Attention, Perception, & Psychophysics
https://doi.org/10.3758/s13414-019-01802-8

How to correctly put the “subsequent” in subsequent search miss errors

Stephen H. Adamo[1] · **Patrick H. Cox**[2] · **Dwight J. Kravitz**[2] · **Stephen R. Mitroff**[2]

Abstract

Visual search, finding targets among distractors, is theoretically interesting and practically important as it involves many cognitive abilities and is vital for several critical industries (e.g., radiology, baggage screening). Unfortunately, search is especially error prone when more than one target is present in a display (a phenomenon termed the satisfaction of search effect or the subsequent search miss effect). The general effect is that observers are more likely to miss a second target if a first was already detected. Unpacking the underlying mechanisms requires two key aspects in analysis and design. First, to speak to the “subsequent” nature of the effect, the analyses must compare performance on single-target trials to performance for a second target in dual-target displays after a first has been found. Second, the design must include single-target displays that are matched in difficulty to each dual-target display to enable fair comparisons. However, it is not clear that prior research has met these two standards simultaneously. Work from academic radiology has primarily used designs with well-matched single- and dual-target trials, but most employed analyses that do not focus solely on performance *after* a first target has been detected. Work from cognitive psychology has generally performed the correct analyses, but relied on unmatched single- and dual-target trials, introducing a confound that could distort the results. In the current paper, we demonstrate the impact of this confound in empirical data and provide a roadmap for proper study design and analyses.

Keywords Visual search

Introduction

Visual search, finding targets among distractors, involves a number of cognitive processes (e.g., perception, attention, decision-making), making it an important area of academic study (for reviews, see Eckstein, 2011; Chan & Hayward, 2013; Nakayama & Martini, 2011). Visual search is also fundamental for many critical professions (e.g., radiology, aviation security, military activities), making it an important skill to understand and potentially improve (e.g., Berbaum, 2012; Biggs, Kramer, & Mitroff, 2018; Wetter, 2013). While there are many factors that can affect search performance, the presence of multiple targets in the same search display is one key factor that consistently hinders successful search amongst both novices and professional searchers (e.g., Adamo, Cain, & Mitroff, 2013; Berbaum et al., 1990, 1991, 1998; Berbaum, 2012; Fleck, Samei, & Mitroff, 2010).

The negative effect of multiple targets on visual search performance was originally referred to as the “satisfaction of search” effect (Smith, 1967), but has more recently been referred to as the “subsequent search miss” effect (SSM; Adamo et al., 2013). Research in both academic radiology (e.g., Ashman, Yu, & Wolfman, 2000; Berbaum et al., 1990, 1991, 1993, 1994; Franken et al., 1994; Samuel et al., 1995) and cognitive psychology (e.g., Adamo, Cain, & Mitroff, 2013, 2015, 2017, 2018; Biggs, Adamo, Dowd, & Mitroff, 2015; Cain, Adamo, & Mitroff, 2013; Cain, Dunsmoor, LaBar, & Mitroff, 2012; Cain & Mitroff, 2013; Fleck et al., 2010; Stothart, Clement, & Brockmole, 2018) has looked to identify the mechanisms of SSM errors to best understand this source of error. Broadly, there are three accounts that have received

---

The original version of this article was revised: The following formatting changes to the figures and table need to be made in order to enhance readability:

1) Delete bounding boxes around the figures
2) The table oriented horizontally
3) The 1st and 3rd lines of the table are headers, the 2nd and 4th are the data. Consistent bolding of font and lines in the table to reflect that.

Stephen H. Adamo and Patrick H. Cox are co-first authors

✉ Stephen H. Adamo
stephen.adamo@ucf.edu

[1] University of Central Florida, Orlando, FL, USA

[2] The George Washington University, Washington, DC, USA

empirical support. First, the "satisfaction" account proposes that observers become "satisfied" with the nature of the evaluation after finding a first target and prematurely terminate their search, causing them to miss additional targets (Smith, 1967; Tuddenham, 1962). Second, the "perceptual set" account suggests that after finding a first target, observers become biased to look for targets that are similar to the just found first target, which consequently makes them more prone to miss dissimilar targets (Biggs et al., 2015; Berbaum, 1990). Finally, the "resource depletion" account (Berbaum et al., 1991; Cain & Mitroff, 2013) suggests that when a first target is found, observers' attention and working memory resources are allocated to the processing of that first target leaving reduced resources available to detect an additional target(s). There is evidence supporting each of these three accounts, with data coming from behavioral (e.g., Adamo et al., 2018; Biggs et al., 2015; Cain & Mitroff, 2013), eye-tracking (e.g., Adamo et al., 2013; Cain et al., 2013; Samuel, Kundel, Nodine & Toto, 1995;), and physiological measures (e.g., Cain et al., 2012).

However, it appears that the research to date has fallen short of being able to directly address the mechanistic accounts of SSM errors. Testing these accounts requires two key factors. First, to test whether the detection of a first target affects the detection of a second target, the analysis must be conditioned on whether the first target was actually detected *before* the second; this is the "subsequent" component of the phenomenon. Second, performance on a second target must be compared to performance on a matched single-target display, where the single- and multiple-target displays are as identical as possible, with the only difference being the presence of an additional target in the multiple-target display. Comparing unmatched trials opens the door to a confound that we lay out in detail, and empirically test, in the current study.

In short, the design of academic radiology studies typically meets the second criterion (having matched and unmatched single- and multiple-target displays), but likely due to limitations in the number of participants (i.e., professional radiologists) and available stimuli (i.e., radiographs) these studies often cannot afford to lose the data necessary to meet the first criterion (e.g., Berabum et al., 1990, 1991). On the other hand, the cognitive psychology literature, which tests novices with artificially generated stimuli, has typically met the first criterion but not the second of using matched displays (e.g., Adamo et al., 2017; Cain & Mitroff, 2013; Fleck et al., 2010; Stothart et al., 2018).

Potential methodological concern in prior "subsequent search miss"(SSM) studies: A participant-driven circularity

Cognitive psychology studies have typically quantified SSM errors as the difference between the average accuracy for finding a target on single-target trials and the average accuracy for finding a second target after finding a first on dual-target trials. Moreover, such studies typically generated random search displays with targets for the single- and dual-target trials drawn from the same distribution of target difficulty. This design should have, on average, equated the overall target difficulty across single- and dual-target displays. Therefore, the logic was that if performance was worse for second targets on average, then finding the first targets had some negative impact on finding the second targets.

Unfortunately, this approach contains a methodological error that inflates the measurement of the SSM effect by failing to account for participants' search tendencies. Consider that on dual-target trials, participants will likely find the less difficult of the two targets first (see Fig. 1a, squares). Which target is easier to detect could be determined by any number of factors (e.g., local contrast, orientation, crowding, location in the display). Therefore, on average, second targets will be more difficult to detect (see Fig. 1b, dark gray distribution) than the average single-target (see Fig. 1b. black distribution). Critically, when comparing average performance across single-target displays (which have relatively easy and difficult targets contributing data) to average performance for the second targets found in dual-target displays (which have a systematic bias towards the relatively more difficult second targets), the data are being pulled from two different underlying difficulty distributions. In effect, the participants' own tendency to detect the relatively easier targets first in dual-target displays interacts with the analysis and experimental design, which we term a "participant-driven circularity," in a way that virtually guarantees an inflation of SSM error estimates since the second targets are sampled from a more difficult distribution than the single targets (for related papers on circularity in functional magnetic resonance imaging experiments, see Kriegeskorte, Simmons, Bellgowan, & Baker, 2009; Vul, Harris, Winkielman, & Pashler, 2009).

Current study

In the current study we quantify this participant-driven circularity by showing its impact on the estimation of SSM errors in actual empirical data, and discuss possible steps for experimental design and analysis that alleviate or eliminate the issue. In general, the solution is to include "matched" single- and dual-target trials and to compare accuracy on the second target from dual-target trials to the accuracy of the single-target displays that were matched to those specific second targets. This better equates the difficulty of the targets included in the measures of single- and multiple-target search, enabling less biased estimates of the size of the SSM effect. In the absence of this design there are analytical steps that can be taken to avoid biased measurements. However, these analyses limit the ability of the experiment to inform claims about the "subsequent" nature of SSM errors, allowing only general statements about the relative difficulty of multiple-target search relative to

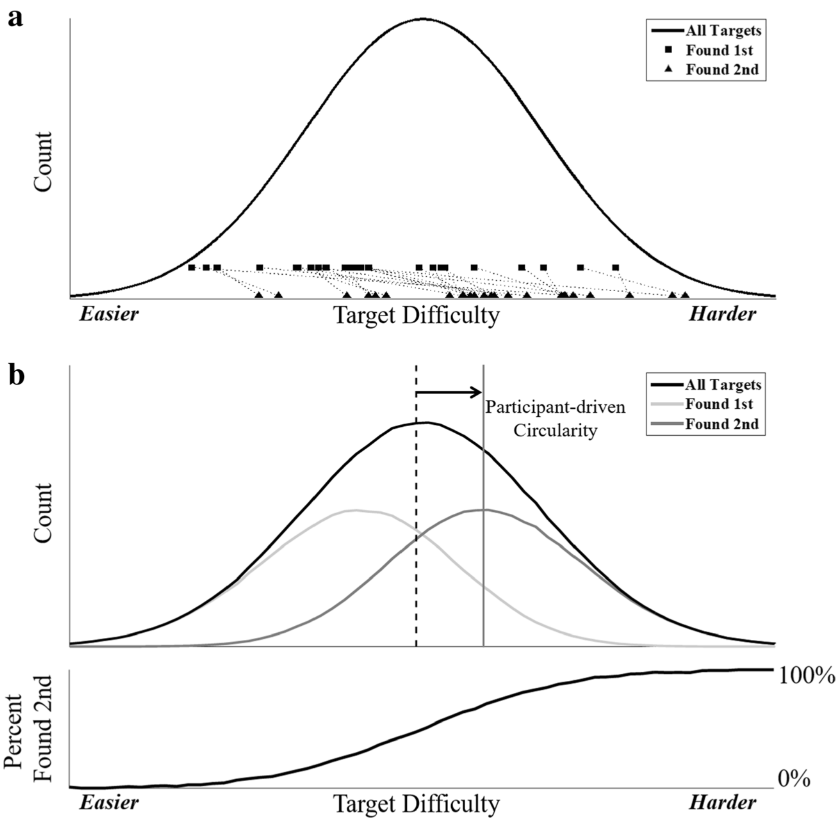

Fig. 1 Demonstration of the theorized participant-driven circularity. (**a**) The black line depicts a theoretical full distribution of search target difficulties. The pairs of square and triangle-shaped data points connected by dashed lines are randomly sampled pairs of targets representing dual-target trials; the less difficult target in the pair (square) is theorized to be the one found first, and the more difficult target (triangle) is taken to be the one found second. (**b**) Sample distributions of the first found (light gray), second found (dark gray), and all targets (black) constructed from 1,000,000 repeats of the sampling of pairs depicted in panel A. The vertical black dashed line depicts the mean difficulty over all targets and the vertical dark-gray solid line depicts the mean difficulty over just the targets found second. The bottom of panel B depicts the probability that a target of a given difficulty was found second

single-target search (see "Takeaway point #2" in the *Discussion*). From a practical standpoint this can still be quite powerful and informative, but from a mechanistic and theoretical standpoint, it is critical to understand the specific contribution of detecting the first target to the errors in detecting the second target.

Methods

Participants

Data were analyzed from 60 undergraduate students from The George Washington University, who voluntarily participated for course credit (age 18–24 years; mean age=19.8 years; 40 female). The participants were evenly split between two conditions. Data from one additional participant in the different salience condition and two additional participants in the same salience condition (see *Stimuli* and *Procedure* below) were removed for being more than two standard deviations from the mean overall accuracy for that condition. An *a priori* participant count of 30 per condition was chosen to provide enough statistical power to reveal possible differences in the circular and matched metrics (see below for details) given that the match metric involves removing one-third of the trials. All experiments were conducted with approval from The George Washington University Institutional Review Board.

Stimuli

Search displays were based on previously published studies of SSM errors (Adamo et al., 2013, 2015, 2017, 2018; Cain & Mitroff, 2013; Stothart et al., 2018). There were 25 items (each 1.3° × 1.3°) per display and each display contained one or two targets with the remaining items being distractors. All items were pairs of perpendicular bars with a small gap between them; targets were perfectly aligned to create a 'T' shape and distractors had the crossbars slightly offset by 1–5 pixels from center to create an 'L' shape (see Fig. 2).

To compare search performance between dual- and single-target trials, displays were generated as matched "triplets." Each triplet was derived from a pair of

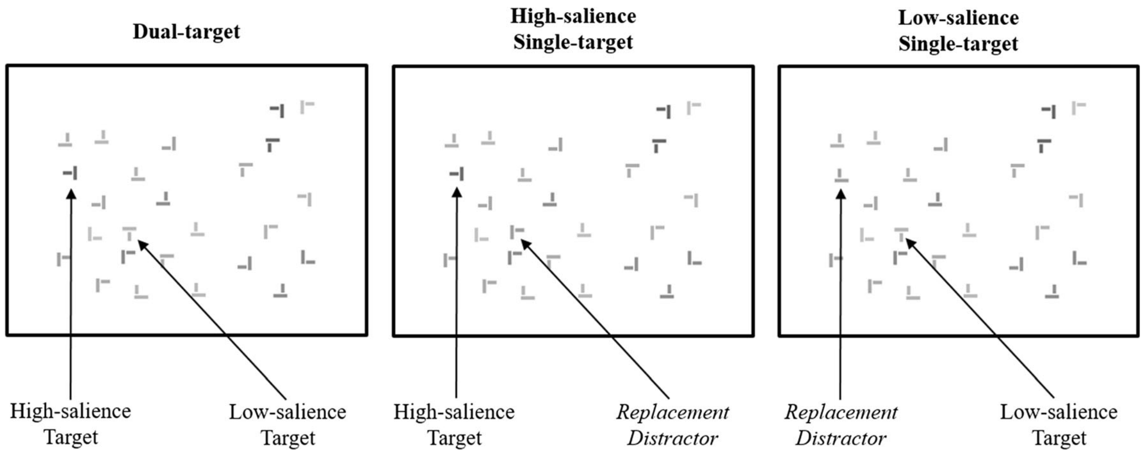

Fig. 2 Example different-salience (DS) trial triplet. Targets and distractor locations are matched across the three stimulus displays. The dual-target display contains a high- and low-salience ‘T’-shaped target among ‘L’-shaped distractors. The high-salience single-target display is matched to the dual-target display with a distractor replacement in the low-salience target’s location. The low-salience single-target display is matched to the dual-target display with a distractor replacement in the high-salience target’s location

targets, which were either present together in the dual-target condition or by themselves in the single-target condition. In all three displays, the targets and distractors had the same color, rotation, and location, except that in the single-target trials, one target remained and the other target was replaced with a distractor (see Fig. 2). This created a set of three displays wherein whichever target was second in the dual-target display there was a matched single-target display to which it could be directly compared.

To connect the current design with the existing literature, two previously used conditions were implemented to investigate the impact of the participant-driven circularity under study on the measurement of the SSM effect. The conditions were run between subjects. In the *different-salience (DS)* condition (Adamo et al., 2013, 2015, 2017, 2018; Cain & Mitroff, 2013), distractors and targets could be high-salience (a gray of 57–65% black; 50% of targets; 5% of distractors) or low-salience (a gray of 22–45% black; 50% of targets; 95% of distractors). Each triplet target pair contained one high-salience target and one low-salience target (see Fig. 2). In the *same-salience (SS)* condition (Stothart et al., 2018), all targets and distractors were the same color (100% black).

Procedure

In both the DS and SS conditions there were a total of 297 experimental trials (99 triplets) split across nine blocks following a practice block of 12 trials. One display from each triplet was assigned randomly to appear in one of three sets of experimental blocks: blocks 1–3, blocks 4–6, or blocks 7–9. This limited the chance that participants would recognize trials from the same triplet by learning the global context and target locations within the displays (Chun & Jiang, 1998).[1]

Participants sat without head restraint approximately 60 cm from a 19-in. LCD monitor and the stimuli were presented with MATLAB via Psychtoolbox (Kleiner, Brainard, & Pelli, 2007). Participants had 15 s to search each display and perform a mouse click on each item they believed was a target. A small blue circle (0.3°) appeared after each click, which has been shown to not affect search performance (Cain & Mitroff, 2013). Participants were instructed to press the spacebar when they were finished searching and encouraged to do so before the 15-s time limit expired. If participants failed to press the spacebar before the time limit expired, this was considered a “time out” and no further clicks could be made. Feedback on target hits and false alarms were provided during the practice trials, but not during the experimental trials.

Planned analyses

To empirically test the assumption underlying the proposed participant-driven circularity – that the targets found first on dual-target trials are easier than the second targets – the single-target trials were grouped according to whether or not their matched target from the dual-target trial was the first target found, and response time and accuracy were assessed. Note that the second target on the dual-target trial may or may not have been found. This analysis ignores single-target trials that are matched to dual-target trials on which both targets were missed, but those were rare (mean ± SD, 5.93% ± 5.74% for DS, 5.50% ± 4.21% for SS).

---

[1] A supplemental analysis revealed no effect of block on accuracy.

To quantify the effect of the participant-driven circularity on estimates of the magnitude of SSM errors, the data from both experimental conditions (DS, SS) were assessed with and without accounting for the matched displays.

First, SSM errors were calculated without incorporating the matched displays in a manner akin to prior cognitive psychology calculations of SSM errors that look at average performance across trial types rather than comparing matching displays in the analyses (e.g., Adamo et al., 2013, 2015, 2017, 2018; Cain & Mitroff, 2013). SSM errors in the DS condition were calculated by comparing the average low-salience target accuracy on single-target trials to the average low-salience target accuracy after a high-salience target was found first on dual-target trials. Trials in which a low-salience target was found first on dual-target trials (mean=28.07%; SD=9.72%) were removed from the analyses in the calculation of this metric. SSM errors in the SS condition were calculated by comparing the average single-target accuracy to the average accuracy for a second target after a first target was found on dual-target trials (e.g., Stothart et al., 2018). In both conditions, comparing the targets found second in the dual-target trials to all of the targets in the single-target trials is the proposed source of participant-driven circularity, therefore we will refer to this as the *circular metric.*

Second, SSM errors were calculated for both the DS and SS conditions when taking into account the matched displays design employed in the current study. This *matched metric* compared the accuracy on the second targets (the targets that were not found first) in the dual-target displays to the accuracy on those same targets in the matched single-target displays. This analysis disregards the remaining, single-target displays (the displays matching the targets that were found first on the dual-target trials) in each triplet, losing these one-third of the total trials in exchange for avoiding potential bias. Here the *matched metric* is reported for the low-salience targets in the DS condition to enable direct comparison to the *circular metric*, but the high-salience targets that are second targets could also be included by including their matched single-target displays in the accuracy calculation for the single targets.

Results

Easier targets are generally found first in dual-target trials

The primary driver of the proposed participant-driven circularity is that the targets found first in the dual-target trials are generally easier than the second targets, resulting in an unfair comparison between performance on those second targets and the overall average performance in single-target trials. To establish that this is true in real data, we took advantage of the matched triplet design and analyzed performance in the single-target trials matched to the individual targets in dual-target trials where participants found at least one target titrated by whether or not the participant found that target first in the dual-target trial. In the DS condition we limited this analysis to the low-salience targets since those are the targets that typically contribute to the SSM metrics. A 2 × 2 ANOVA (order found on dual-target trial, first vs. second, by salience condition, DS vs. SS) on single-target trial accuracy revealed a main effect of order found (*f*(1)=11.38; p<0.001) and salience condition (*f*(1)=35.92; p<0.0001), but no interaction effect (f(1)=1.17; p>0.05). Planned *post hoc* paired two-tailed t-tests showed that the targets found first in the dual-target trials were more likely to be detected in the single-target trials than the second targets for both the DS (*t*(29)=4.64; *p*<0.0001; Cohen's d=0.85) and SS (*t*(29)=5.26; *p*<0.0001; Cohen's d= 0.96) conditions (Fig. 3a). The same 2 × 2 ANOVA on response time revealed a main effect of order found (*f*(1)=26.68; *p*<0.0001), but no main effect of salience condition (*f*(1)=1.75; *p*>0.05) or an interaction effect (*f*(1)=0.20; p>0.05). Planned *post hoc* paired two-tailed t-tests (Fig. 3b) again revealed the advantage for the targets found first in both salience conditions (DS: (*t*(29)=5.52; *p*<0.0001; Cohen's d=1.00; SS: (*t*(29)=6.33; *p*<0.0001; Cohen's d=1.16). Taken together, these results confirm a significant difference in difficulty between the targets that are found first versus second in dual-target displays, which could contribute to an inflated SSM effect in the circular metric.

The participant-driven circularity inflates SSM error estimates but does not account for the entire effect

To quantify the inflation of the SSM effect we analyzed the data from the DS and SS conditions with the circular and matched metric. In the DS condition (Fig. 4a), a significant SSM effect was found for both the circular (*t*(29)=8.95; *p*<0.0001; Cohen's d=1.63) and matched metric (*t*(29)=8.22; *p*<0.0001; Cohen's d=1.50). The matched metric was significantly smaller than the circular metric (*t*(29)=3.32; *p*<0.01; Cohen's d=0.61). The same pattern of results held in the SS condition (Fig. 4b), with significant SSM effects with the circular (*t*(29)=11.60; *p*<0.0001; Cohen's d=2.12) and the matched metrics (*t*(29)=10.65; *p*<0.0001; Cohen's d=1.94). Again, the matched metric SSM effect was significantly smaller than the circular metric SSM effect (*t*(29)=4.10; *p*<0.001; Cohen's d=0.75). Furthermore, at the individual participant level in both the DS and SS conditions, the matched

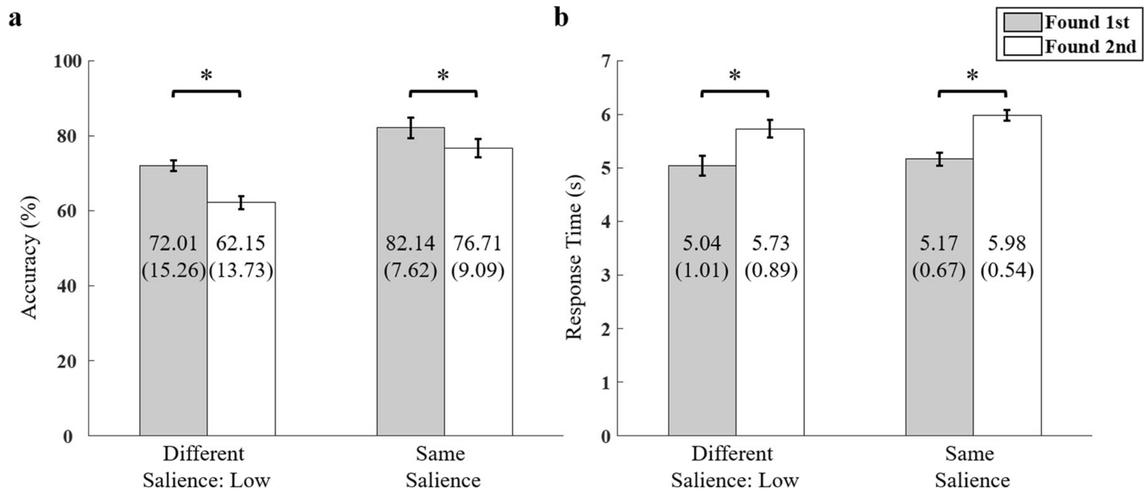

Fig. 3 Targets found first in the dual-target trials are easier than the targets found second in both DS and SS conditions. Accuracy (**a**) and response time (**b**) for single-target trials grouped by whether the matched target was found first in the dual-target trials for the low-salience stimuli in the DS condition and the all black stimuli from the SS condition. Error bars represent the standard error of the mean. Means (and standard deviations) are presented on the column bars; asterisks represent $p<0.01$.

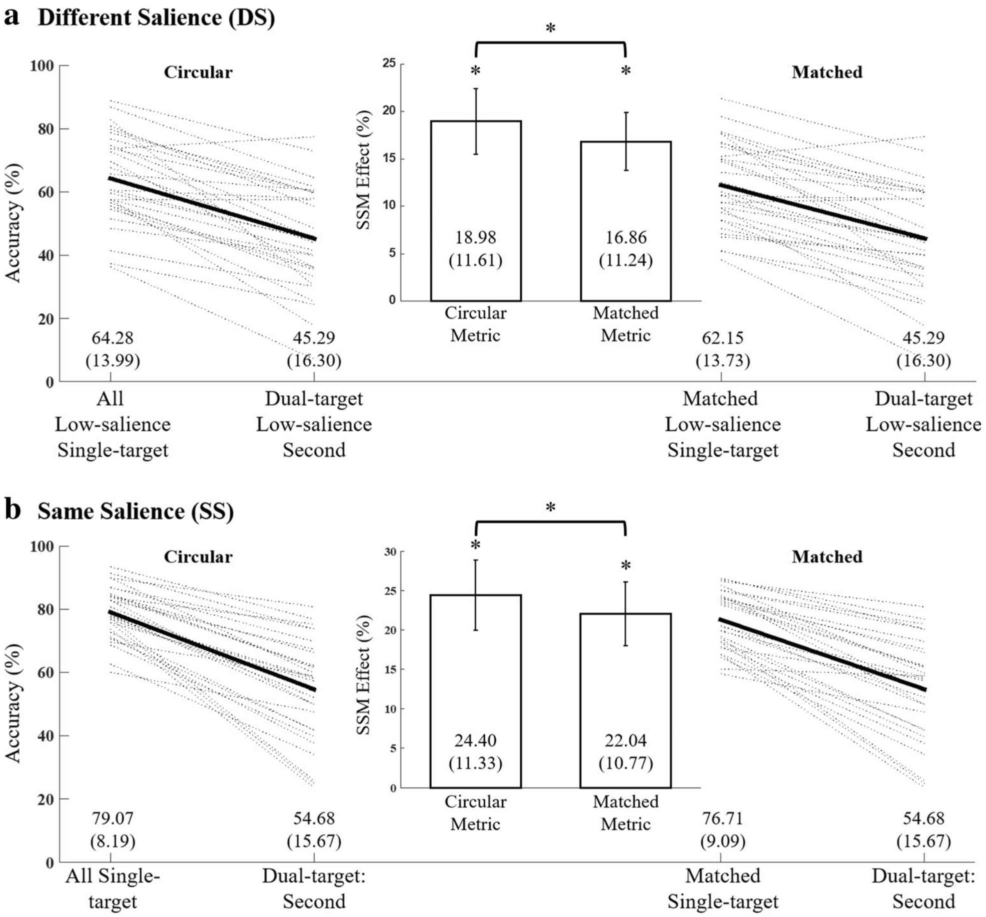

Fig. 4 Accuracy by trial type by condition with circular and matched measures of SSM errors. (**a**) Different-salience (DS) accuracy for single-target low-salience targets and dual-target low-salience targets given the high-salience target was found first when using the circular measure of SSM (left) and the matched measure (right). (**b**) Same-salience (SS) accuracy for all single-target trials and the second target on dual-target trials when using the circular measure of SSM (left) and the matched measure (right). Black lines represent the average data for each analysis and dashed gray lines represent each participant's performance (n=30). Insets display the two SSM measures (error bars represent standard error of the mean; asterisks represent $p<0.01$).

Table 1 Accuracy (and standard deviation) for trial types and circular and matched SSM effect for different-salience (DS) and same-salience (SS) conditions.

Different-Salience Condition				
All low-salience single-target trials (A)	***Matched low-salience single-target trials (B)***	***Dual-target low-salience given high found first (C)***	***Circular metric SSM effect (A–C)***	***Matched metric SSM effect (B–C)***
64.28 (13.99)	62.15 (13.73)	45.29 (16.30)	18.98 (11.61) $t(29)=8.95$; $p<0.0001$	16.86 (11.24) $t(29)=8.22$; $p<0.0001$
Same-Salience Condition				
All single-target trials (A)	***Matched single-target trials (B)***	***Dual-target second target (C)***	***Circular metric SSM effect (A–C)***	***Matched metric SSM effect (B–C)***
79.07 (8.19)	76.71 (9.09)	54.68 (15.67)	24.40 (11.33) $t(29)=11.60$; $p<0.0001$	22.04 (10.77) $t(29)=10.65$; $p<0.0001$

metric consistently produced a smaller estimate of the SSM effect than the circular metric (22/30 participants in the DS condition; 25/30 in the SS condition). However, the matched metric still showed an incredibly consistent SSM effect even without the inflation from the participant-driven circularity (28/30 participants in the DS condition; 30/30 in the SS condition). Given the serial response nature of the task, response times were not analyzed. Accuracy by condition and SSM metric values are depicted in Table 1.

Discussion

Visual search is an important construct as it relies on several cognitive processes (e.g., perception, attention, decision making) and is vital to many critical, real-world professions that demand high levels of accuracy (e.g., radiology and baggage screening). Here, we examined one issue of particular importance: SSM errors, a phenomenon defined as searchers showing worse target detection after having already found another target in the same search. This issue has been studied in academic radiology since the 1960s and is the subject of more recent work in cognitive psychology. However, we argue here that both literatures have, for the most part, not yet provided clear and unbiased quantifications of this phenomenon.

In academic radiology, SSM errors (i.e., satisfaction of search errors) are often theoretically discussed as reduced accuracy for a target given that another target is in the same display compared to when the same target was the only one present (e.g., Berbaum et al., 1990, 1991). A typical approach to studying SSM errors in academic radiology is to take a radiograph that has an abnormality (e.g., a lesion) and then create a second "matched" version of the same radiograph with another abnormality artificially added (e.g., a nodule, see, e.g., Berbaum et al., 1991). This approach allows the researchers to compare performance in finding the native abnormality when it was the only target present to performance when another abnormality was present. This use of matched displays is exactly what the current study suggests doing, however, it is vital to then ensure that data are only included in the analyses when the added abnormality was detected *first*. Some radiology studies do not condition the detection of the native abnormality on whether the artificial abnormality was detected (e.g., Berbaum et al., 1990, 1991, 1993, 1998; Samuel et al., 1995) and others do condition on whether it was detected, but not whether it was detected first (e.g., Berbaum et al., 1994, 2000, 2001). This limits the ability to make inferences from these results about the precise mechanisms involved (e.g., satisfaction, perceptual set, resource depletion) because they do not directly measure whether the detection of the artificial abnormality caused radiologists to miss the native abnormality.

In contrast, cognitive psychology studies have generally run the necessary analyses to isolate the subsequent nature of SSM errors, but have not used matched displays. Without using matched displays, it is possible that the analyses introduce a participant-driven circularity that artificially inflates estimates of SSM errors. Here, we demonstrated that this participant-driven circularity actually impacts estimates of SSM errors in real data. First, two studies showed that the second target found in dual-target trials was systematically harder to find in matched single-target trials (i.e., lower accuracies, longer response times). When married to this participant-driven circularity, the analysis used in previous cognitive psychology experiments caused significantly inflated SSM estimates. Critically, while significant, this inflation did not explain the entire SSM effect – even when the matched analyses that controlled for this bias were used the SSM effect was still clearly observed.

One potential concern of using a matched-display design is there will always be a higher proportion of single-target trials than dual-target trials (each dual-target trial has two accompanying single-target versions). It is possible that participants could learn the statistical properties of the experiment and begin to expect more single-target trials (Cain, Vul, Clark, & Mitroff, 2012), which could potentially impact the SSM effect (Chen & Rich, 2018). Previous studies have found SSM effects from a variety of single- to dual-target trial ratios (e.g., Adamo et al., 2017; Cain et al., 2011; Fleck et al., 2010); however, steps can be taken to reduce this concern. For example, “filler” dual-target trials can be added to equate the single-to dual-target ratio. These added dual-target trials would not have matching single-target trials and therefore would not contribute to SSM error calculations. Future studies could also make the dual-target trials even rarer with the use of “filler” single-target or target-absent trials.

There are several takeaway points from this work, both for how to interpret prior research and how best to move forward.

Takeaway point #1: SSM errors are real and a meaningful source of errors in visual search

The first, and perhaps most important, takeaway from the current study is that even when the participant-driven circularity explored here is avoided using matched displays, the SSM effect remains. The current study serves as a methodological correction for cognitive psychology, but does not in any way suggest these errors are not real. The absolute magnitude of the errors is likely exaggerated, but the effect remains. Further, prior reports of significant variation in SSM errors across conditions (e.g., Cain & Mitroff, 2013; Stothart et al., 2018) are likely qualitatively correct as the participant-driven circularity should be equally present across conditions assuming equivalent variance in target difficulty across conditions.

Takeaway point #2: Prior SSM studies cannot quantify the “subsequent” aspect of SSM errors

The most pressing concern caused by the inflation of absolute estimates of SSM errors is how to accurately quantify the extent of these errors when at least some of the reported SSM errors are artifactual. Moreover, in the absence of a matched design like the one suggested here, it is difficult to estimate how much of an inflation the participant-driven circularity is creating. While the literature has suggested a few different analysis paths for calculating SSM errors (Biggs, 2017), the analyses cannot remove the impact of the participant-driven circularity if the study was not properly designed. Given that prior studies in radiology (e.g., Ashman et al., 2000; Berbaum et al., 1990; Franken et al., 1994; Samuel et al., 1995) and cognitive psychology (e.g., Adamo et al., 2018; Cain et al., 2012; Fleck et al., 2010; Gorbonova, 2017) do not simultaneously condition their analyses on detecting the first target and have a matched design, it is unknown whether the results reported speak to the “subsequent” aspect of SSM errors. In other words, they cannot attribute the difference between single- and multiple-target search performance to whether a first target impacted the detection of a subsequent target in a multiple-target search.

However, even if the suggested criteria are not met, results can still speak to whether multiple-target search performance is worse than single-target performance. One possibility is to calculate a criterion based on single-target trial performance that can be used to establish that the dual-target trials are harder. The simplest and most conservative of those estimates is to simply assume that the expected average chance of successfully detecting both targets in dual-target trials (i.e., dual-target accuracy) is the average accuracy in single-target trials squared. For example, in the current SS condition the average single-target accuracy was 79.07%, so the expected dual-target accuracy would be 62.52%. In the DS condition, we would take the expected dual-target accuracy to be the product of the average accuracies for the high-salience (84.61%) and low-salience (64.28%) single-targets, which equaled 54.39%. Dual-target accuracy below these expectations would indicate that dual-target search is harder than two independent single-target searches, which was true for both the SS (55.25% < 62.52%) and DS (45.29% < 54.39%) conditions. However, even if this standard is met it only allows the study to claim that dual-target search is harder than single-target search. Because the analysis is not conditioned on finding a first target, it is not possible to speak to the “subsequent” nature of the observed SSM effect. The reduction in dual-target performance could result from changes specific to either target and/or their simultaneous interaction. Thus, much of the current literature is not able to quantify the effect of *finding* a first target on finding an additional target.

Takeaway point #3: Prescription for how best to explore SSM errors

The experiments presented here provide a viable framework for cognitive psychology and radiology studies to effectively address SSM errors *above and beyond speaking generally about multiple-target search performance*. Experiments must (1) have matched single-target and dual-target displays to enable fair comparisons and (2) quantify SSM errors as the difference in performance between the second target in dual-target displays given the first is already detected and the matched single target.

Specifically, the suggestion here is to create matched "triplets" – a display is generated with two targets present and the remaining items as distractors. Two additional copies of the same display are then created, one with one of the targets replaced with a distractor and one with the other target replaced with a distractor. This creates three trials that are perfectly matched except for whether they contain both targets or one of the two targets. The three trials should all be presented to participants, but spread across the experiment so as not to appear too close together in time. Researchers can then analyze the data based on the participants' performance and quantify SSM errors with properly matched single-target trials as a baseline while directly addressing the "subsequent" part of the effect. The method we describe here controls for the stimulus-driven (e.g., target position, distractor layout) factors that can impact SSM estimates. There are undoubtedly a wide-variety of other factors, both paradigmatic and statistical (e.g., Becker, Anderson, & Brascamp, 2019), that also play a role that will need to be considered and controlled for as the exploration of SSM errors continues.

Conclusion

SSM errors are a real problem where the finding of one target causes observers to miss an additional target. Finding ways to reduce their occurrence is important for a wide variety of fields including radiology and baggage screening. The literature to date has come close and has provided important insights into the nature of multiple-target visual search, but the work is not yet done. The methodological suggestions put forth here can hopefully help the academic radiology and cognitive psychology literatures to continue to inform this thorny issue.

Acknowledgements We thank Laura Schubel and the Visual Cognition Lab for their help with data collection. We thank Michelle Kramer, Samoni Nag, Courtney Porfido, Joseph Schmidt, Michael Hout, Chad Peltier, and two anonymous reviewers for helpful feedback.

Funding This work was supported by a postdoctoral fellowship from the Ford Fellowship Foundation, the Army Research Office (grant W911NF-16-1-0274), and a Cross Disciplinary Research Fund award from The George Washington University's Office of the Vice President of Research.

Compliance with ethical standards

Open Practices Statement The data for all experiments can be made available upon request.

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*(12), 2569–2574.

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2017). An individual differences approach to multiple-target visual search errors: How search errors relate to different characteristics of attention. *Vision Research, 141*, 258–265.

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2018). Satisfaction at last: Evidence for the "satisfaction" account for multiple-target search errors. *Proceedings SPIE: Medical Imaging, Image Perception, Observer Performance, and Technology Assessment.* 105770A.

Ashman, C. J., Yu, J. S., & Wolfman, D. (2000). Satisfaction of search in osteoradiology. *American Journal of Roentgenology*, *175*(2), 541–544.

Becker, M. W., Anderson, K., & Brascamp, J. W. (2019, May). Eye tracking during search for two unique targets to investigate categorical effects in subsequent search misses & the typical method of calculating SSM is biased; An unbiased method suggests some SSM effects may be spurious. Presentation at the annual meeting of the Vision Sciences Society, St. Pete Beach.

Berbaum, K. S. (2012). Satisfaction of search experiments in advanced imaging. *Proceedings of SPIE, 8291*(1), 82910V–82910V–8.

Berbaum, K. S., Brandser, E. A., Franken, E. A., Dorfman, D. D., Caldwell, R. T., & Krupinski, E. A. (2001). Gaze dwell times on acute trauma injuries missed because of satisfaction of search. *Academic Radiology, 8*(4), 304–314.

Berbaum, K. S., El-Khoury, G. Y., Franken Jr., E. A., Kuehn, D. M., Meis, D. M., Dorfman, D. D., … Kathol. (1994). Missed fractures resulting from satisfaction of search effect. *Emergency Radiology*, *1*(5), 242–249.

Berbaum, K. S., Franken Jr., E. A., Anderson, K. L., Dorfman, D. D., Erkonen, W. E., Farrar, G. P., … Phillips, M. E. (1993). The influence of clinical history on visual search with single and multiple abnormalities. *Investigative Radiology,* 28(3), 191–201.

Berbaum, K. S., Franken Jr., E. A., Dorfman, D. D., Caldwell, R. T., & Krupinski, E. A. (2000). Role of faulty decision making in the satisfaction of search effect in chest radiography. *Academic Radiology, 7*(12), 1098–1106.

Berbaum, K. S., Franken Jr., E. A., Dorfman, D. D., Miller, E. M., Caldwell, R. T., Kuehn, D. M., & Berbaum, M. L. (1998). Role of faulty visual search in the satisfaction of search effect in chest radiography. *Academic Radiology, 5*(1), 9–19.

Berbaum, K. S., Franken Jr., E. A., 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.

Berbaum, K. S., Franken Jr., E. A., 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.

Biggs, A. T. (2017). Getting satisfied with "satisfaction of search": How to measure errors during multiple-target visual search. *Attention, Perception, & Psychophysics, 79*(5), 1352–1365.

Biggs, A. T., Adamo, S. H., Dowd, E. W., & Mitroff, S. R. (2015). Examining perceptual and conceptual set biases in multiple-target visual search. *Attention, Perception, & Psychophysics, 77*(3), 844–855.

Biggs, A. T., Kramer, M. R., & Mitroff, S. R. (2018). Using cognitive psychology research to inform professional visual search operations.

Journal of Applied Research in Memory and Cognition, 7(2), 189–198.

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(7)*, 866–871.

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*(5), 1398–1408.

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.

Chan, L. K. H., & Hayward, W. G. (2013). Visual search. *WIREs Cognitive Science, 4*(4), 415–429.

Cheng, P., & Rich, A. N. (2018). More is better: Relative prevalence of multiple targets affects search accuracy. *Journal of Vision*, *18*(4), 1–14.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*(1), 28–71.

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision, 11*(5):1–36.

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(*1), 60–71.

Franken, E. A., Jr, Berbaum, K. S., Lu, C. H., Kannam, S., Dorfman, D. D., Warnock, N. G., … Pelsang, R. E. (1994). Satisfaction of search in the detection of plain-film abnormalities in abdominal contrast studies. *Investigative Radiology, 29*(4), 403–409.

Gorbunova, E. S. (2017). Perceptual similarity in visual search for multiple targets. *Acta Psychologica*, *173*, 46–54.

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

Kriegeskorte, N., Simmons, W. K., Bellgowan, P. S. F., & Baker, C. I. (2009). Circular analysis in systems neuroscience: The dangers of double dipping. *Nature Neuroscience*, *12*(5) 535–540.

Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research, 51*(13), 1526–1537.

Samuel, S., Kundel, H. L., Nodine, C. F., & Toto, L. C. (1995). Mechanism of satisfaction of search: Eye position recordings in the reading of chest radiographs. *Radiology, 194*(3), 895–902.

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

Stothart, C., Clement, A., & Brockmole, J. R. (2018). Satisfaction in motion: Subsequent search misses are more likely in moving search displays. *Psychonomic Bulletin & Review, 25*(1), 409–415.

Tuddenham, W. J. (1962). Visual search, image organization, and reader error in roentgen diagnosis. *Radiology*, *78*, 694–704.

Vul E., Harris C., Winkielman P., Pashler H. (2009) Puzzlingly high correlations in fMRI studies of emotion, personality, and social cognition. *Perspectives on Psychological Science, 4*, 274–290.

Wetter, O. E. (2013). Imaging in airport security: Past, present, future, and the link to forensic and clinical radiology. *Journal of Forensic Radiology and Imaging, 1*(4), 152–160.