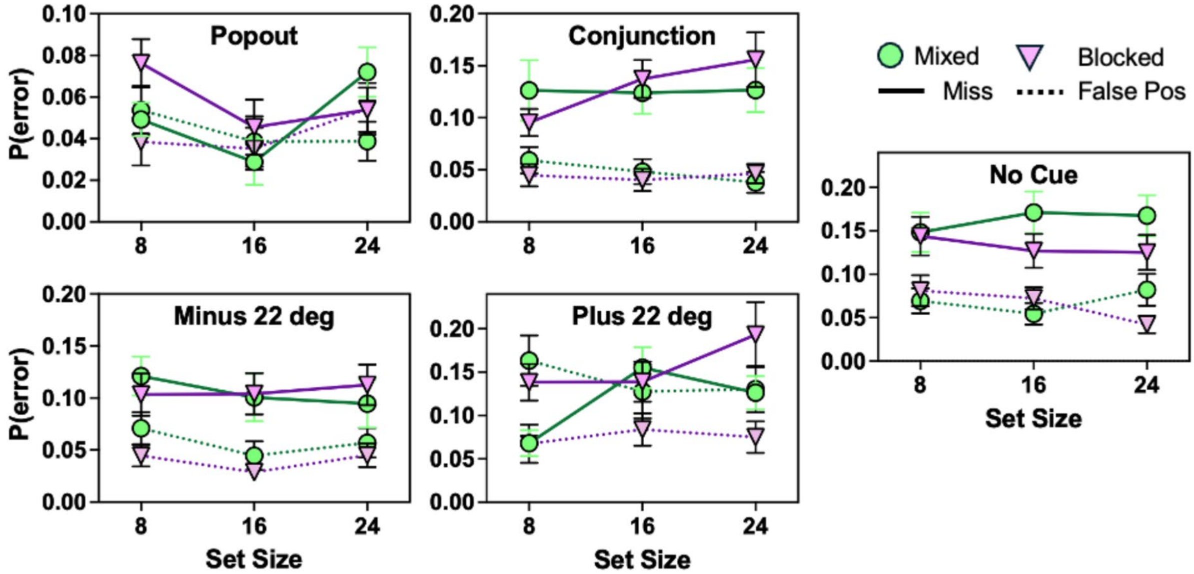


Fig. 14 Error × Set Size functions for the five tasks of Experiment 4. Mixed conditions are shown with green circles; blocked, with purple triangles. Correct present (hit) trials are shown with solid lines; true negative (TNeg), with dashed lines. Error bars, where visible are ± 1 s.e.m. Note the different *y*-axes for these tasks of different difficulty. (Color figure online)

interaction of mixed/block with target presence, $F(1, 23) = 14.08$, $p = 0.0010$, reflecting a higher rate of false-positive errors in the mixed condition. The triple interaction including set size is also significant, $F(2, 46) = 3.862$, $p = 0.0282$. No other tasks show any significant main effects or interactions of the mixed/block variable. Supplementary Table S9 shows three-way ANOVA results for the arcsin-transformed error data.

In this experiment, we can look at the effects of the order of mixed and blocked conditions. Supplementary Table S9 shows three-way ANOVAs with Mixed vs. Block (Within *O*s) × Order (1 st vs. 2nd, Between *O*s) × Set Size (Within *O*s). ANOVAs are done for each of the five conditions and for target present and absent separately. Participants get faster with practice. The main effect of order is significant in all cases (all p values < 0.05) except for the pop-out task where observers are fast from the start. This basic practice effect sometimes interacts with set size indicating that the slopes of RT × Set Size functions sometimes become shallower with practice. The order variable only interacts with the mixed/blocked variable for the absent trials for the plus 22° target task. In this task, blocked RTs are slower than mixed when in the first session and faster in the second. The blocked RTs change fairly dramatically from first to second session. The mixed RTs do not. Overall, however, the effect of order is that observers get faster with practice. They do not qualitatively change the relationship between mixed and blocked conditions.

Turning to the evidence for an adaptive quitting rule, Fig. 15 shows the pooled results for the change in true negative RTs following a miss.

In Experiment Four, there is evidence for an increase in RT immediately after a miss in blocked and mixed conditions. The effect is more robust in the blocked condition, but the RTs in the mixed condition are reliably elevated for the first TNeg after a miss and for the first TNeg of the same task as the miss.

The overall conclusion from Experiment 4 is that, as with Experiments 1–3, there are some modest effects of the mixed/blocked manipulation and, in this case, of the order of mixed and blocked conditions. However, none of these effects qualitatively change the search performance. We have now found similar results, sampling three of four cells in a 2 × 2 set of mixed vs. blocked experiments. Here, we held the distractors constant and varied the target (search for different items in the closet). In Experiment 2 and 3, we held the target constant and varied the distractors (search for the cat in different rooms). In Experiment 1, we varied both targets and distractors as in the daily run of searches for the coffee in cupboard, the coffee maker on the counter, the spoon in the drawer, and so forth. The fourth cell in which targets and distractors remain constant is not meaningful as a mixed/block condition.

In the final experiment, we try one, orthogonal manipulation. Does it matter if the observers can choose how trials are mixed? After all, in the real world, the sequence of searches is often under volitional control (coffee—> coffee

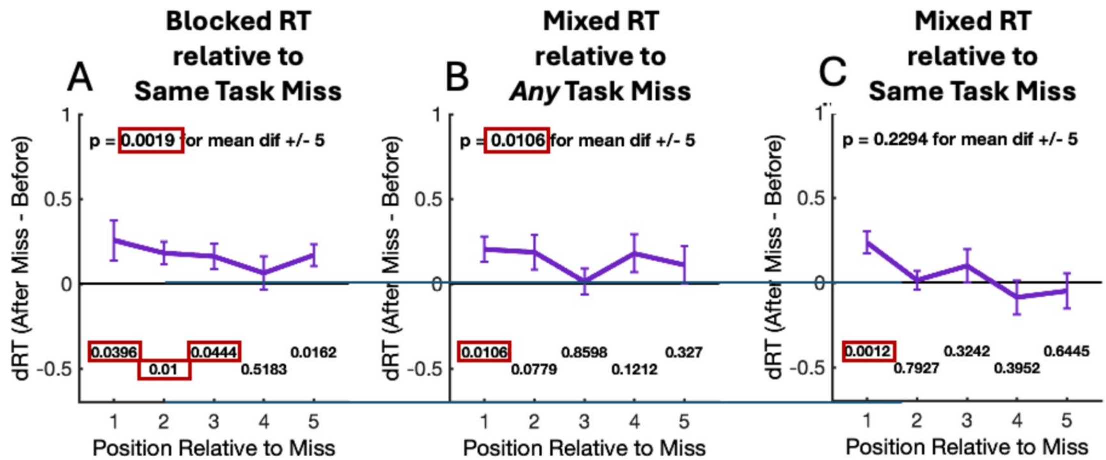


Fig. 15 Difference in RT for matched positions before and after a miss error for blocked and mixed condition of Experiment 4. Results are pooled over all five tasks. Panel **A** shows the blocked condition. Panel **B** shows the results for all misses regardless of whether the miss and TNeg trials involve the same task. Panel **C** shows the results where the TNeg trial task matches the miss trial task. (Color figure online)

maker—> spoon). Perhaps having some sense of agency makes mixed search easier.

Experiment 5: Choosing your task

Method

Experiment 5 was preregistered on OSF (https://osf.io/3ejm6). Observers completed 100 trials of each of four different visual search tasks. In all cases, the target was present on 50% of trials. The search field was square 0.9 of display height on a side. The four tasks were:

(1) TvsL task: The task was to find a *T* among nine *L*s placed at equidistant positions on an imaginary circular ring (radius: 0.3 of search field). Each letter has a height and width 0.05 of search field. *T*s and *L*s were randomly rotated between 0 and 359°. The location of a target, if present, was randomly selected from ten possible locations on each trial.

(2) Bouba–kiki search: The task was to find a "kiki" shape among "bouba" shapes positioned on an invisible circle (radius: 0.4 of search field). Boubas and kikis are pregenerated orange blobs whose outer contours are composed of the sum of several radial sine functions of differing amplitude (Bell et al., 2007). Ten different levels of bumpiness were created by adding more higher radial frequencies to create more bumpy items. For each level, 30 different exemplars were generated. Items' height and width was 0.1 of patch size. The five less bumpy levels were designated as boubas and would be considered a distractor. The five levels of the bumpier stimuli were kikis. Any of these could be the target, if present. Figure 16A shows examples. A search display consisted of a ring of 12 items. All were boubas except for one target kiki, when present.

(3) Complex conjunction task: As shown in Fig. 16B, the target in this task was a "plus" composed of green horizontal and red vertical bars. Distractors were plusses with red horizontal and blue vertical bars or crosses with green vertical and red horizontal bars. The search field was divided into a 10 × 10 grid and 16 locations were randomly selected as locations for stimuli. Each plus was 0.05 × 0.05 of the field size. This is an inefficient conjunction search since distinguishing red vertical/green horizontal from green vertical/red horizontal is known to be laborious (Wolfe & Bennett, 1997).

(4) Categorical search task: In this task, participants searched for one animal target among multiple distractor objects from different categories. The animal could be any of 50 photographic exemplars. All distractor items were also photographs of isolated real-world objects on a white background. Target and distractor images were taken from Brady et al. (2008). There were 14 distractor categories, each with 50 exemplars (700 potential distractors). Set size was 8 items. On each trial, each item was drawn from a unique category. Items were placed in a 5 × 5 grids, with each item having a size of 0.125 × 0.125 of the field size.

Figure 16C shows the display as seen by the participants. Buttons for the four tasks were presented on the left. Use of buttons is described below. Figures on the button show the

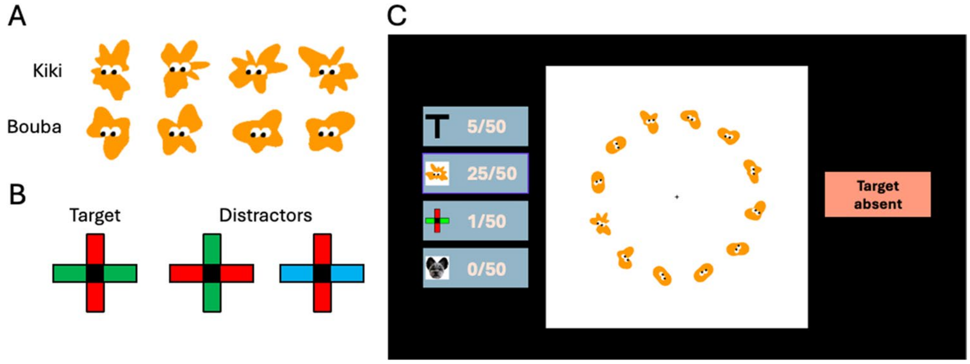


Fig. 16 Stimuli for two of the four tasks of Experiment 5. **A** bouba–kiki, **B** complex conjunction. The other two tasks are self-explanatory. Panel **C** shows the layout of the screen seen by participants (fonts enlarged for visibility here). (Color figure online)

number of trials completed over the total number of trials for that task. Participants clicked on the target or, if the target was absent, they clicked on the target-absent button to the right. Trial by trial feedback was provided by a "good" tone for correct responses or a "bad" tone for errors.

Each observer was randomly assigned to one of five conditions:

(1) Trial choice condition: In this condition, participants were allowed to choose which task to perform on each trial. At the beginning of each trial, the message "Choose your task" appeared in the center of the screen. Participants chose one of four tasks by clicking on the corresponding button on the left side of the patch (Fig. 16C). The button became deactivated when the number of trials completed in a task reached 50. If the deactivated button was clicked, the message "Choose another task" appeared at the center again.

(2) Block choice condition: In this condition, participants began by choosing one of the four tasks. Participants did not need to choose a task before each trial; the current trial would be drawn automatically from the most recently chosen task until the participant actively chose to switch tasks or until the full complement of trials for that task was exhausted. At the start of each trial, a "switch" button appeared at the center of the display for 700 ms. If participants clicked on that button within the 700 ms, they would then be prompted to select the task. If participants did not click the switch button within the 700 ms window, the button vanished and a trial from the current task would be presented. When the number of trials of one task reached its limit, participants were asked to select another task.

(3) Random condition: In this condition, participants did not choose which task to perform on each trial. The order of the trials was randomized and assigned to the participant before each trial. Before each trial, the target task was highlighted and activated, while the other buttons remained deactivated. The trial began only when the participant clicked on the correct task button.

(4) Blocked condition: In this condition, participants also had no choice. Here, trials of each task type were presented in a blocked fashion. After 50 trials of one type, the task was changed to another task. To keep the motor demands similar to the choice conditions, at the beginning of each trial, participants needed to click the highlighted button for the assigned task type. The order of blocks was counterbalanced across participants.

(5) Yoked condition: Participants in this condition saw trials in the order chosen by a participant in the trial choice condition. Again, to keep the motor demands similar to the choice conditions, at the beginning of each trial, participants needed to click the highlighted task button for the assigned task type though they had no choice in the matter.

Each participant was tested in two blocks of 200 trials (50 for each task). This took an average of 30 min per participant.

Participants

Across conditions, 51 participants were tested. The intention was to test 10 participants in each condition, but an error led to 11 participants in the random condition. All participants were recruited from the BU SONA Psych 101 pool, gave informed consent as approved by Mass General Brigham IRB Protocol #2009P001253, and received class credit. The average age was 18.8 years ($SD = 0.9$, range: 18–21). Thirty-four identified as female, 17 as male.

Results

RTs were excluded from analysis if they were shorter than 200 ms or longer than 7,000 ms. These exclusions were not preregistered. The RT exclusions left 98.6% of the data. No participants needed to be removed for poor performance.

Switching behavior

Given the choice, our participants did not choose to switch tasks often. In the random condition, where the tasks were randomly assigned on each trial, there were an average of 300 switches per participant, because choosing randomly among four tasks means that the remaining 100 trials will be repeats of the same task. In the blocked choice condition, where the task remained the same unless participants chose to switch, the range of switches ran from 1 (the switch when a task was finished does not count) to 14, with an average of 6.3 switches per participant. In the trial choice condition, where participants had to choose their task on each trial, one participant managed to make 396 switches, having apparently decided to move to a new task on each trial. Another produced 155 switches. The remaining eight participants averaged just 5.8 switches.

Response times

Figure 17 shows the average response times (RTs) for each type of search task in each of the conditions. The statistic at the bottom of each graph is the one-way ANOVA, testing for differences between conditions. None of these tests suggested a significant effect of condition. None of the pairwise comparisons were statistically significant once corrected for multiple comparison. There are RT differences between tasks (e.g., the Color × Color conjunction is slower) and between present and absent trials (absent is slower), but these standard results are not of particular interest in this case. The blocked and random conditions repeat the blocked and mixed conditions of the prior experiments. Adding choice, in either the BlockedChoice or TrialChoice variations, did not significantly change the results.

Errors

Figure 18A shows miss error rates, pooled across all four search tasks. There are quite dramatic differences in the error rates between tasks but none of the one-way ANOVAs across conditions within a task were significant for either miss errors or false-positive/false-alarm errors. The apparent trend for more errors in the blocked condition was significant by paired t tests (corrected for multiple comparison), but only for the TvL task and only for the comparisons between Blocked and BlockChoice ($p = 0.0269$), and Blocked and TrialChoice ($p = 0.0114$). Miss errors in the yoked condition, while apparently a bit elevated, were not significantly greater than other error rates.

With error data, it is suggested that the data can be arcsine transformed to make the data more normally distributed. That is shown in Fig. 18B. Here, the effect of condition is modestly significant in the one-way ANOVA. Still, none of the pairwise comparisons are significant. Thus, there is a possibly reliable effect of condition on miss errors. Perhaps participants became bored and a bit careless in the blocked condition. In any case, the effect of condition on error is not dramatic and certainly does not suggest that errors increase when participants have choice. Nor are the choice condition error rates particularly low. It is clear that they are essentially identical to the errors in the random condition.

Discussion

The results of Experiment 5 constitute a reasonably interesting, essentially null result. In this situation, giving people the ability to choose their tasks had little or no effect, for better or for worse. Either it is that choice does not matter in these tasks or that participants suspected that switching was a bad idea and chose not to switch. We do not have enough data to determine if those who chose to switch behaved differently from those who did not.

General discussion

This paper began by raising the disturbing possibility that decades of research on blocks of trials of one search task might describe rules of search that would not apply to the real world of constantly changing search tasks. The main message of the experiments described here is that the situation does not appear to be that dire. For methodological reasons, we could not have participants perform a unique search on each trial and have them do a block of trials for each unique target, but we could ask them to perform four different tasks in a randomly mixed manner and in blocks of the same task. Switching randomly amongst four search tasks produced RT and error data that were qualitatively similar to the results when each of the four tasks was performed in its own block of trials. This was true when targets and distractors were different for each task (Experiment 1). It remained true when the same target was presented on all tasks but the distractors and, thus, the nature of the search changed from task to task (Experiments 2 & 3). Similar results were obtained when the distractor set was fixed for all tasks and only the target changed (Experiment 4). Finally, giving participants control over their choice of task on a trial-by-trial basis made little, if any, difference (Experiment 5).

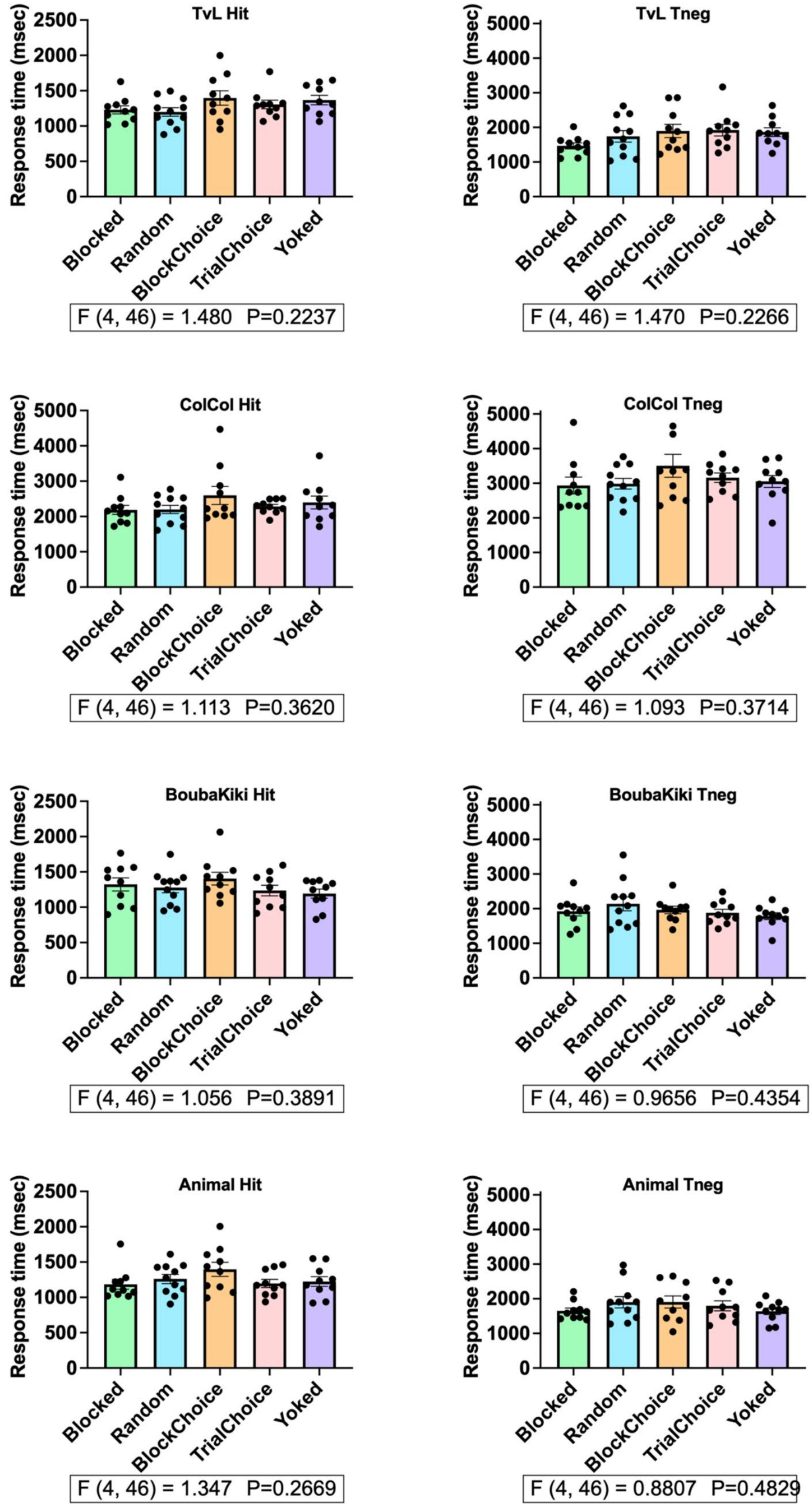


Fig. 17 Response times for each condition and each search task. Black dots show data for individual observers. Left column shows true positive (hit) trials. The right column shows true negative RTs. "Colcol" is an abbreviation for Color × Color conjunctions. (Color figure online)

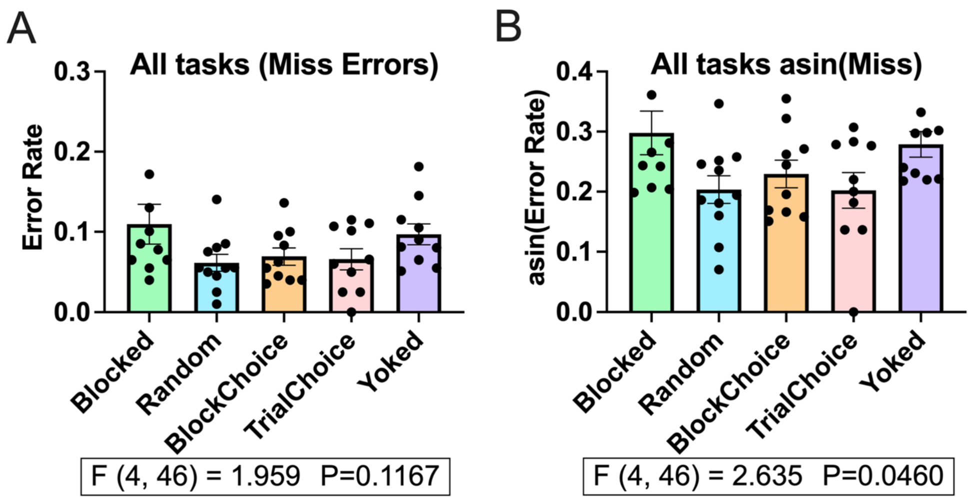


Fig. 18 Miss error rates pooled over all search tasks. **A** Raw error data, **B** arcsine transformed error data. Data points are individual participants. (Color figure online)

There are a few effects that rise to the level of statistical significance. The most interesting of these is the possibility, seen most clearly in Experiment 3 (Fig. 9), that RTs in the mixed condition move toward the average RT across tasks. The easiest tasks become a bit slower, and the harder tasks may become a bit faster when all the tasks are intermixed. This effect is only hinted at in other experiments but might be worth further exploration with more extreme conditions. One could imagine that a hard task among several easy ones would show a more dramatic change, for example.

The lack of major effects of mixing tasks has implications for theories of search. In particular, these essentially null results suggest that we should not overemphasize the role of search history in search performance. In recent years, there has been a strong interest in history effects such as feature priming. Awh et al. (2012) published an influential paper entitled "Top-down Versus Bottom-Up Attentional Control: A Failed Theoretical Dichotomy" that promoted trial by trial effects as an important source of attentional control. Our Guided Search model, which had relied solely on top-down and bottom-up guidance in its earlier versions (Wolfe, 1994; Wolfe et al., 1989), has adopted "history" as one of five types of guidance (Wolfe & Horowitz, 2017) in its most recent incarnation (Wolfe, 2021). A large and growing literature is concerned with this issue (e.g., Anderson et al., 2021; Becker et al., 2023; Kristjánsson & Campana, 2010; Kristjansson & Driver, 2008; Kruijne & Meeter, 2015; Lamy et al., 2010; Maljkovic & Nakayama, 1994; Ramgir & Lamy, 2020; Wolfe et al., 2019; Wolfe et al., 2022a, 2022b; Zehetleitner et al., 2012). There is no doubt that the history of search exerts a significant effect on search behavior, biasing attention toward or away from some locations and priming or suppressing different basic features, though we did not find repetition priming when we looked in Experiment 2a. The present results do not show strong trial-by-trial effects. If trial history was an especially strong determinant of search behavior, we would expect more of an effect from major changes in that history. This is not to say that search history is not a force in guiding attention in visual search. The massive body of literature noted above makes it clear that what happened on recent trials is important. However, these results argue that these history effects can be secondary to other factors.

The present results raise a similar issue regarding the termination of target-absent trials. There have been multiple efforts to model search termination rules (Chun & Wolfe, 1996; Moran et al., 2013; Schwarz & Miller, 2016; Zenger & Fahle, 1997). As a general rule, they have based the quitting rule on some version of an adaptive process, based on performance on the previous trials. It has always been clear that the quitting threshold is not set by a simple 'staircase' procedure where the observer gets faster after a correct response and slower after an error. Search termination must take the set size into account (or some equivalent of set size in the case of search in scenes). Observers manage to terminate search successfully even when they do not get trial by trial feedback. Indeed, Mazo and Fleming (2022) showed that observers can end a target-absent search before ever seeing a target-present search. Observers were looking for a red dot, either amongst blue dots (feature search) or amongst blue dots and red squares (conjunction search). The clever wrinkle was that the first four trials were all target absent (2 tasks × 2 set sizes). On those first four trials, observers clearly showed that they knew they could quit more quickly in the feature search and when the set size was smaller. In subsequent trials, RTs did tend to get faster, but the observers could start the task with a sensible quitting rule.

The present data also suggest that observers have reasonable priors that allow them to end unsuccessful searches in a reasonable manner. The data might have shown that all absent trials in the Mixed condition used a single, compromise quitting rule. This would have produced absent RTs that were too long for easy tasks and probably would have produced elevated miss errors for harder tasks. This result was not seen. In general, mixed and blocked target-absent RTs and errors are comparable. There is some evidence for an adaptive process that increases RT after an error. The evidence is fairly consistent for the blocked conditions and is less clear for the mixed conditions. The results are consistent with the idea that, given a new search trial, observers have an immediate idea about how long that search should take. This estimate is based on a lifetime of search and not merely on experience in an experiment. Experience in the experiment can probably fine-tune the termination threshold, but the prior with which the observer entered the task is reasonable, at least for the tasks used here. More research would be worthwhile. In particular, it would be interesting to examine the changes in search termination rules that must accompany learning a new task (e.g., search in medical images). One could imagine that it would be possible to create search tasks where target-absent responses were initially markedly too fast or slow. In such a case, the proposed adaptive adjustment of quitting thresholds might be much more obvious.

Limitations and future directions

The present work has several limitations that point in the direction of future work. First, the set of tasks that were used were necessarily limited. The classic search literature is filled with other search tasks that could be compared in the mixed and blocked conditions. Among the more interesting cases would be search in continuous scenes, rather than in arrays of simple objects and search by experts where it would be interesting to know if grouping searches by type would be helpful. For instance, would it help to read mammograms of dense (harder) breasts in one block and "fatty" (easier) breasts in another block (Gommers et al., 2024)?

We used four tasks at a time. It could be informative to vary that number, though using larger numbers of tasks would require longer experiments. With larger numbers of tasks, the average time between instances of any one task would become longer. Another way to manipulate the time between instances is to vary the relative prevalence of tasks. Would we see more of an effect of mixing tasks if one of the tasks was relatively rare. Certainly, we know that search performance changes when targets are rare (Horowitz, 2017; Wolfe et al., 2005). There is data showing that there can be a cost for the first "surprise" appearance of a task (Ernst et al., 2024; Horstmann & Ansorge, 2016). We don't know if performance would be disrupted for a rare task if it were intermixed with more common tasks.

In all of the experiments reported here, we provided trial-by-trial feedback. Obviously, that is not the case for many real-world tasks. In many cases in the real world, finding the target is its own feedback, but not finding it does not tell the searcher if they correctly terminated a target-absent search or missed the target on a target-present search. It would be valuable to manipulate feedback as it is known that it can have a significant effect in other search contexts (Lyu et al., 2024).

There are methodological limitations in the present work. While it seemed like a good idea at the time, it was unwise to have observers perform the Blocked condition before the Mixed condition in Experiments 1 and 2. This confounded a learning effect with a mixed/block effect. Our later experiments indicate that this confound did not hide a penalty for the mixed condition, but counterbalancing would have been the correct approach from the start. These experiments also reveal some of the perils of online testing. In several experiments, we lost unusually large number of participants (~ 50%) because they seem to have decided not to perform one or the other task—typically the hardest of the tasks. In future work, we will want to clarify the instructions (and rewards) to discourage such behavior.

Conclusion

In sum, the primary message of this paper is that the search rules obtained from testing participants on blocks of the same type of search appear to remain qualitatively similar when different search tasks are intermixed. Though there are many further directions to explore, this finding should be reassuring to authors who have started search papers with an example of a single-trial, real-world search only to proceed to study the question with hundreds of trials of one task.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13414-025-03077-8.

Funding Jeremy Wolfe, Injae Hong, and Ava Mitra were supported by National Institute of Health-National Eye Institute: EY017001 and National Science Foundation: 2146617. Eduard Objio Hula Khalifa, and Yousra Ali were supported by Harvard Medical School's Project Success and by Supplement to J.W.'s NSF grant.

Data availability All the experiments were preregistered on the Open Science Framework, and all study data, materials, and code can be accessed at: Experiment 1: https://osf.io/26jsr/, Experiment 2: https://osf.io/cmjrb/, Experiment 3: https://osf.io/m9z57/, Experiment 4: https://osf.io/yv73p, Experiment 5: https://osf.io/3ejm6

Declarations

Ethics approval and consent to participate The research was approved by the Mass General Brigham IRB Protocol #2007P000646, and all participants gave consent at the start of the online experiments.

Consent for publication All authors and participants consent to publication of this work.

Conflicts of interest All authors declare no conflict of interest.

References

Anderson, B. A., Kim, H., Kim, A. J., Liao, M.-R., Mrkonja, L., Clement, A., & Grégoire, L. (2021). The past, present, and future of selection history. *Neuroscience & Biobehavioral Reviews, 130*, 326–350. https://doi.org/10.1016/j.neubiorev.2021.09.004

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences, 16*(8), 437–443. https://doi.org/10.1016/j.tics.2012.06.010

Bauer, B., Jolicoeur, P., & Cowan, W. B. (1996a). Distractor heterogeneity versus linear separability in colour visual search. *Perception, 25*(1), 1281–1294.

Bauer, B., Jolicœur, P., & Cowan, W. B. (1996b). Visual search for colour targets that are or are not linearly-separable from distractors. *Vision Research, 36*(10), 1439–1466.

Becker, M. W., Rodriguez, A., & Pontious, D. (2022). Quitting thresholds in visual search are impacted by target present detection times but not their variability. *Attention, Perception, & Psychophysics, 84*, 2461–2471.

Becker, S. I., Grubert, A., Horstmann, G., & Ansorge, U. (2023). Which processes dominate visual search: Bottom-up feature contrast, top-down tuning or trial history? *Cognition*, *236*, Article 105420. https://doi.org/10.1016/j.cognition.2023.105420

Bell, J., Badcock, D. R., Wilson, H., & Wilkinson, F. (2007). Detection of shape in radial frequency contours: Independence of local and global form information. *Vision Research, 47*(11), 1518–1522. https://doi.org/10.1016/j.visres.2007.01.006

Berbaum, K. S., Franken, E. A., Jr., Dorfman, D. D., Rooholamini, S. A., Kathol, M. H., Barloon, T. J., . . . Montgomery, W. J. (1990). Satisfaction of search in diagnostic radiology. *Investigative Radiology*, *25*(2), 133–140. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=2312249

Berge, P., Darsonval, A., Nedelcu, C., Paisant, A., & Aubé, C. (2020). Incidental findings on emergency CT scans: Predictive factors and medico-economic impact. *European Journal of Radiology, 129,* Article 109072. https://doi.org/10.1016/j.ejrad.2020.109072

Biederman, I., Glass, A. L., & Stacy, E. W. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology, 97*, 22–27.

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 of the United States of America, 105*(38), 14325–14329.

Buetti, S., Cronin, D. A., Madison, A. M., Wang, Z., & Lleras, A. (2016). Towards a better understanding of parallel visual processing in human vision: Evidence for exhaustive analysis of visual information. *Journal of Experimental Psychology: General, 145*(6), 672–707. https://doi.org/10.1037/xge0000163

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology, 30*, 39–78.

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

Egeth, H. (1977). Attention and preattention. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11, pp. 277–320). Academic Press.

Egeth, H., Jonides, J., & Wall, S. (1972). Parallel processing of multielement displays. *Cognitive Psychology, 3*, 674–698.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for conjunctively defined targets. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 32–39.

Enns, J. T., & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science, 247*, 721–723.

Ernst, D., Horstmann, G. A., Hulleman, J., & Wolfe, J. M. (2024). *Inattentional blindness for a salient target in visual search: Finding a surprisingly easy target can be surprisingly hard.* Paper presented at Vision Sciences Society meeting, St. Petersburg Beach, FL.

Fahle, M. (1991a). A new elementary feature of vision. *Investivative Ophthalmology & Visual Science, 32*(7), 2151–2155.

Fahle, M. (1991b). Parallel perception of Vernier offests, curvature, and chevrons in humans. *Vision Research, 31*(12), 2149–2184.

Findlay, J. M. (1973). Feature detectors and Vernier acuity. *Nature*, *241*(5385), 135–137. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=4695545

Gale, A. G., Mugglestone, M. D., Purdy, K. J., & McClumpha, A. (2000). Is airport baggage inspection just another medical image? In E. A. Krupinski (Ed.), *Proceedings of Medical Imaging 2000: Image Perception and Performance, 3981,* 184–192.

Gandomkar, Z., & Mello-Thoms, C. (2019). Visual search in breast imaging. *British Journal of Radiology*, *92*(1102), Article 20190057. https://doi.org/10.1259/bjr.20190057

Godwin, H. J., Menneer, T., Cave, K. R., Helman, S., Way, R. L., & Donnelly, N. (2010). The impact of relative prevalence on dual-target search for threat items from airport X-ray screening. *Acta Psychologica, 134*(1), 79–84. https://doi.org/10.1016/j.actpsy.2009.12.009

Gommers, J. J. J., Verboom, S. D., Duvivier, K. M., van Rooden, J.-K., van Raamt, A. F., Houwers, J. B., . . . Wolfe, S. (2024). Enhancing radiologist reading performance by ordering screening mammograms based on characteristics that promote visual adaptation. *Radiology*, *313*(1), Article e240237. https://doi.org/10.1148/radiol.240237

Green, B. F., & Anderson, L. K. (1956). Color coding in a visual search task. *Journal of Experimental Psychology, 51*, 19–24.

Gur, D., Rockette, H. E., Warfel, T., Lacomis, J. M., & Fuhrman, C. R. (2003). From the laboratory to the clinic: The "prevalence effect". *Academic Radiology*, *10*(11), 1324–1326. http://www.ncbi.nlm.nih.gov/pubmed/14626308

Hatfield, G. (1998). Attention in early scientific psychology. In R. D. Wright (Ed.), *Visual attention* (pp. 3–25). Oxford University Press.

Hogg, R. V., & Craig, A. T. (1995). *Introduction to mathematical statistics* (5th ed.). Prentice-Hall.

Horowitz, T. S. (2006). Revisiting the variable memory model of visual search. *Visual Cognition, 19*(4/8), 668–684.

Horowitz, T. S. (2017). Prevalence in visual search: From the clinic to the lab and back again. *Japanese Psychological Research, 59*(2), 65–108. https://doi.org/10.1111/jpr.12153

Horowitz, T. S., & Wolfe, J. M. (1998). Visual search has no memory. *Nature*, *394*(Aug 6), 575–577.

Horstmann, G., & Ansorge, U. (2016). Surprise capture and inattentional blindness. *Cognition, 157*, 237–249. https://doi.org/10.1016/j.cognition.2016.09.005

Huang, L., Holcombe, A. O., & Pashler, H. (2004). Repetition priming in visual search: Episodic retrieval, not feature priming. *Memory & Cognition*, *32*(1), 12–20. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=15078040

Hwang, A. D., Wang, H.-C., & Pomplun, M. (2011). Semantic guidance of eye movements in real-world scenes. *Vision Research, 51*(10), 1192–1205.

Kingsley, H. L. (1926). Search: A function intermediate between perception and thinking. *Psychological Monographs, 35*, 16–55.

Kingsley, H. L. (1932). An experimental study of 'search.' *American Journal of Psychology, 44*, 314–318.

Koopman, B. O. (1956). The theory of search. *II. Target Detection. Operations Research, 4*(5), 503–531.

Kristjánsson, A., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception & Psychophysics, 72*(1), 5–18.

Kristjansson, A., & Driver, J. (2008). Priming in visual search: Separating the effects of target repetition, distractor repetition and role-reversal. *Vision Research, 48*(10), 1217–1232.

Kruijne, W., & Meeter, M. (2015). The long and the short of priming in visual search. *Attention, Perception, & Psychophysics, 77*(5), 1558–1573.

Krupinski, E. A. (1996). Visual scanning patterns of radiologists searching mammograms. *Academic Radiology*, *3*(2), 137–144. http://www.ncbi.nlm.nih.gov/pubmed/8796654

Kundel, H. L., Nodine, C. F., & Carmody, D. (1978). Visual scanning, pattern recognition and decision-making in pulmonary nodule detection. *Investigative Radiology*, *13*(3), 175–181. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=711391

Lamy, D., Yashar, A., & Ruderman, L. (2010). A dual-stage account of inter-trial priming effects. *Vision Research*, *50*(14), 1396–1401. http://www.doi.org/https://doi.org/10.1016/j.visres.2010.01.008

Lyu, W., Trueblood, J. S., & Wolfe, J. M. (2024). Effects of prevalence and feedback in the identification of blast cells in peripheral blood: Expert and novice observers. *Cognitive Research: Principles and Implications (CRPI).* Manuscript under review.

Mackworth, N. H. (1948). The breakdown of vigilance during prolonged visual search. *Quaterly Journal of Experimental Psychology, 1*, 6–21.

Maljkovic, V., & Nakayama, K. (1994). Priming of popout: I. *Role of Features. Memory & Cognition, 22*(6), 657–672.

Mazor, M., & Fleming, S. M. (2022). Efficient search termination without task experience. *Journal of Experimental Psychology: General, 151*(10), 2494–2510. https://doi.org/10.1037/xge0001188

Mendes, M., Kuhnb, M., & Schwaninger, A. (2015). X-ray security screening at airports: Enhancement of human factors. *Ergonomics*.

Moran, R., Zehetleitner, M. H., Mueller, H. J., & Usher, M. (2013). Competitive guided search: Meeting the challenge of benchmark RT distributions. *Journal of of Vision*, *13*(8), Article 24. https://doi.org/10.1167/13.8.24

Muhl-Richardson, A., Parker, M. G., Recio, S. A., Tortosa-Molina, M., Daffron, J. L., & Davis, G. J. (2020). Improved threat detection in baggage X-rays using 'targetless' search training. *Cognitive Research: Principles and Implications (CRPI)*.

Nagy, A. L., & Sanchez, R. R. (1990). Critical color differences determined with a visual search task. *Journal of the Optical Society of America-A, 7*(7), 1209–1217.

Neider, M. B., & Zelinsky, G. J. (2008). Exploring set size effects in scenes: Identifying the objects of search. *Visual Cognition*, *16*(1), 1–10. http://www.doi.org/https://doi.org/10.1080/13506280701381691

Neisser, U. (1963). Decision time without reaction time: Experiments in visual scanning. *American Journal of Psychology, 76*, 376–385.

Neisser, U., Novick, R., & Lazar, R. (1963). Searching for ten targets simultaneously. *Perceptual and Motor Skills, 17*, 955–961. https://doi.org/10.2466/pms.1963.17.3.955

Peltier, C., & Becker, M. W. (2016). Decision processes in visual search as a function of target prevalence. *Journal of Experimental Psychology: Human Perception and Performance, 42*(9), 1466–1476. https://doi.org/10.1037/xhp0000248

Ramgir, A., & Lamy, D. (2020). Top-down, bottom-up and selection-history-based control of attention: A failed trichotomy? The case of Priming of pop-out (PoP). *PsyArXiv Preprints.*

Ramgir, A., & Lamy, D. (2021). Top-down, bottom-up and selection-history-based control of attention: A failed trichotomy? The case of Priming of pop-out (PoP). *Psychonomic Bulletin & Review*. https://doi.org/10.3758/s13423-021-01997-8

Rose, R. M., Teller, D. Y., & Rendleman, P. (1970). Statistical properties of staircase estimates. *Perception & Psychophysics, 8*(4), 199–204.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review, 84*, 1–66.

Schwarz, W., & Miller, J. O. (2016). GSDT: An integrative model of visual search *Journal of Experimental Psychology: Human Perception and Performance*, *42*(10), 1654–1675. Advance online publication. https://doi.org/10.1037/xhp0000247

Shiffrin, M. R., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychonomic Review, 84*, 127–190.

Shiffrin, R. M., & Gardner, G. T. (1972). Visual processing capacity and attentional control. *Journal of Experimental Psychology, 93*(1), 72–82.

Sperling, G., Budiansky, J., Spivak, J., & Johnson, M. C. (1971). Extremely rapid visual search. The maximum rate of scanning letters for the presence of a numeral. *Science, 174*, 307–311.

Sternberg, S., & Scarborough, D. L. (1971, July 23–26). Parallel testing of stimuli in visual search. Paper presented at the Visual Information Processing and Control of Motor Activity, Sofia, Bulgaria.

Sun, J., & Perona, P. (1998). Where is the sun? *Nature Neuroscience*, *1*(3), 183–184. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10195141

Taylor, P. M. (2007). A review of research into the development of radiologic expertise: Implications for computer-based training [Review]. *Academic Radiology, 14*(10), 1252–1263.

Thomas, E. L., & Lansdown, E. L. (1963). Visual search patterns of radiologists in training. *Radiology, 81*, 288–291. https://doi.org/10.1148/81.2.288

Townsend, J. T. (1971). A note on the identification of parallel and serial processes. *Perception and Psychophysics, 10*, 161–163.

Treisman, A., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of seperable features. *Journal of Experimental Psychology-General, 114*, 285–310.

Tuddenham, W. J., & Calvert, W. P. (1961). Visual search patterns in roentgen diagnosis. *Radiology*, *76*, 255–256. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=13778547

Vighneshvel, T., & Arun, S. P. (2013). Does linear separability really matter? Complex visual search is explained by simple search *Journal of Vision*, *13*(11), Article 10. http://www.journalofvision.org/cgi/content/abstract/13/11/10?etoc

Wolfe, J. M. (1994). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Wolfe, J. M. (2012). When do I quit? The search termination problem in visual search. In M. D. Dodd spsampsps J. H. Flowers

(Eds.), *The influence of attention, learning, and motivation on visual search* (pp. 183–208). [Nebraska Symposium on Motivation]. Springer.

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review, 28*, 1060–1092. https://doi.org/10.3758/s13423-020-01859-9

Wolfe, J. M. (2023). Visual Search. In O. Braddick (Ed.), *Oxford research encyclopedia of psychology*. Oxford University Press.

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011a). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics, 73*(6), 1650–1671. https://doi.org/10.3758/s13414-011-0153-3

Wolfe, J. M., & Bennett, S. C. (1997). Preattentive object files: Shapeless bundles of basic features. *Vision Research, 37*(1), 25–43.

Wolfe, J. M., Brunelli, D. N., Rubinstein, J., & Horowitz, T. S. (2013). Prevalence effects in newly trained airport checkpoint screeners: Trained observers miss rare targets, too. *Journal of Vision*, *13*(3:33), 1–9. https://doi.org/10.1167/13.3.33

Wolfe, J. M., Cain, M. S., & Aizenman, A. M. (2019). Guidance and selection history in hybrid foraging visual search [journal article]. *Attention, Perception, & Psychophysics, 81*(3), 637–653. https://doi.org/10.3758/s13414-018-01649-5

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

Wolfe, J. M., Friedman-Hill, S. R., Stewart, M. I., & O'Connell, K. M. (1992). The role of categorization in visual search for orientation. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 34–49. https://doi.org/10.1037/0096-1523.18.1.34

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*(6), 495–501.

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search [Review Article]. *Nature Human Behaviour, 1*, 0058. https://doi.org/10.1038/s41562-017-0058

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare targets are often missed in visual search. *Nature, 435*(7041), 439–440. https://doi.org/10.1038/435439a

Wolfe, J. M., Kosovicheva, A., & Wolfe, B. A. (2022a). Normal blindness: When we look but fail to see. *Trends in Cognitive Science, 26*, 809–819. https://doi.org/10.1016/j.tics.2022.06.006

Wolfe, J. M., Palmer, E. M., & Horowitz, T. S. (2010). Reaction time distributions constrain models of visual search. *Vision Research, 50*, 1304–1311. https://doi.org/10.1016/j.visres.2009.11.002

Wolfe, J. M., Suresh, S. B., Dewulf, A. W., & Lyu, W. (2022b). Priming effects in inefficient visual search: Real, but transient. *Attention, Perception, & Psychophysics, 84*, 1417–1431. https://doi.org/10.3758/s13414-022-02503-5

Wolfe, J. M., Vo, M.L.-H., Evans, K. K., & Greene, M. R. (2011b). Visual search in scenes involves selective and non-selective pathways. *Trends in Cognitive Science, 15*(2), 77–84. https://doi.org/10.1016/j.tics.2010.12.001

Xu, Z. J., Lleras, A., Shao, Y., & Buetti, S. (2021). Distractor-distractor interactions in visual search for oriented targets explain the increased difficulty observed in nonlinearly separable conditions. *Journal of Experimental Psychology: Humman Perception and Performance, 47*(9), 1274–1297. https://doi.org/10.1037/xhp0000941

Yang, Z., Mondal, S., Ahn, S., Zelinsky, G., Hoai, M., spsampsps Samaras, D. (2022). Target-absent human attention. *European Conference on Computer Vision*, 52–68.

Zehetleitner, M., Rangelov, D., & Müller, H. (2012). Partial repetition costs persist in nonsearch compound tasks: Evidence for multiple-weighting-systems hypothesis. *Attention, Perception, & Psychophysics, 74*(5), 879–890. https://doi.org/10.3758/s13414-012-0287-y

Zenger, B., & Fahle, M. (1997). Missed targets are more frequent than false alarms: A model for error rates in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 23*(6), 1783–1791.