Journal of Experimental Psychology: Applied

A Little Bit of History Repeating: Splitting up Multiple-Target Visual Searches Decreases Second-Target Miss Errors

Matthew S. Cain, Adam T. Biggs, Elise F. Darling, and Stephen R. Mitroff

Online First Publication, April 7, 2014. http://dx.doi.org/10.1037/xap0000014

Journal of Experimental Psychology: Applied

1076-898X/14/$12.00 DOI: 10.1037/xap0000014

A Little Bit of History Repeating: Splitting up Multiple-Target Visual Searches Decreases Second-Target Miss Errors

Matthew S. Cain, Adam T. Biggs, Elise F. Darling, and Stephen R. Mitroff
Duke University

Visual searches with several targets in a display have been shown to be particularly prone to miss errors in both academic laboratory searches and professional searches such as radiology and baggage screening. Specifically, finding 1 target in a display can reduce the likelihood of detecting additional targets. This phenomenon was originally referred to as "satisfaction of search," but is referred to here as "subsequent search misses" (SSMs). SSM errors have been linked to a variety of causes, and recent evidence supports a working memory deficit wherein finding a target consumes working memory resources that would otherwise aid subsequent search for additional targets (Cain & Mitroff, 2013). The current study demonstrated that dividing 1 multiple-target search into several single-target searches, separated by three to five unrelated trials, effectively freed the working memory resources used by the found target and eliminated SSM errors. This effect was demonstrated with both university community participants and with professional visual searchers from the Transportation Security Administration, suggesting it may be a generally applicable technique for improving multiple-target visual search accuracy.

Keywords: visual search, satisfaction of search, multiple-target search, Transportation Security Administration

Many visual searches involve looking for more than one item at a time. When looking for your shoes on a messy floor, you are unlikely to be content to end your search having found only one of them. A similar problem confronts professional visual searchers; for example, airport security screeners must find all potentially dangerous items in a baggage X-ray image, and radiologists must find all abnormalities in a medical MRI scan. However, unlike when looking for your shoes, professional searchers typically do not know in advance how many targets are actually present.

Most laboratory investigations of visual search have focused on the case where, at most, one target is present in any given display (see Eckstein, 2011; Nakayama & Martini, 2011 for recent reviews). If a target is found, the search can be terminated immediately. If no target is found, the searcher must decide when they have searched "long enough," which is usually about twice as long as it takes to find a target when one is present. This quitting threshold for single-target search has been extensively investigated (e.g., Chun & Wolfe, 1996; Wolfe, 2007, 2012).

When there is a possibility of additional targets being present in a search, the situation becomes more complicated. After finding one target, the searcher still needs to determine if a second target is present. The found target also significantly diminishes available cognitive resources because its location and/or identity is likely being stored in working memory (Cain & Mitroff, 2013). The decision about when to quit is also more complicated, and can be modeled as foraging behavior (Cain, Vul, Clark, & Mitroff, 2012; Wolfe, 2013), which maximizes the rate of finding targets rather than minimizing the rate of missing targets. Foraging strategies can be beneficial in some circumstances (e.g., acquiring as many berries as possible), but can be highly problematic for searches where any missed target could have dire consequences (e.g., making sure absolutely no bombs get onto airplanes).

One frequent outcome in dual-target search is that after a target has been found, a second target in the display is less likely to be found than if it were the only target in the display. This problem has long been studied in academic radiology and was originally referred to as "satisfaction of search" (Berbaum, Franken, Caldwell, & Schartz, 2010; Smith, 1967; Tuddenham, 1962). We have recently proposed the more theory-neutral term "subsequent search misses" (SSMs; Adamo, Cain, & Mitroff, 2013; Cain, Adamo, & Mitroff, 2013) as the original name does not accurately

Matthew S. Cain, Adam T. Biggs, Elise F. Darling, and Stephen R. Mitroff, Center for Cognitive Neuroscience, Department of Psychology and Neuroscience, Duke University.

Matthew S. Cain is now at the Visual Attention Lab, Brigham & Women's Hospital and Harvard Medical School, Cambridge, Massachusetts.

This work was partially supported by the Army Research Office (54528LS) and partially through a subcontract with the Institute for Homeland Security Solutions, a research consortium sponsored by the Resilient Systems Division in the Department of Homeland Security (DHS). This material is based upon work supported by the DHS under Contract HSHQDC-08-C-00100. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the official policy or position of DHS or of the U.S. Government. The study is approved for public release. We thank Michael Zunk, Hung Nguyen, Bob Kingan, Kellis Turner, Ken Macdonald, Richard Menard, and Everett Vaughn for making it possible to conduct research at RDU airport. We especially thank Dave Evans and Stuart Campbell for their extraordinary efforts to support the research at RDU.

Correspondence concerning this article should be addressed to Matthew S. Cain, 64 Sidney St., Suite 170, Cambridge, MA 02139. E-mail: mcain@partners.org

account for the phenomenon's causes. SSM errors are a substantial problem in radiology, accounting for up to a third of radiological misses (see Berbaum et al., 2012 for a recent review), and may impact related accuracy-focused professional searches such as security screening (Biggs & Mitroff, 2013) and safety inspections. While much effort has been spent trying to understand SSM errors, there does not appear to be a single cause, but rather a number of underlying problems that manifest as increased miss errors (Cain et al., 2013). While some potential causes have largely been ruled out, such as time on task (Berbaum et al., 1991, but see 2013), many other factors have been shown to contribute to SSM errors, such as misdirection of attention (Berbaum et al., 1996), time pressure (Fleck, Samei, & Mitroff, 2010), anticipatory anxiety (Cain, Dunsmoor, LaBar, & Mitroff, 2011), global contextual pressure (Clark, Cain, Adcock, & Mitroff, 2014), working memory resource depletion (Cain & Mitroff, 2013), and an attentional refractory period (Adamo et al., 2013).

With so many contributing factors, SSM errors have proven stubbornly difficult to eliminate. A number of experimental interventions have been attempted, such as manipulating the order of reporting abnormalities (Berbaum, Franken, Dorfman, Caldwell, & Lu, 2005) using checklists (Berbaum, Franken, Caldwell, & Schartz, 2006), and removing found items from the search array (Cain & Mitroff, 2013), but nothing has been identified that truly eliminates SSM errors when searchers are under any sort of time pressure. Here we demonstrate that splitting a dual-target search into multiple single-target searches effectively abolishes SSMs and provides single-target-search levels of accuracy.

The inspiration for the current intervention came from two sources. First, SSM errors seem to operate at the level of the clinical case or whole trial rather than the level of the individual display. For example, SSM errors occurred in a radiological study where abnormalities were present on separate X-ray films that purportedly belonged to the same patient (Berbaum et al., 1994). What if, rather than combining multiple displays into one trial, we reversed this and split one display into multiple separate trials? Would searchers treat them as separate or related searches, and might this be modulated by searchers' awareness of this manipulation? The second inspiration was the process of luggage screening at U.S. airport checkpoints. When a U.S. Transportation Security Administration officer notices something of interest in a baggage X-ray image, the item is removed and the bag is X-rayed for a second, notionally independent search. Is this technique effective? It is a notably different approach than is used in radiology, for example, where an entire image or series of images is searched and all anomalies noted before other medical action is taken.

Furthermore, dividing search into multiple trials may limit the depletion of cognitive resources. In a previous study (Cain & Mitroff, 2013), we showed that the presence of a found target in a search array disrupted subsequent search. Importantly, it was not the salience of the features of the found target that caused the interference (changing the found target into a color singleton actually lessened its negative impact), but rather that searchers automatically remember the location and/or features of the first target they find. This memory representation persisted even when the target changed into a distractor—after being identified as a target—and was no longer task-relevant. Thus, automatic memory representations of found targets likely reduce available memory resources during subsequent search (Cain & Mitroff, 2013). While this resource depletion appears to degrade performance over the course of a multiple-target trial, single-target performance remains high throughout the course of search experiments. Thus, resources must be "reset" at some point or overall performance would suffer. Perhaps when one trial's search array is replaced with those of the following trials, previously found targets are cleared from memory. Thus, breaking a multiple-target search into several single-target searches could potentially reset memory resources and eliminate resource depletion errors. However, there is a possibility that when a search display reappears, it may cause elements of the search to be recalled (cf. contextual cuing: Chun & Jiang, 1998; and rapid resumption: Lleras, Rensink, & Enns, 2005), lessening any benefits of splitting the search.

Here we explore this intervention and its associated theoretical and applied questions across seven experiments with abstract search displays. Experiments 1–5 involved nonprofessional searchers from Duke University to establish the basic findings, whereas Experiments 6–7 involved professional searchers from the Transportation Security Administration to extend the generality of the findings. To preview the results, splitting multiple-target search into multiple single-target searches completely eliminated SSM errors in both nonprofessional and professional searchers.

General Method

All experiments described here were methodologically similar and the general methods and procedures are provided here. All deviations from this general method are noted in the relevant experimental section.

Participants

For each experiment we aimed to collect data from 12 participants, based on our previous work which used groups of 10 (Fleck et al., 2010) and 15 (Cain & Mitroff, 2013) participants. Participants in Experiments 1–5 were members of the Duke University community and participated for a cash payment or for partial credit toward a course requirement. Participants in Experiments 6 and 7 were Transportation Security Administration (TSA) officers at Raleigh-Durham International Airport (RDU). TSA officers participated in groups of up to six at a time (which led to >12 participants in those experiments). The TSA officers participated in the experiments while at work, and could opt out without their managers being aware of their choice to participate or not (see Biggs, Cain, Clark, Darling, & Mitroff, 2013; Biggs & Mitroff, 2013 for detailed descriptions of airport testing procedures). Each participant took part in only one multiple-target search experiment.

Stimuli and Apparatus

All experiments were programmed in MATLAB using the Psychophysics Toolbox version 3.0.8 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007), and were presented on either Dell Inspiron computers with 20-inch CRT monitors (Experiments 1–5) or Dell Vostro 260 computers and 23.6-inch widescreen LCD monitors (Experiments 6 & 7). Stimuli were adjusted to appear at the same size in each testing environment.

The core stimuli and experiment parameters were based on previous works (e.g., Cain & Mitroff, 2013; Fleck et al., 2010).

Target and distractor stimuli were pairs of rectangles (width of 0.3° of visual angle based upon an approximate viewing distance of 57 cm) contained within the bounds of an invisible 1.3° × 1.3° square (see Figure 1). The members of each pair were oriented perpendicularly to each other and slightly separated. Targets were perfect T shapes and appeared in one of two salience levels (high salience: 57%–65% black; low salience: 22%–45% black). Distractors were non-T shapes drawn from the same salience ranges (5% high-salient distractors). Each display initially contained 25 total items randomly arranged within an invisible 8 × 7 grid, with each item slightly offset spatially from perfect grid alignment. Each item randomly appeared in one of four possible rotations. On initial presentation, one-sixth of displays had a single, high-salience target, one-sixth had a single, low-salience target, one third had both a low-salience and a high-salience target, and the remaining one third of displays had no targets.

Procedure

Participants were instructed to use the computer mouse to click on any T that they found or press the space bar to indicate that no Ts were present. Displays remained on the screen for 15 s or until a response was made. If no response was made within 15 s, the trial ended and a message appeared encouraging participants to respond more quickly. Participants could only respond that no targets were present after at least 3 s had elapsed. This delay was introduced to minimize motor-based miss errors; that is, to prevent the target-absent response from becoming prepotent and executed habitually (Fleck & Mitroff, 2007; Rich et al., 2008).

If participants clicked on an item (either a hit click on a target or a false alarm click on a distractor), the trial would end and the display would be reshown with the clicked item removed either on the next trial (Experiment 1) or three to five trials later (Experiments 2–7). Clicks on empty space had no effect. If participants indicated that no targets were present (either correctly or as a miss error), the display was not reshown. Displays would continue to be re-presented until participants indicated that no targets were present or until the end of the block was reached. Assuming correct performance, a zero-target display would be presented once (no-target response), a single-target display twice (target click; no-target response), and a dual-target display three times (target click; target click; no-target response). This procedure, combined with the distribution of targets across displays, would lead to a target click on half of trials and a no-target-present button press on the other half. In all experiments except Experiments 3 and 6, participants were informed that displays would be reshown. There was a 500-ms intertrial interval (ITI) with a blank white screen.

The first block of each experiment was considered practice and not analyzed. It included 24 trials and feedback about errors was provided. The remaining 10 experimental blocks included 48 trials each and no accuracy feedback was provided. A self-paced break was provided between blocks. No previously viewed displays were carried over from one block to the next (e.g., a display in which a target was identified on Trial 47 of one block would not be reshown on Trial 2–4 of the subsequent block, but rather, fresh displays would be presented).

Data Analysis

In all experiments target present accuracy (hits vs. misses) was the dependent variable of interest. Response time data for all experiments are given in the Appendix. Trials with false alarm responses (i.e., clicks on areas that did not contain a target) were excluded from analysis. Data from three participants were excluded from analysis (one each from Experiments 2, 5, and 6) for committing false alarms on more than 25% of trials (all other participants committed false alarms on less than 20% of trials).

For the calculation of accuracy for low-salience targets on dual-target trials, only trials in which the high-salience target was correctly found first were included in analysis (see Cain et al.,

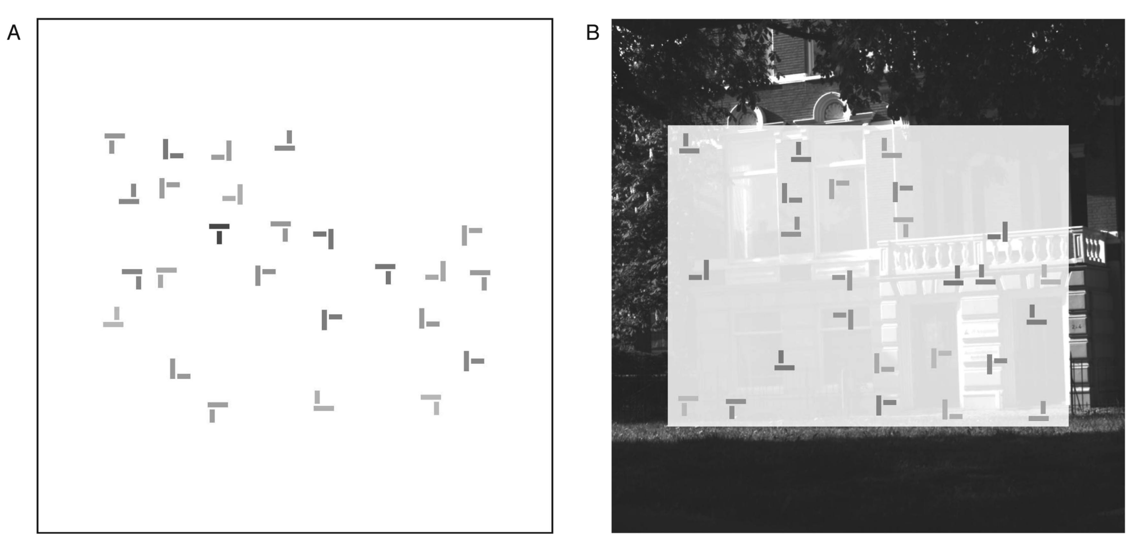

Figure 1. Example displays for (A) Experiments 1 and 2 and (B) Experiments 3–7. Targets were perfectly aligned T shapes, (two present in A, none in B). In Experiments 3–7, background images were unique for each search display and were selected from the van Hateren Natural Image Database (van Hateren & van der Schaaf, 1998).

2013). This included 81.0% of all dual-target trials and was intended to give a conservative estimate of SSM errors (e.g., eliminating errors due to simple inattentiveness).

Statistical Testing

To argue for an absence of SSM errors we need to demonstrate equivalent low-salience target accuracy in single-target and dual-target contexts. Standard null-hypothesis statistical test such as *t* tests are useful for arguing that performance in two experimental conditions differs, but they cannot be used to make the case that performance across two conditions is equivalent. Despite this limitation, *t* tests are presented for each analysis for completeness. High-salience target comparisons are all two-tailed, as there were no a priori predictions about the direction of any effect from previous studies. Low-salience target comparisons are all one-tailed based on strong a priori hypotheses that performance in the dual-target condition would be worse than performance in the single-target condition. All tests were conducted with an alpha-level of 0.05, uncorrected for multiple comparisons. Note that between the large number of tests performed and the one-tailed testing, our null-hypothesis significance tests were biased toward finding differences between conditions (i.e., toward finding SSM errors).

To better argue that low-salience target accuracy in the single-target and dual-target conditions is practically equivalent (i.e., demonstrating elimination of SSM errors), we also performed Bayesian parameter estimation (Kruschke, 2013). This technique uses Markov chain Monte Carlo estimation to determine the distribution of likely means, standard deviations, and effect sizes, given the data. This, in turn, provides ranges of credible values; the sharper the distributions, the more precise an estimate that can be endorsed. We used these distributions to test within-participant differences between single- and dual-target performance against a difference of 0. We present the limits of the highest-density interval (HDI) of the distribution that contains 95% of the estimated mean difference values. If there is no difference between single- and dual-target performance, we would expect 0 to be near the center of this interval. If there is a robust SSM effect, 0 should fall well outside the HDI; the further outside the HDI, the more confidently we can claim to have found a difference.

In addition to estimating the mean difference, we also estimate distributions of effect size of the difference between conditions, which we use for making claims that performance in two conditions is equivalent. Specifically, we define a Region of Practical Equivalence (ROPE) for effect size (Cohen's *d*) of 0 ± 0.2. This is a conservative ROPE for arguing for equivalence, as 0.2 indicates a small effect and SSM effects tend to be much larger than $d = 0.2$ (e.g., $d = 1.08$ in Cain et al., 2013). The proportion of the effect size distribution outside ROPE reflects the confidence with which we can claim that an effect with a size of at least $d = 0.2$ is present. Conversely, if most of the distribution falls within the ROPE, then it is unlikely that there is any meaningful difference between conditions.

Experiment 1—Immediate Repeat

In Experiment 1 we sought to establish a baseline effect of subsequent search misses using a near-replication of Experiment 1 of Cain and Mitroff (2013). In the previous study, participants searched for Ts among pseudo-Ls against a cloudy background, with zero, one, or two targets present on each trial. When a participant clicked on a target (or distractor), it disappeared from the display and search was allowed to continue. This removal of found targets led to a reduction in SSM errors compared to a control condition without removal. However, importantly, a significant quantity of SSM errors remained.

Here, we presented items on a white background (instead of cloudy; as done in Adamo et al., 2013), inserted a 500-ms blank screen after each click (instead of continuous searching), and presented a greater proportion of dual-target trials, but all other parameters remained the same as in Cain and Mitroff (2013, Experiment 1).

Method

Twelve members of the Duke University community participated (mean age = 21.3 years, *SD* = 2.0 years; nine females). Methods were as described in the General Method section, except that when an item was clicked (either a target or a distractor), the trial temporarily ended, and then the display was re-presented after a 500-ms blank ITI with the clicked item no longer present.

Results

Mean accuracy data for all experiments are provided in Table 1 and response time data are in the Appendix. Accuracy for high-

Table 1
Mean Accuracy Data for All Experiments

	High-salience targets					Low-salience targets				
	Single target		Dual target			Single target		Dual target		
Ex.	Mean	*SD*	Mean	*SD*	*SD* of Difference	Mean	*SD*	Mean	*SD*	*SD* of Difference
1	83.59	10.39	88.77	7.55	10.82	59.57	20.23	50.42	23.14	7.46
2	82.52	16.20	87.26	8.27	11.81	66.25	20.46	61.27	17.51	17.32
3	89.46	9.24	92.04	6.20	5.53	57.43	21.00	60.45	16.67	17.51
4	88.76	6.82	93.39	4.23	6.33	58.99	16.76	56.56	18.25	14.74
5	89.97	7.11	91.91	8.43	7.44	62.38	19.65	62.14	17.18	14.45
6	84.17	8.38	87.99	8.62	3.84	57.38	21.29	60.91	16.59	13.84
7	90.93	5.59	95.41	2.82	4.83	64.93	25.02	65.65	20.95	19.43

Note. Accuracy (percent) means and standard deviations (*SD*) for all experiments, broken down by target salience and single- vs. dual-target trial type and *SD* of the differences between single- and dual-target accuracy for both target types.

salience targets averaged 83.59% (*SD* = 10.39%) on single-target trials and 88.77% (*SD* = 7.55%) on dual-target trials. There was no significant difference between these means, $t(11) = 1.648$, $p = .128$. The 95% HDI for the difference in means was −12.1% to 2.51% and the 95% HDI for the effect size of the difference was −1.05 to 0.202, with 21% of the distribution inside the ROPE.

Accuracy for low-salience targets averaged 59.57% (*SD* = 20.23%) on single-target trials and 50.42% (*SD* = 23.14%) on dual-target trials in which the high-salience target was found first. This difference was statistically significant, $t(11) = 3.899$, $p = .001$, and represents a significant SSM error effect for the experiment. The 95% HDI for the difference in means was 3.18% to 12.4% (Figure 2A). The 95% HDI for the effect size of the difference was 0.329 to 2.08, with only 1% of the distribution within the ROPE (Figure 2B).

Discussion

There was a significant amount of SSM errors for low-salience targets, with worse accuracy in dual-target trials than in single-target trials, even with a conservative measure of dual-target accuracy (see General Method). This replicates the results of our previous experiment that removed targets from the display after they were identified (Cain & Mitroff, 2013). It is important to note that this similar result suggests that simply inserting a 500-ms blank period after a target is clicked does not eliminate the SSM effect. This is consistent with the idea of rapid resumption of visual search: Participants are able to resume a visual search interrupted by a blank interval with very little cost (e.g., Lleras et al., 2005).

Unlike low-salience targets, high-salience targets were identified more accurately in dual-target trials than single-target trials. This finding is likely an artifact of the way dual-target trials were analyzed and will be discussed in detail in Experiment 4.

Experiment 2—Delayed Repetition

Adding a blank interval to a visual search in Experiment 1 did not change a dual-target search into two single-target searches, suggesting it was not a meaningful manipulation for reducing SSM errors. Experiment 2 increased the time and visual information between the initial and subsequent presentations; instead of a trial being immediately re-presented one trial later (i.e., after a 500-ms blank interval), it was redisplayed three to five trials later. This setup better mimics the actual dynamics of baggage screening, where several intervening bags might be searched between the inspections of a bag where an object of interest was spotted.

Method

Twelve members of the Duke University community participated (mean age = 20.6 years, *SD* = 1.8 years; eight females). The procedure was identical to Experiment 1, except that displays were repeated three to five trials later, rather than one trial later.

Results

Accuracy for high-salience targets averaged 82.52% (*SD* = 16.20%) on single-target trials and 87.26% (*SD* = 8.27%) on dual-target trials. There was no significant difference between these means, $t(11) = 1.136$, $p = .280$. The 95% HDI for the difference in means was −11.80% to 4.09% and the 95% HDI for the effect size of the difference was −0.906 to 0.310, with 32% of the distribution inside the ROPE.

Accuracy for low-salience targets averaged 66.25% (*SD* = 20.46%) on single-target trials and 61.27% (*SD* = 17.51%) on dual-target trials in which the high-salience target was found first. Unlike Experiment 1, this difference was not statistically significant, $t(11) = 0.575$, $p = .288$. The 95% HDI for the difference in means was −9.52% to 14.2% (Figure 2C). The 95% HDI for the effect size of the difference was −0.465 to 0.748, with 45% of the distribution within the ROPE (Figure 2D).

Discussion

Unlike in Experiment 1, low-salience target accuracy was no worse on dual-target trials than on single-target trials (i.e., there were few to no SSM errors). While there was still a numerical difference between single-target and dual-target trials for low-salience targets, it was not robust and the actual effect size of such a difference is likely to be small. This suggests that searchers were likely treating the dual-target displays as two separate searches rather than as two parts of a single search. Perhaps the separation of the initial and subsequent presentations of each search display eliminated the memory for the location and identity of the previously found target (cf. Cain & Mitroff, 2013), causing searchers to effectively treat the two presentations as independent.

An alternate explanation is that the types of displays employed here are highly confusable, with identical plain-white backgrounds and very similar shapes—perhaps less distinguishable than real baggage or medical radiographs—and participants may not have been able to notice, either explicitly or implicitly, that a trial involved a representation. To address this, the next set of experiments added a unique, memorable background to each display to make them better resemble many real-world searches.

Experiment 3—Memorable Background

Real-world searches rarely occur against a plain white background without any scene context; even relatively sparse baggage and medical searches have potentially distinctive overall bag and body shapes. While the background is not part of the search per se, it may give cues to the searcher about the previous search in that display.

To mimic those sorts of searches, displays in Experiment 3 had photographs of outdoor scenes behind each search array (Figure 1B). Each display had a unique background so that searchers could potentially use this scene context to guide their second search. One possibility is that the background scene might automatically load information about the previous search into working memory, enhancing search. Participants were not explicitly told about the purpose of the background scenes, so any impact of the scenes on performance could possibly be mediated by something akin to contextual cueing (cf. Chun, 2000). Alternatively, it may give searchers the sense that they had already searched the display, leading participants to search less of the display (Hout & Goldinger, 2012) or to strategically terminate search early to maximize their search efficiency rather than their accuracy—motivation, especially to finish quickly, is a problem in laboratory studies which might not be present to the same degree in more conse-

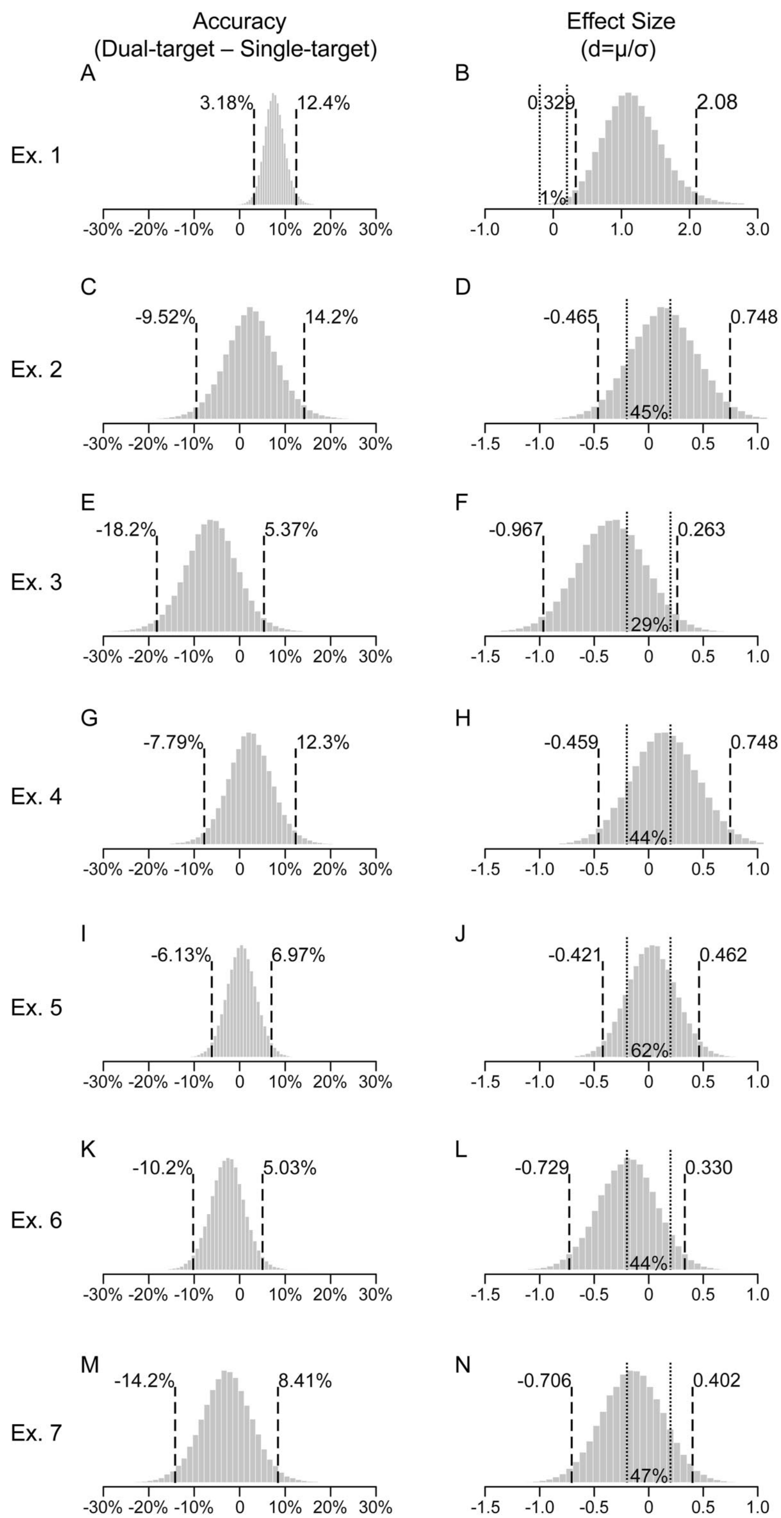

Figure 2. Estimated differences between means of low-salience target accuracy on single-target and dual-target trials (left column; positive numbers mean better performance on single-target trials) and estimated effect sizes (right column) for each experiment. Dashed lines represent the 95% highest-density interval (HDI) of each distribution. Dotted lines represent the region of practical equivalence (ROPE) for effect size of $d \pm 0.2$. Note the scale difference for panel (B).

quential, professional searches (Clark, Cain, Adamo, & Mitroff, 2012).

Method

Twelve members of the Duke University community participated (mean age = 21.4 years, *SD* = 2.8 years; three females). Search arrays were presented on a translucent white rectangle superimposed over an image of an outdoor scene. To make the backgrounds, 176 grayscale images were selected from the van Hateren Natural Image Database (van Hateren & van der Schaaf, 1998), cropped to a 1024 × 1024 pixel square and intensity normalized (Figure 1B). Background images were unique for each search array and were redisplayed whenever the corresponding search array was redisplayed. Participants were instructed that search arrays "will be presented in front of a picture background," and that "there will *not* be a memory test for the backgrounds" [emphasis in original]. Unlike Experiments 1 and 2, there was no mention of displays being repeated or rerun in the instructions. All other procedures were identical to Experiment 2.

Results

Accuracy for high-salience targets averaged 89.46% (*SD* = 9.24%) on single-target trials and 92.04% (*SD* = 6.20%) on dual-target trials. There was no significant difference between these means, $t(11) = 1.649$, $p = .127$. The 95% HDI for the difference in means was −6.40% to 1.19% and the 95% HDI for the effect size of the difference was −1.09 to 0.172, with 20% of the distribution inside the ROPE.

Accuracy for low-salience targets averaged 57.43% (*SD* = 21.00%) on single-target trials and 60.45% (*SD* = 16.67%) on dual-target trials in which the high-salience target was found first; this difference was not statistically significant, $t(11) = 1.336$, $p = .104$. The 95% HDI for the difference in means was −18.20% to 5.37% (see Figure 2E). The 95% HDI for the effect size of the difference was −0.967 to 0.263, with 29% of the distribution within the ROPE (Figure 2F).

Discussion

The results of Experiment 3 mirrored those of Experiment 2; there was no significant difference between single-target trials and dual-target trials for either high-salience targets or for low-salience (i.e., no SSM errors). There is somewhat more evidence for a difference between single-target and dual-target trials for low-salience targets than was seen in Experiment 2, but, if anything, the difference was in the opposite direction from what was predicted, with better performance on dual-target trials.

Thus, adding a distinguishing picture background to each display was not a strong cue to the display, or at least not strong enough to overcome the disruption of several intervening searches. Perhaps the implicit nature of this manipulation was simply too subtle to have an effect. Experiment 4 made the manipulation explicit to searchers.

Experiment 4—Explicit Memorable Background

This experiment was identical to Experiment 3, except that participants were informed about the nature and role of the background scenes. Participants may have been implicitly cued about search display repetitions in Experiment 3, but that may not have been a strong enough manipulation to affect search performance. By allowing for explicit recall of search displays using the background images, we hypothesized that searchers would notice the repetitions and terminate search in repeated displays too early, leading to more motivation- or "satisfaction"-induced SSM errors.

Method

Twelve members of the Duke University community participated (mean age = 21.3 years, *SD* = 2.3 years; nine females). Procedures were identical to those in Experiment 3, except for the addition of one line in the instructions informing participants that after clicking on a target, "a few trials later the same screen will reappear with the 'T' you clicked on removed so you can search the display for any additional 'T's."

Results

Accuracy for high-salience targets averaged 88.76% (*SD* = 6.82%) on single-target trials and 93.39% (*SD* = 4.23%) on dual-target trials, with the accuracy on dual-target trials being significantly greater, $t(11) = 2.491$, $p = .030$. The 95% HDI for the difference in means was −8.44% to 0.205% and the 95% HDI for the effect size of the difference was −1.40 to 0.002, with 7% of the distribution inside the ROPE.

Accuracy for low-salience targets averaged 58.99% (*SD* = 16.76%) on single-target trials and 56.56% (*SD* = 18.25%) on dual-target trials in which the high-salience target was found first. Once again, this difference was not statistically significant, $t(11) = 0.525$, $p = .305$. The 95% HDI for the difference in means was −7.79% to 12.3% (Figure 2G). The 95% HDI for the effect size of the difference was −0.459 to 0.748, with 44% of the distribution within the ROPE (Figure 2H).

Discussion

As in Experiments 1–3, high-salience target accuracy was greater in dual-target trials than in single-target trials. In the present experiment, that difference was statistically significant. This significant difference even persisted when only the first presentations of each display were examined (i.e., no repeated displays; $t(11) = 3.370$, $p = .006$), so it is not a result of participants having an extra chance to find the target on a redisplay. What it may well be is an artifact of the manner in which trials were categorized for analysis. Specifically, on a portion of dual-target displays (approximately 18% of the time), the low-salience target was found in the first presentation. These first presentations were not counted as either hits or misses for high-salience targets—an exclusion that has no counterpart for single-target trials. In this paradigm, the most common strategy is likely that of searching for high-salience targets and then to search for low-salience targets, as evidenced by the high-salience target being found faster than the low-salience target in previous studies (Cain & Mitroff, 2013; e.g., Fleck et al., 2010) and all experiments reported here. With this search strategy, any trial where searchers located the low-salience target first indicates that they missed the high-salience target in the display, at least initially. By removing the trials with the highest likelihood of miss errors for high-salience targets for our current analysis purposes, we have likely artificially inflated the hit rate—in a nonmeaningful way—for high-salience targets on dual-target trials. While this cannot be conclusively demonstrated, it suggests that the apparent dual-target advantage for high-salience targets should be taken lightly.

Most important, there was once again no evidence for SSM errors. Even with explicit instructions, participants seemed to treat the repeating displays on a subsequent trial as separate searches and did not automatically load previously found targets into working memory or strategically terminate the searches. However, it could be that the backgrounds, while more memorable than a white screen, were still too subtle a cue. The next experiment included an explicit marker of repetition trials.

Experiment 5—Marked Repetitions

This experiment built on the design of Experiment 4, but included colored borders to mark which displays were novel and which were repetitions. As before, we hypothesized that the strong cue combination of the colored border and memorable background scene would lead to poorer performance. In particular it would cause automatic recall of the previously found target, interfering with subsequent search or lead to searchers strategically terminating the search too early, or both. Additionally, this experimental design is the closest in the present investigation to an actual baggage security search, as X-ray screeners are often aware when a bag is rescanned and may well remember background characteristics, such as the bag's size and shape.

Method

Participants. This experiment included 23 members of the Duke University community, one of whom was excluded for making false alarm responses on >25% of trials. The 22 participants who remained in the analysis had a mean age of 21.1 year (*SD* = 2.1 year) and included eight males, 13 females, and one person who did not list a gender. This experiment included more participants than the others described here because data from the first group of 11 participants demonstrated a strong, but not statistically significant, trend of better low-salience target performance on dual-target trials than on single-target trials (i.e., an "anti-SSM" effect, not unlike that seen in Experiment 3). This was not predicted. Intrigued, we collected data from another 11 participants to attempt to verify this trend. However, that pattern did not replicate; thus, we collapsed the two datasets.

Stimuli and procedure. The stimuli and procedure were identical to those of Experiment 4, except that every background image had a 0.8°-thick border around the edge. On the first presentation of a given display the border was brown (60% red, 56% green, and 14% blue) and on subsequent presentations it was blue (25% red, 29% green, and 69% blue). Note that this was visually salient, as all other elements of the display were gray. Participants were informed of the meanings of the borders.

Results

Accuracy for high-salience targets averaged 89.97% (*SD* = 7.11%) on single-target trials and 91.91% (*SD* = 8.43%) on dual-target trials, but this difference was not significant, $t(21) = 1.220$, $p = .236$. The 95% HDI for the difference in means was −1.55% to 5.25% and the 95% HDI for the effect size of the difference was −0.196 to 0.702, with 38% of the distribution inside the ROPE.

Accuracy for low-salience targets averaged 62.38% (*SD* = 19.65%) on single-target trials and 62.14% (*SD* = 17.18%) on dual-target trials in which the high-salience target was found first. Once again, this difference was not statistically significant, $t(21) = 0.079$, $p = .469$. The 95% HDI for the difference in means was −6.13% to 6.97% (Figure 2I). The 95% HDI for the effect size of the difference was −0.421 to 0.462, with 62% of the distribution within the ROPE (Figure 2J).

Discussion

As in previous experiments, there was a slight advantage for high-salience target accuracy in dual-task trials compared to single-task trials, but unlike in Experiment 4, it did not reach significance. Again, there was no evidence of SSM errors, as low-salience target accuracy was indistinguishable in single-target and dual-target trials. Taken with the results of Experiments 2–4, this suggests that splitting up multiple-target searches is an effective intervention, even when repeats are marked and cues are given to aid memorability. Repeated search arrays do not appear to automatically reload the features or locations of previously found targets into working memory. Also, there was good task compliance, as participants were not overeager to end their search on repeated displays, alleviating one of the major concerns of translating these results to a professional context (Clark et al., 2012).

Experiment 6—Memorable Background With Professional Searchers

The broader goals of the current study were motivated by professional baggage screening, where typical procedures involve rerunning bags through the X-ray scanner after an item is found and removed. Experiments 1–5 involved nonprofessional searchers, who may differ from professionals who conduct visual searches as part of their job (Clark et al., 2012). In Experiments 6 and 7, we explored these questions with Transportation Security Administration (TSA) officers. Compared to the Duke University community sample, these individuals were older, had a more variable level of education, and were performing a work-like task during the course of their employment (see Biggs et al., 2013 for a full discussion of population differences).

Method

This experiment was identical to Experiment 3 (i.e., memorable background, no information about the purpose of the background, and no colored border), except with professional visual searchers as participants. As in Experiment 3, the participants were not informed of the repeated-display manipulation. The participants were 17 TSA Officers tested at Raleigh-Durham International Airport. One participant's data were excluded from analysis because of false alarms on >25% of trials. The 16 participants remaining in the analysis included 10 males and six females and had a mean age of approximately 46 years (a categorical survey with age ranges was administered). See Biggs et al. (2013) for details about the participation and recruitment process for the TSA Officers.

Results

Accuracy for high-salience targets averaged 84.17% (*SD* = 8.38%) on single-target trials and 87.99% (*SD* = 8.62%) on dual-target trials, with the accuracy on dual-target trials being significantly greater, $t(15) = 4.322$, $p = .001$. The 95% HDI for the difference in means was −6.23% to −2.01% and the 95% HDI for the effect size of the difference was −1.74 to −0.393, with 0% of the distribution inside the ROPE.

Accuracy for low-salience targets averaged 57.38% (*SD* = 21.29%) on single-target trials and 60.91% (*SD* = 16.59%) on dual-target trials in which the high-salience target was found first. As with nonprofessional searchers, this difference was not statistically significant, $t(15) = 0.767$, $p = .228$. The 95% HDI for the difference in means was −10.20% to 5.03% (Figure 2K). The 95% HDI for the effect size of the difference was −0.729 to 0.330, with 44% of the distribution within the ROPE (Figure 2L).

Discussion

This experiment with professional searchers produced the same pattern of results seen with nonprofessional searchers in Experiment 3. Searchers were better at detecting high-salience targets in dual-target trials than in single-target trials, though this has the potential to be an artifact, as described in the discussion of Experiment 4. There was again no difference in low-salience performance between single-target and dual-target trials (i.e., no evidence of SSM errors). If anything, performance was better on dual-target trials, in line with the high-salience target performance, but opposite what was predicted. This stands in contrast to previous work that has shown a robust SSM effect in TSA personnel on more typical dual-target searches (Biggs & Mitroff, 2013). While these results suggest that separating dual-target trials effectively turns them into single-target trials and eliminates SSM errors—and may even do so in a professional search context—it should be noted that these are still abstract stimuli and not baggage images, so further work is needed to demonstrate full generality.

Experiment 7—Explicit Memorable Background With Professional Searchers

In an actual baggage security screening environment, or other similar professional search environments, searchers would typically be aware that revisiting previous search displays could occur. To better mimic this typical search environment, we administered the methods of Experiment 4 (i.e., memorable background, instructions about the background, no colored border) to professional searchers.

Method

This experiment is an exact replication of Experiment 4, except with professional searchers rather than nonprofessional searchers. The participants were 13 TSA Officers at Raleigh Durham Airport. This group contained nine males and four females and had a mean age of approximately 46 years.

Results

Accuracy for high-salience targets averaged 90.93% (*SD* = 5.59%) on single-target trials and 95.41% (*SD* = 2.82%) on dual-target trials, with the accuracy on dual-target trials being significantly greater, $t(12) = 3.261$, $p = .007$. The 95% HDI for the difference in means was −6.70% to −0.703% and the 95% HDI for the effect size of the difference was −1.38 to −0.12, with 4% of the distribution inside the ROPE.

Accuracy for low-salience targets averaged 64.93% (*SD* = 25.02%) on single-target trials and 65.65% (*SD* = 20.95%) on dual-target trials in which the high-salience target was found first. As with nonprofessional searchers, this difference was not statistically significant, $t(12) = 0.594$, $p = .282$. The 95% HDI for the difference in means was −14.20% to 8.41% (Figure 2M). The 95% HDI for the effect size of the difference was −0.706 to 0.402, with 47% of the distribution within the ROPE (Figure 2N).

Discussion

As in Experiment 6, we observed the same pattern in professional searchers as we did in nonprofessional searchers. Searchers were significantly better at detecting high-salience targets in dual-target displays than in single-target displays, and were no worse at detecting low-salience targets in dual-target displays than in single-target trials. Again, there was no evidence for SSM errors.

Combined Analyses

The results of Experiments 2–7 are quite similar and present a good argument for dividing searches as an effective way to eliminate SSM errors. However, one potential problem when making the case for a null finding is a lack of statistical power. Here, we combine the data from Experiments 2–5 and 6–7 into two larger analyses, where we can increase the power through larger population sizes. These analyses are obviously not independent from the original results, but do offer additional insight.

Nonprofessional Searcher Studies

Combining Experiments 2–5 resulted in a total of 58 participants. Results were analyzed first with a pair of 2 × 4 mixed-model ANOVAs with number of targets (single- or dual-target) as a within-participants factor and Experiment (2, 3, 4, or 5) as a between-participants factor. Accuracy for high-salience targets averaged 88.4% (*SD* = 10.0%) on single-target trials and 91.4% (*SD* = 7.4%) on dual-target trials. There was a significant main effect of number of targets ($F(1, 54) = 8.872$, $p = .004$, $\eta_p^2 = 0.141$), with the accuracy on dual-target trials being significantly greater, but no significant main effect of Experiment ($F(3, 54) = 1.548$, $p = .213$, $\eta_p^2 = 0.079$). Accuracy for low-salience targets averaged 61.2% (*SD* = 19.3%) on single-target trials and 61.4% (*SD* = 17.1%) on dual-target trials in which the high-salience target was found first. There was no significant main effect of number of targets ($F(3, 54) = 1.548$, $p = .213$, $\eta_p^2 = 0.079$) or of Experiment ($F(3, 54) = 0.331$, $p = .803$, $\eta_p^2 = 0.018$). For both analyses, there was no significant interaction between factors ($ps > 0.4$).

For parameter estimation, the larger number of participants allows for much more precise estimation of likely value distributions. For high-salience target accuracy, the 95% HDI for the difference in means was −3.61% to 0.76% and the 95% HDI for the effect size of the difference was −0.454 to 0.0959, with 54% of the distribution inside the ROPE, suggesting a small-to-medium sized effect in the direction of superior performance on dual-target trials. For low-salience targets, the 95% HDI for the difference in means was −4.22% to 4.14% (Figure 3A). The 95% HDI for the effect size of the difference was −0.277 to 0.261, with 85% of the distribution within the ROPE (Figure 3B), suggesting that most plausible effect sizes are small and as likely to favor dual-target trial performance as single-target trial performance.

Professional Searcher Studies

Combining Experiments 6–7 resulted in a total of 29 participants. Results were analyzed first with a pair of 2 × 2 mixed-model ANOVAs with number of targets (single- or dual-target) as a within-participants factor and Experiment (6 or 7) as a between-participants factor. Accuracy for high-salience targets averaged 87.36% (SD = 7.99%) on single-target trials and 91.61% (SD = 7.54%) on dual-target trials. There was a significant main effect of number of targets ($F(1, 27) = 28.017$, $p < .001$, $\eta_p^2 = 0.509$), with the accuracy on dual-target trials being significantly greater, and a significant main effect of Experiment ($F(1, 27) = 8.345$, $p = .008$, $\eta_p^2 = 0.236$), with better performance in Experiment 7. Accuracy for low-salience targets averaged 60.48% (SD = 22.78%) on single-target trials and 63.38% (SD = 18.58%) on dual-target trials in which the high-salience target was found first. There was no significant main effect of number of targets ($F(1, 27) = 0.896$, $p = .352$, $\eta_p^2 = 0.032$) or of Experiment ($F(1, 27) = 0.648$, $p = .428$, $\eta_p^2 = 0.023$). For both analyses, there was no significant interaction between factors ($ps > 0.8$).

For parameter estimation, the larger number of participants again allows for more precise estimation of likely value distributions. For high-salience target accuracy, the 95% HDI for the difference in means was −5.57% to −2.26% and the 95% HDI for the effect size of the difference was −1.38 to −0.477, with 0% of the distribution inside the ROPE, suggesting a medium-to-very-large sized effect in the direction of superior performance on dual-target trials. For low-salience targets, the 95% HDI for the difference in means was −8.91% to 3.29% (Figure 3C). The 95% HDI for the effect size of the difference was −0.56 to 0.193, with 53% of the distribution within the ROPE (Figure 3D), suggesting that most plausible effect sizes are small-to-medium and slightly more likely to favor dual-target trial performance than single-target trial performance.

Discussion

These combined analyses tell essentially the same story that the individual analyses told. Removing a search display after one target is found and redisplaying it three to five trials later, effectively turns it into two single-target trials, abolishes subsequent search miss errors and even boosts high-salience target performance. The absence of SSM effects in the individual experiments was not due to a lack of power, as pooling data across experiments revealed no hint of SSM errors, and the more precisely estimated parameter distributions suggested even smaller potential effect sizes than were endorsable for the individual experiments.

General Discussion

In multiple-target visual search, once one target is found, additional targets are less likely to be found than if they were the only targets present in the display. This phenomenon of subsequent search misses has been a known problem in professional searches for over half a century (e.g., Tuddenham, 1962), but has resisted efforts at explanation and abatement. Here we demonstrated that breaking a multiple-target search into several single-target searches effectively eliminated SSM errors. Note that miss errors still occurred, but they were no more likely in a dual-target context than in a single target context, suggesting that it is SSM errors in

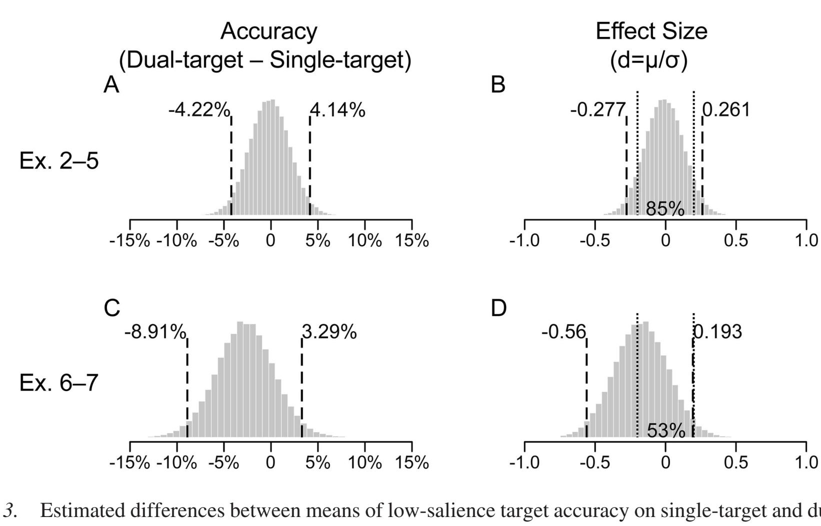

Figure 3. Estimated differences between means of low-salience target accuracy on single-target and dual-target trials (left column; positive numbers mean better performance on single-target trials) and estimated effect sizes (right column) for data combined across experiments. Dashed lines represent the 95% highest-density interval (HDI) of each distribution. Dotted lines represent the region of practical equivalence (ROPE) for effect size of $d \pm 0.2$.

particular that were affected by this technique. This elimination of SSMs was true even when searchers were aware of display repetitions and were given explicit cues about which trials were repeated and memorable backgrounds to individuate displays. It is important that this manipulation improved performance for professional searchers as well as nonprofessional searchers, suggesting that it might have broad applicability in real-world searches.

Potential Alternative Explanations

Lack of statistical power. One potential problem with arguing for a null hypothesis of no difference between conditions is that subtle effects are often difficult to discern with few participants. However, SSM errors have not tended to be subtle and often have large effect sizes (e.g., Cain et al., 2013) and are easily found with 10 participants (e.g., Fleck et al., 2010). Here, we had at least 12 participants in each experiment and found no evidence of SSM errors using 1-tailed t tests for Experiments 2–7 individually or when combined into larger analyses with 58 and 29 total participants.

Additionally, we provide Bayesian estimations of the difference in means between conditions and the effect size of these differences. From the combined analysis of Experiments 2–5, over 90% of the distribution of credible values of d was less than 0.2 (Figure 3B). This suggests that even if an SSM effect does still exist with this manipulation, it is dramatically smaller than those observed previously.

Contextual cuing. Contextual cuing is an effect where search speed improves when search displays are repeated, compared with novel displays (e.g., Chun & Jiang, 1998; Chun, 2000). While the current paradigm involves repeated search arrays, contextual cuing is unlikely to be driving our effects. First, displays were shown on a maximum of three trials (i.e., the trial on which the first target is found, the trial on which the second target is found, and the trial on which it is indicated that no targets are present) and the crucial presentation for our purposes is the second one, whereas contextual cuing usually requires three to five exposures (e.g., Chun & Jiang, 2003), even with natural scene backgrounds (Brockmole & Henderson, 2006). More importantly, contextual cuing primarily speeds the finding of targets in their previous locations, and has not been found to speed search of the whole array for stimuli such as Ts and Ls (Chun & Jiang, 1998). With more informative distractors, such as photographs of objects, display repetition has been shown to speed search across the whole array, but this effect is driven most strongly by target-absent trials and likely occurs on a slower timescale than the second presentation after a target-present display (Hout & Goldinger, 2010). Here, once a target is found it is never redisplayed and participants must search the rest of the array, which should not be aided by the repetition. It is possible that searchers may have been able to use the repeated background to avoid some re-searching of the area where the first target was located (e.g., Peterson, Boot, Kramer, & McCarley, 2004), or recall some of previous search path, but the resource depletion theory of SSMs (Berbaum et al., 1991; Cain & Mitroff, 2013) would predict that this recall would overall hinder, rather than help, performance. Future studies with methods such as eye tracking may be able to settle this question directly.

Rapid resumption. Rapid resumption is the idea that interrupted searches can be restarted more quickly than a new search can be initiated (e.g., Lleras et al., 2005). This effect may have contributed to the SSM errors seen in Experiment 1: searchers likely considered the 500-ms blank interval after clicking a target to be a break in the current search rather than the start of a new search. Rapid resumption has been demonstrated across blank intervals of over 3 s (Lleras et al., 2005) and with large changes to the search array (Lleras, Rensink, & Enns, 2007), but no research has looked for resumption across three to five interfering searches of many seconds each.

Additionally, the reported effects of rapid resumption are an improvement in the first 500 ms of resumed search. This is a much different timescale than the current search, where the mean response time for second displays of search arrays was 7.3 s. Thus, it is unlikely that rapid resumption is driving the main effect here, but could have had some benefit on search that was too subtle to detect in the current paradigm, which was more focused on accuracy than response time.

Implications

Given the results with both nonprofessional and professional searchers, the intervention of dividing up multiple-target searches shows promise for generalizability. One caveat is that the stimuli used here were abstract Ts and pseudo Ls and bear little resemblance to the stimuli in most real-world searches. This allowed us to use the same task with both the nonprofessional and professional groups, but may have introduced other complications. In particular, these stimuli are quite difficult to remember ($k = 1.3$; Cain et al., 2012, supplementary experiment) and real scenes with real objects may be easier to remember and, thus, be more likely to reactivate memory of previous search. Additionally, while there were memorable backgrounds in Experiments 3–7, they did not actually provide structure to the search arrays. Real scenes and radiographs may have more meaningful structure that could be more easily memorized by expert searchers (e.g., chess experts memorize valid chess boards more easily than random chess boards; Chase & Simon, 1973). Thus, we advise verifying that this effect holds within a given search domain before enacting a policy change within that domain.

Implications for visual search theory. The finding that people can treat a multiple-target search that is interrupted as two separate searches has a few interesting implications for theories of search. First, it demonstrates that the cognitive resources posited by resource depletion theory to be used to represent a found target (Cain & Mitroff, 2013) are freed after the trial is over. For this to occur, a half-second blank interval is not a sufficient break in search to clear working memory resources, whereas a period including three to five similar searches was sufficient. The current paradigm does not allow for a more fine-grained analysis of when the contents of memory are released, but perhaps it may be with the presentation of a new search array, similar to the way a previous task set is likely released upon viewing a stimulus that prompts a new task set (e.g., Yeung, Nystrom, Aronson, & Cohen, 2006).

Additionally, this suggests that the working memory contents of one trial may be present at the beginning of the subsequent trial. This may lead to effects such as priming of target features (e.g., Maljkovic & Nakayama, 1994) or locations (e.g., Maljkovic & Nakayama, 1996), or more complex guidance of search (e.g., Dowd & Mitroff, 2013). While these trial history effects have been studied separately, the current findings raise the question of what impact they may have on typical search tasks.

Implications for applied visual search. The present finding suggests that the procedures used in airport luggage screening of rescanning bags after suspicious items are removed are likely effective in combating SSM errors. Further, it cautions that bags with threat images projected on them for assessment purposes—a common practice known as "Threat Image Projection" in airport screening (e.g., Cutler & Paddock, 2009)—should be rerun to reduce the risk of inducing SSM errors. More broadly, it suggests that taking such an approach with other applied searches, such as medical X-ray screenings may prove fruitful. While there are potential complications to splitting up such real-world examinations (e.g., reviewing the medical history for a patient may be a far more potent mnemonic cue than any tested here), it speaks toward the efficacy of screening techniques such as double-reading (e.g., Dinnes et al., 2001), where the second search is performed by a different searcher than the first. It is also in line with previous suggestions for searchers to conduct separate searches for different categories of items (e.g., search first for guns then search for bombs) rather than search for all categories simultaneously (Menneer et al., 2012; Menneer, Barrett, Phillips, Donnelly, & Cave, 2004).

References

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2013). Self-induced attentional blink: A cause of errors in multiple-target search. *Psychological Science, 24,* 2569–2574. doi:10.1177/0956797613497970

Berbaum, K. S., El-Khoury, G. Y., Franken, E. A., Kuehn, D. M., Meis, D. M., Dorfman, D. D., . . . Kathol, M. H. (1994). Missed fractures resulting from satisfaction of search effect. *Emergency Radiology, 1,* 242–249. doi:10.1007/BF02614935

Berbaum, K. S., Franken, E. A., Jr., Caldwell, R. T., & Schartz, K. M. (2006). Can a checklist reduce SOS errors in chest radiography? *Academic Radiology, 13,* 296–304. doi:10.1016/j.acra.2005.11.032

Berbaum, K. S., Franken, E. A., Jr., Caldwell, R. T., & Schartz, K. M. (2010). Satisfaction of search in traditional radiographic imaging. In E. Samei & E. Krupinski (Eds.), *The handbook of medical image perception and techniques* (pp. 107–138). New York, NY: Cambridge University Press.

Berbaum, K. S., Franken, E. A., Jr., Dorfman, D. D., Caldwell, R. T., & Lu, C. H. (2005). Can order of report prevent satisfaction of search in abdominal contrast studies? *Academic Radiology, 12,* 74–84. doi:10.1016/j.acra.2004.11.007

Berbaum, K. S., Franken, E. A., Jr., Dorfman, D. D., Miller, E. M., Krupinski, E. A., Kreinbring, K., . . . Lu, C. H. (1996). Cause of satisfaction of search effects in contrast studies of the abdomen. *Academic Radiology, 3,* 815–826. doi:10.1016/S1076–6332(96)80271-6

Berbaum, K. S., Franken, E. A., Jr., Dorfman, D. D., Rooholamini, S. A., Coffman, C. E., Cornell, S. H., . . . Kao, S. C. (1991). Time course of satisfaction of search. *Investigative Radiology, 26,* 640–648. doi:10.1097/00004424–199107000-00003

Berbaum, K. S., Schartz, K. M., Caldwell, R. T., El-Khoury, G. Y., Ohashi, K., Madsen, M., & Franken, Jr., E. A. (2012). Satisfaction of search errors detecting subtle fractures diminish in the presence of more serious injuries. *Proceedings of SPIE, 8318,* 83180L–83180L–12. doi:10.1117/12.913630

Berbaum, K. S., Schartz, K. M., Caldwell, R. T., Madsen, M. T., Thompson, B. H., Mullan, B. F., . . . Franken Jr., E. A. (2013). Satisfaction of search from detection of pulmonary nodules in computed tomography of the chest. *Academic Radiology, 20,* 194–201. doi:10.1016/j.acra.2012.08.017

Biggs, A. T., Cain, M. S., Clark, K., Darling, E. F., & Mitroff, S. R. (2013). Assessing visual search performance differences between Transportation Security Administration Officers and nonprofessional visual searchers. *Visual Cognition, 21,* 330–352. doi:10.1080/13506285.2013.790329

Biggs, A. T., & Mitroff, S. R. (2013). Different predictors of multiple-target search accuracy between non-professional and professional visual searchers. *Quarterly Journal of Experimental Psychology, Advance online publication*. doi:10.1080/17470218.2013.859715

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436. doi:10.1163/156856897X00357

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition, 13,* 99–108. doi:10.1080/13506280500165188

Cain, M. S., Adamo, S. H., & Mitroff, S. R. (2013). A taxonomy of errors in multiple-target visual search. *Visual Cognition, 21,* 899–921. doi:10.1080/13506285.2013.843627

Cain, M. S., Dunsmoor, J. E., LaBar, K. S., & Mitroff, S. R. (2011). Anticipatory anxiety hinders detection of a second target in dual-target search. *Psychological Science, 22,* 866–871. doi:10.1177/0956797611412393

Cain, M. S., & Mitroff, S. R. (2013). Memory for found targets interferes with subsequent performance in multiple-target visual search. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 1398–1408. doi:10.1037/a0030726

Cain, M. S., Vul, E., Clark, K., & Mitroff, S. R. (2012). A Bayesian optimal foraging model of human visual search. *Psychological Science, 23,* 1047–1054. doi:10.1177/0956797612440460

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55–81. doi:10.1016/0010–0285(73)90004-2

Chun, M. M. (2000). Contextual cueing of visual attention. *Trends in Cognitive Sciences, 4,* 170–178. doi:10.1016/S1364–6613(00)01476-5

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36,* 28–71. doi:10.1006/cogp.1998.0681

Chun, M. M., & Jiang, Y. (2003). Implicit, long-term spatial contextual memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 224–234. doi:10.1037/0278–7393.29.2.224

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. doi:10.1006/cogp.1996.0002

Clark, K., Cain, M. S., Adamo, S. H., & Mitroff, S. R. (2012). Overcoming hurdles in translating visual search research between the lab and the field. In M. D. Dodd & J. H. Flowers (Eds.), *The influence of attention, learning, and motivation on visual search* (Vol. 59, pp. 147–181). New York, NY: Springer. doi:10.1007/978-1-4614–4794-8_7

Clark, K., Cain, M. S., Adcock, R. A., & Mitroff, S. R. (2014). Context matters: The structure of task goals affects accuracy in multiple-target visual search. *Applied Ergonomics, 45,* 528–533. doi:10.1016/j.apergo.2013.07.008

Cutler, V., & Paddock, S. (2009). Use of threat image projection (TIP) to enhance security performance. In *43rd Annual International Carnahan Conference on Security Technology,* 2009 (pp. 46–51). IEEE. doi:10.1109/CCST.2009.5335565

Dinnes, J., Moss, S., Melia, J., Blanks, R., Song, F., & Kleijnen, J. (2001). Effectiveness and cost-effectiveness of double reading of mammograms in breast cancer screening: Findings of a systematic review. *The Breast, 10,* 455–463. doi:10.1054/brst.2001.0350

Dowd, E. W., & Mitroff, S. R. (2013). Attentional guidance by working memory overrides salience cues in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 39,* 1786–1796. doi:10.1037/a0032548

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision, 11*(5), article 14. doi:10.1167/11.5.14

Fleck, M. S., & Mitroff, S. R. (2007). Rare targets are rarely missed in correctable search. *Psychological Science, 18,* 943–947. doi:10.1111/j.1467–9280.2007.02006.x

Fleck, M. S., Samei, E., & Mitroff, S. R. (2010). Generalized “satisfaction of search”: Adverse influences on dual-target search accuracy. *Journal of Experimental Psychology: Applied, 16,* 60–71. doi:10.1037/a0018629

Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception, & Psychophysics, 72,* 1267–1282. doi:10.3758/APP.72.5.1267

Hout, M. C., & Goldinger, S. D. (2012). Incidental learning speeds visual search by lowering response thresholds, not by improving efficiency: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 90–112. doi:10.1037/a0023894

Kleiner, M., Brainard, D. H., & Pelli, D. (2007). What’s new in Psychtoolbox-3? *Perception, 36,* ECVP Abstract Suppl.

Kruschke, J. K. (2013). Bayesian estimation supersedes the *t* test. *Journal of Experimental Psychology: General, 142,* 573–603. doi:10.1037/a0029146

Lleras, A., Rensink, R. A., & Enns, J. T. (2005). Rapid resumption of interrupted visual search: New insights on the interaction between vision and memory. *Psychological Science, 16,* 684–688. doi:10.1111/j.1467–9280.2005.01596.x

Lleras, A., Rensink, R. A., & Enns, J. T. (2007). Consequences of display changes during interrupted visual search: Rapid resumption is target specific. *Perception & Psychophysics, 69,* 980–993. doi:10.3758/BF03193936

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition, 22,* 657–672. doi:10.3758/BF03209251

Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. The role of position. *Perception & Psychophysics, 58,* 977–991. doi:10.3758/BF03206826

Menneer, T., Barrett, D. J. K., Phillips, L., Donnelly, N., & Cave, K. R. (2004). Search efficiency for multiple targets. *Cognitive Technology, 9,* 22–25.

Menneer, T., Stroud, M. J., Cave, K. R., Li, X., Godwin, H. J., Liversedge, S. P., & Donnelly, N. (2012). Search for two categories of target produces fewer fixations to target-color items. *Journal of Experimental Psychology: Applied, 18,* 404–418. doi:10.1037/a0031032

Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research, 51,* 1526–1537. doi:10.1016/j.visres.2010.09.003

Peterson, M. S., Boot, W. R., Kramer, A. F., & McCarley, J. S. (2004). Landmarks help guide attention during visual search. *Spatial Vision, 17,* 497–510. doi:10.1163/1568568041920230

Rich, A. N., Kunar, M. A., Van Wert, M. J., Hidalgo-Sotelo, B., Horowitz, T. S., & Wolfe, J. M. (2008). Why do we miss rare targets? Exploring the boundaries of the low prevalence effect. *Journal of Vision, 8*(15), article 15. doi:10.1167/8.15.15

Smith, M. J. (1967). *Error and variation in diagnostic radiology*. Springfield, IL: C. C. Thomas.

Tuddenham, W. J. (1962). Visual search, image organization, and reader error in roentgen diagnosis. Studies of the psycho-physiology of roentgen image perception. *Radiology, 78,* 694–704. doi:10.1148/78.5.694

van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society B: Biological Sciences, 265,* 359–366. doi:10.1098/rspb.1998.0303

Wolfe, J. M. (2007). Guided search 4.0: Current progress with a model of visual search. In W. Gray (Ed.), *Integrated models of cognitive systems* (pp. 99–119). New York, NY: Oxford.

Wolfe, J. M. (2012). When do I quit? The search termination problem in visual search. In M. D. Dodd & J. H. Flowers (Eds.), *The influence of attention, learning, and motivation on visual search* (Vol. 59, pp. 183–208). New York, NY: Springer. doi:10.1007/978-1-4614–4794-8_8

Wolfe, J. M. (2013). When is it time to move to the next raspberry bush? Foraging rules in human visual search. *Journal of Vision, 13*(3), article 10. doi:10.1167/13.3.10

Yeung, N., Nystrom, L. E., Aronson, J. A., & Cohen, J. D. (2006). Between-task competition and cognitive control in task switching. *The Journal of Neuroscience, 26,* 1429–1438. doi:10.1523/JNEUROSCI.3109-05.2006

(*Appendix follows*)

Appendix

Response Times

	High-salience targets				Low-salience targets			
	Single target		Dual target		Single target		Dual target	
Ex.	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
1	3.91	1.34	3.66	1.05	5.25	1.24	4.66	1.22
2	3.12	0.62	2.94	0.64	4.87	0.95	4.80	0.75
3	3.38	1.58	3.06	1.48	5.16	1.51	5.18	1.37
4	3.01	1.04	3.04	0.96	5.48	1.24	5.19	0.80
5	3.00	1.01	2.71	0.84	5.31	1.22	5.04	1.03
6	4.77	1.82	4.45	1.67	7.35	1.54	7.04	1.28
7	4.60	1.13	4.20	0.97	7.09	1.60	6.83	0.93

Note. Mean and standard deviations (*SD*) of response times (seconds) to correctly click on each target type in each display type.

High-Salience Targets

We conducted two repeated-measures ANOVAs with number of targets and experiment as variables. One test included data from the undergraduate participants (Experiments 2–5) and the other included data from the professional searchers (Experiments 6 and 7). There was a main effect of number of targets, as high-salience targets were identified more quickly overall in dual-target trials [Duke studies: $F(1, 54) = 10.149$, $p = .002$, $\eta_p^2 = 0.158$; RDU studies: $F(1, 27) = 7.235$, $p = .012$, $\eta_p^2 = 0.211$]. There was no main effect of experiment [Duke studies: $F(3, 54) = 0.339$, $p = .797$, $\eta_p^2 = 0.019$; RDU studies: $F(1, 27) = 0.149$, $p = .702$, $\eta_p^2 = 0.006$] and no interaction between factors [Duke studies: $F(3, 54) = 1.717$, $p = .174$, $\eta_p^2 = 0.087$; RDU studies: $F(1, 27) = 0.079$, $p = .781$, $\eta_p^2 = 0.001$].

Low-Salience Targets

For low-salience targets we conducted the same pair of repeated-measures ANOVAs as for high-salience targets. There were neither a main effects of number of targets [Duke studies: $F(1, 54) = 0.815$, $p = .371$, $\eta_p^2 = 0.015$; RDU studies: $F(1, 27) = 1.383$, $p = .250$, $\eta_p^2 = 0.049$] nor of experiment [Duke studies: $F(3, 54) = 0.606$, $p = .614$, $\eta_p^2 = 0.033$; RDU studies: $F(1, 27) = 0.272$, $p = .607$, $\eta_p^2 = 0.010$]. There were no interactions between factors [Duke studies: $F(3, 54) = 0.204$, $p = .893$, $\eta_p^2 = 0.011$; RDU studies: $F(1, 27) = 0.012$, $p = .914$, $\eta_p^2 < 0.001$].

Received October 17, 2013
Revision received December 26, 2013
Accepted December 26, 2013 ■