

Getting satisfied with “satisfaction of search”: How to measure errors during multiple-target visual search

Adam T. Biggs¹

Published online: 28 March 2017
© Psychonomic Society, Inc. (outside the USA) 2017

Abstract Visual search studies are common in cognitive psychology, and the results generally focus upon accuracy, response times, or both. Most research has focused upon search scenarios where no more than 1 target will be present for any single trial. However, if multiple targets can be present on a single trial, it introduces an additional source of error because the found target can interfere with subsequent search performance. These errors have been studied thoroughly in radiology for decades, although their emphasis in cognitive psychology studies has been more recent. One particular issue with multiple-target search is that these subsequent search errors (i.e., specific errors which occur following a found target) are measured differently by different studies. There is currently no guidance as to which measurement method is best or what impact different measurement methods could have upon various results and conclusions. The current investigation provides two efforts to address these issues. First, the existing literature is reviewed to clarify the appropriate scenarios where subsequent search errors could be observed. Second, several different measurement methods are used with several existing datasets to contrast and compare how each method would have affected the results and conclusions of those studies. The evidence is then used to provide appropriate guidelines for measuring multiple-target search errors in future studies.

Keywords Visual search · Visual attention · Multiple-target search · Satisfaction of search

✉ Adam T. Biggs
adam.biggs.1@us.af.mil

¹ Naval Medical Research Unit Dayton, Wright-Patterson Air Force Base, 2624 Q Street, Bldg. 851, Area B, Fairborn, OH 45433, USA

Significance

Multiple-target search studies are becoming more common in cognitive psychology. These studies sometimes better parallel existing, real-world conditions where multiple targets in a single search represent a significant issue (e.g., airport security screening, radiology). However, improper measurement methods can overestimate or underestimate multiple-target search errors, depending on the situation and the method used. This inaccuracy is especially troubling when considering that these situations are specifically designed to address real-world policies. Thus, this study fills an important gap by providing guidance and clarification on when and how multiple-target search errors should be measured.

Visual search, or the act of finding targets among distractors, is a common process conducted countless times each day. As with any cognitive task, numerous factors influence performance and induce errors. Visual search is prone to a particular type of error that has gone largely understudied in the cognitive psychology literature: errors made when more than one target is present in a display. Such errors are pervasive and have been suggested to account for up to one-third of errors in radiology (Anbari, 1997; Berbaum, Franken, Caldwell, & Schartz, 2010; Krupinski, 2010) and a potential majority of errors in emergency medicine (Kuhn, 2002; Voytovich, Rippey, & Suffredini, 1985). Many laboratory experiments have used visual search or general attention tasks with only one target present per display, and although this manipulation controls for some factors, it ignores the potential impact of multiple-target search errors. However, multiple-target search accuracy can be more difficult to assess than single-target search accuracy. Myriad methods have been used, each adapting to the particular situation. This investigation is designed to assess various forms of measuring multiple-target search errors. Specifically, this examination will focus on the

differences in measurement style, any corresponding impact different measures would have on existing interpretations of data, and finally, a set of recommendations for any future studies measuring multiple-target search errors. The investigation will begin with a discussion of differences between previous uses of the term *multiple targets* in the literature as well as some of the factors known to influence multiple-target search.

Multiple targets versus multiple categories

There is an important distinction in visual search studies between the possible target set and the possible number of targets present. The target set comprises the various identities of possible targets during the search task, whereas the number of targets present indicates the possible number of targets that could appear in a single search display. Unfortunately, the nomenclature does not always differentiate between these two because “multiple targets” has been used to reference both the target set and the number of targets present. Some attention tasks may even include target sets with multiple objects, yet neither the target set nor the number of targets are critical factors in the research. For example, a modified flanker paradigm (e.g., Biggs et al., 2012; Cosman & Vecera, 2009, 2010a, 2010b; Lavie & Cox, 1997) often has participants search for one of two possible targets. Although a target is present in every display and there is more than one target type, the research focus is on distractor processing and not target processing. Multiple possible targets are used as a means to vary responses and target-distractor compatibility conditions. Therefore, these examples would not qualify as investigations into multiple-target search.

Another core example of multiple targets includes the cost associated with searching for different target types. For example, participants are slower and less accurate when searching for two target types versus a single target type—such as during a search for guns and knives versus searching for either guns alone or knives alone (Godwin, Menneer, Cave, & Donnelly, 2010; Godwin, Menneer, Cave, Helman, et al., 2010; Menneer, Barrett, Phillips, Donnelly, & Cave, 2007; Menneer, Cave, & Donnelly, 2009; Menneer, Donnelly, Godwin, & Cave, 2010). These results are often discussed in terms of “multiple target search” or “multiple category search,” although others have used the label “hybrid search” to describe multiple target types—that is, hybrid search describes a visual search through the display with additional searches through the memory set of target types (Boettcher, Drew, & Wolfe, 2013; Boettcher & Wolfe, 2015; Drew, Boettcher, & Wolfe, 2015; Drew & Wolfe, 2014; Wolfe, 2012a; Wolfe, Aizenman, Boettcher, & Cain, 2016; Wolfe, Boettcher, Josephs, Cunningham, & Drew, 2015). This evidence does tap into a larger debate about whether simultaneous attentional control settings can be set for multiple target

types (cf. Irons, Folk, & Remington, 2012; Stroud, Menneer, Cave, & Donnelly, 2012; Wolfe, 2012a). However, this example once again demonstrates the impact of variability within the target set and not the presence of multiple targets in a single display.

The multiple target distinction—that is, between multiple targets in the target set and multiple targets in a single display—is also important because many visual search studies require only a present or absent decision (cf. Biggs, Cain, Clark, Darling, & Mitroff, 2013; Hout & Goldinger, 2010, 2012; Nakashima, Kobayashi, Maeda, Yoshikawa, & Yokosawa, 2013). The search task ends upon reaching this decision, which is the most important difference between these disparate uses of “multiple targets.” Specifically, search must continue *after* finding the target for a search display that can contain more than one target. This component of visual search has gone largely understudied, and it will be the focus of this investigation. The phrase “multiple-target search” will be used throughout the remainder of the article to describe studies wherein more than one target can be present in a single search display.

Multiple-target search or foraging?

Different circumstances fulfill the definition offered here for multiple-target search, yet these circumstances are not all equivalent. Perhaps the most notable distinction involves the example of foraging. In this case, visual search proceeds in a target-rich environment until the searcher elects to leave a particular area and move onto the next (e.g., Cain, Vul, Clark, & Mitroff, 2012; Ehinger & Wolfe, 2016; Wolfe, 2012b). Results of foraging experiments are often compared to the predictions of marginal value theorem (Charnov, 1976), which suggests that an optimal searcher will abandon search in the current location (i.e., in the current display) when the rate of return reaches or is about to fall below the overall rate of return for the environment. The eventual decision to terminate a foraging search can be affected by various factors. For example, if the searcher is looking for berries, systematic variance in patch quality (e.g., changing seasons) can significantly influence the decision to remain at a particular patch (Fougny, Cormiea, Zhang, Alvarez, & Wolfe, 2015; Zhang, Gong, Fougny, & Wolfe, 2015).

Foraging situations clearly fit the broader description of multiple-target visual search as multiple targets—even dozens of targets—can appear in a single search display. However, this example is notably different from the multiple-target search challenges present in airport security screening (for a review, see Biggs & Mitroff, 2015a). One difference is in the sheer number of targets. Even though it is theoretically possible that a bag going through airport security might contain dozens of contraband items, the situation is less plausible than a bush with many ripened berries.

Still, the overall number of targets is not the most important difference between these scenarios. The key difference between foraging and standard multiple-target search involves the quitting rules, or the criterion by which the searcher chooses to terminate search without having found all possible targets. Quitting thresholds are particularly important in visual search as they have been proposed as one of the primary reasons searchers fail to find rarely appearing targets (Wolfe & van Wert, 2010) in addition to imposing a significant cognitive burden on the observer (Dougherty, Harbison, & Davelaar, 2014). In foraging situations, searchers often voluntarily terminate search despite knowing that additional targets are present. Search could be discontinued for a variety of reasons, such as the rate of return falling below the optimal level, but the key point is that the searcher knowingly does not collect or identify all possible targets present.

Other multiple-target search scenarios, such as in radiology or airport security screenings, do not have a similarly optimal quitting threshold. These multiple-target search scenarios require searchers to continue searching until all possible targets have been located—or what could be described as an exhaustive search strategy. Thus, the key difference between foraging and exhaustive multiple-target search is whether the quitting rule involves optimal search strategies as in foraging or exhaustive search strategies as in airport security screenings.¹

One example raises questions about this exhaustive search rule as it pertains to foraging. Specifically, searchers will switch freely between target types in a foraging environment with many conspicuous targets, yet when focused attention is required to identify individual targets, searchers tend to exhaust an entire category of targets before switching to another (Kristjánsson, Jóhannesson, & Thornton, 2014). This scenario appears to be foraging with an exhaustive search rule, although the critical distinction is that the exhaustive search procedure is once more an optimal strategy—one searchers are using to maximize yield through seemingly efficient search procedures. Even if the searchers then select an exhaustive or near-exhaustive search rule to select targets from a particular category, the quitting threshold remains optimal, and searchers could select a different strategy if a more

effective one is identified. A true exhaustive search, however, will always require that all targets be found with no option to terminate search based on an optimal yield rate. Radiology and airport baggage screening are exemplars of this exhaustive search scenario because attaining an optimal yield does not fulfill the purpose of the task or identify good performance—there can be serious consequences with even a single missed target.

Because most visual search operates on the exhaustive search principle (i.e., the instructions are to find all possible targets), foraging, while fitting the broader definition of multiple-target search, will be considered a unique example and outside the scope of this investigation. The focus will be on multiple-target search studies with an exhaustive search termination rule.

Known causes of multiple-target search errors

Multiple-target search errors remain largely understudied in cognitive psychology, although the problem has received far more significant attention from the radiological literature over the last 50 years (for a review, see Berbaum et al., 2010). With this previous evidence, there has been some significant headway in understanding the causes behind these errors. Notably, the issue appears to be multifaceted, where multiple sources can simultaneously contribute to the likelihood of these errors (Cain, Adamo, & Mitroff, 2013). The original explanation involved the possibility that searchers became “satisfied” upon finding the first target and quit searching (Smith, 1967; Tuddenham, 1962). This belief gave rise to the original moniker for multiple-target search errors—“satisfaction of search,” or SOS errors. However, significant evidence has demonstrated that searchers continue to search after finding the first target (Berbaum, Dorfman, Franken, & Caldwell, 2000; Berbaum et al., 1991; Fleck, Samei, & Mitroff, 2010). Only recently has evidence demonstrated that effort expended after finding one target might be related to multiple-target search errors (for a more thorough discussion, see Adamo, Cain, & Mitroff, 2016). Given the lack of a predominant relationship between multiple-target search errors and the satisfaction explanation, “subsequent search misses” (SSM; Adamo et al., 2013) has been suggested as an alternative and more accurate label.

Another proposed explanation has been the perceptual set bias (Berbaum et al., 2010), where the searcher becomes biased to look for additional targets consistent with the found target. For example, a searcher might become biased to look for other beverage containers after finding a water bottle—or biased to look for guns after finding bullets. Although the idea is logical and has been presented for many years, only recently has any strong evidence come out in support of the perceptual set bias. Using a substantially varied target set and the possible presence of multiple-targets, the mobile application Airport Scanner has proven to be a remarkably versatile research tool

¹ It is hypothetically possible to have a multiple-target search scenario with many different targets, as in foraging, and a need to find every single target, as in exhaustive search. This instance then raises the question whether foraging is better defined by the sheer number of targets or by the optional (or optimal) quitting threshold. If foraging were defined by the sheer number of targets, then the cutoff between foraging and nonforaging search becomes an arbitrary number (e.g., nonforaging search has under 10 targets in a display and foraging search has more than 10 targets in a display). Conversely, the optimal versus exhaustive quitting threshold provides a more meaningful and theoretical distinction between the two. As such, the opinion put forth here is that “foraging” requires an optimal quitting threshold, and even a search with dozens of different targets and an exhaustive quitting rule would not truly be a “foraging” search.

(Mitroff & Biggs, 2014; Mitroff et al., 2015). Airport Scanner data is particularly useful in perceptual set bias assessments because it included target types of similar and different colors (i.e., a perceptual match or mismatch) as well as functional similarity between target types (i.e., a conceptual match or mismatch). In this larger data set, SSM errors were shown to be fewer when the found target and subsequent target matched in color (e.g., two blue targets) or when the two targets matched conceptually (e.g., guns and bullets; Biggs et al., 2015). This finding suggests that the found target may indeed bias search priorities during subsequent search.

A third explanation involves the resource depletion argument (Berbaum et al., 1991; Cain & Mitroff, 2013). According to this argument, the found target places a cognitive burden on the searcher because he or she must then maintain the location and identity of the found target. This burden limits the searcher's cognitive capacity during subsequent search and reduces accuracy (e.g., highlighting; Cain & Mitroff, 2013). Additionally, searchers—even professional visual searchers such as airport security screeners—do not exhibit as many SSM errors when presented with a single display multiple times rather than continuing to search the same display after finding a target (Cain, Biggs, Darling, & Mitroff, 2014). Subsequent search accuracy thus improves without the need to maintain either the location or identity of the found target, which supports the idea that resource depletion is one of the primary factors contributing to SSM errors. Numerous situational factors can likewise affect SSM error rates, including anxiety (Cain, Dunsmoor, LaBar, & Mitroff, 2011), task structure (Clark, Cain, Adcock, & Mitroff, 2014), and decision-making criteria (Biggs & Mitroff, 2015b). However, these other factors can be generally aligned to support one or more of the three theoretical positions: satisfaction, perceptual bias, or resource depletion.

Methods of measuring multiple-target search errors

The discussion thus far has centered on describing the specific definition of multiple-target search, how the presence of multiple targets differs in foraging search, and the known causes of SSM errors. This prerequisite information allows for a focused examination of the core element: how different studies have measured multiple-target search errors, and which measurement is best.

The most straightforward method would be to compare search accuracy for a particular target on single-target trials versus dual-target trials. This method has been used previously to measure SSM errors, albeit with the caveat that at least one target had been detected on the dual-target trials (Fleck et al., 2010). However, there is one particular problem with this approach. SSM errors are not simply a measure of search accuracy but rather a specific error that occurs after finding at least one target in a multiple-target search. If the requirement

involves at least one found target, then the target of interest could have been found first or second. For example, consider a multiple-target search for a water bottle and a gun. If the critical question involves search accuracy for the gun after the water bottle has been found, then the prerequisite is that the water bottle must be found first.

This chief requirement establishes the first possible method of measuring SSM errors, which will be described as the “baseline method.” In the baseline method, single-target accuracy for a given target is used as the baseline and compared against dual-target accuracy for the same target—assuming that another target had been found first on the dual-target trial. However, assessing SSM errors across different target types requires taking certain asymmetries into account. A second method, known as the “weighted method,” does precisely that by taking into account the number of instances where each target type appeared first. For example, if you have two target types (e.g., Target A and Target B), then measuring overall SSM errors requires adjusting the baseline method for the percentage of dual-target trials where Target A was found first versus dual-target trials where Target B was found first.

Two more methods have been proposed during the course of these studies, although neither has been explicitly used in published empirical assessments. One measure will be described as the “adjusted method,” which takes baseline accuracy into account differently than the baseline method. For example, two target types can have significantly different single-target accuracy rates, such as 95 % accuracy for Target Type A and 35 % accuracy for Target Type B. Either target might exhibit a 10 % decline in accuracy after another target had been found, albeit the relative decline is substantially different when compared to 95 % baseline accuracy versus 35 % baseline accuracy. The adjusted method accounts for this disparity by incorporating the relative difference into the calculation [e.g., $(35\% - 25\%) / 35\% = 28.57\%$ SSM errors]. A final method takes into account not just single-target accuracy but also attempts to incorporate the accuracy variability of multiple target types into the SSM calculation. This method, described as the “dependent method,”² incorporates single-target accuracy for both the target in question and the found target. For example, if a searcher only has an 80 % chance of finding Target A and a 50 % chance of finding Target B during a single-target trial, then the searcher should only have a 40 % chance of finding Target B after finding Target A (e.g., $80\% * 50\% = 40\%$). The intent is to incorporate multiple sources of error variance into the expected baseline for comparison; that is, if a search has multiple targets, then it also has multiple individual searches that should both impact the expected accuracy rate.

² Credit for creating the dependent method, and the active/passive SSM error distinction in the discussion section, belongs to Jeremy Wolfe.

All four methods (baseline, weighted, adjustment, and dependent) represent viable ways of measuring SSM errors. Each has its own merits in theory and could apply to a wide variety of situations. The intent of this investigation is to assess each method within a real dataset and compare how each measurement would lead to a particular interpretation of the data.

Method

Both datasets used in these analyses come from previously published studies (Biggs & Mitroff, 2014; Biggs et al., 2015; Mitroff et al., 2015). They are used here for illustrative purposes to assess how different SSM measures might have influenced the corresponding interpretations of the results. Basic methodological details are given to provide insight into how these paradigms differ. Full methodological details are described in the original publications.

Ts and Ls Search (Biggs & Mitroff, 2014)

This study examined different predictors of dual-target accuracy between professional and nonprofessional visual searchers. The final dataset included 103 nonprofessional participants (members of the Duke University community) and 72 professional participants (members of the Transportation Security Administration). The primary hypothesis involved whether similar factors predicted search accuracy in single-target and multiple-target search for both professional and nonprofessional searchers (see also Biggs et al., 2013). Results indicated that similar factors could predict accuracy in both single-target and multiple-target search, although professional search experience continued to determine which factors best predicted search accuracy.

Search displays included an invisible grid with 25 items randomly arranged upon the 8×7 positions. Individual items were randomly offset from perfect grid alignment by 0–10 pixels (see Fig. 1). Target items were two perpendicular bars formed into a perfect “T” shape, whereas distractor items were two perpendicular bars offset to form pseudo-“L” shapes. Target items appeared in one of two salience levels (high salience: 57%–65% black; low salience: 22%–45% black), although distractors were drawn from the same salience range. Experimental trials included 125 trials with a single, high-salience target; 40 trials with a single, low-salience target; 40 trials with both a high-salience and a low-salience target (dual-target trials); and 50 trials with no target present. This trial distribution was based on previous manipulations that had demonstrated significant SSM effects (e.g., Fleck et al., 2010). Participants were informed that 0, 1, or 2 targets could be present on any given trial, and responded to targets by making a mouse click directly on a target to identify its

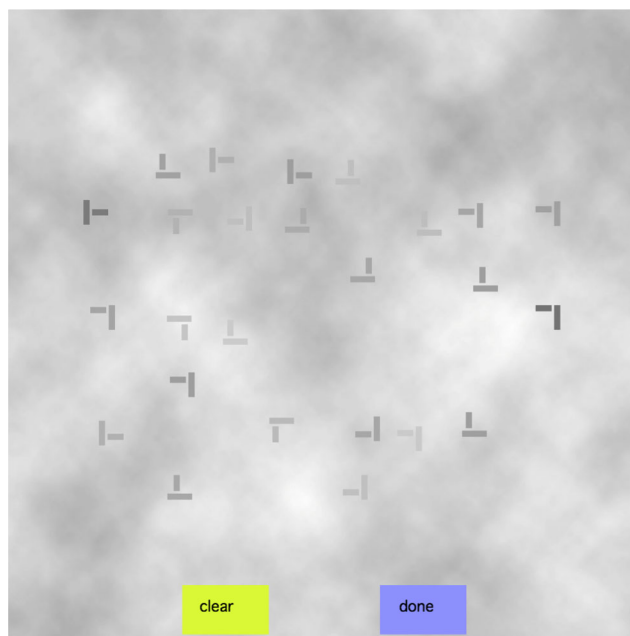


Fig. 1 Sample figure of the stimuli used in the Biggs and Mitroff (2014) “T”s and “L”s search. A perfect “T” target is displayed in the upper left-hand corner of the display

presence. Participants ended a trial by clicking the “Done” button at the bottom of the screen. The computer then proceeded to the next trial.

Airport Scanner (Biggs, Adamo, Dowd, & Mitroff, 2015; Mitroff, Biggs, Adamo, Dowd, Winkle, & Clark, 2015)

The mobile application Airport Scanner provides a highly versatile platform capable of collecting a monumental amount of data. Players consented to the terms and conditions of the standard Apple User Agreement and those provided by Kedlin Co. (<https://www.airportscannergame.com>) upon installing the Airport Scanner application. The Duke University Institutional Review Board approved the analytical investigation of the original dataset.

During Airport Scanner gameplay, players act as security officers at various airports and are tasked with finding prohibited items in simulated X-rays of airport luggage. Bags are viewed one at a time, and target presence is reported by directly tapping onto each illegal item. Individual bags could contain zero to three “illegal” target items and zero to 20 “legal” items, with the legal items serving as distractor stimuli. The target set included 94 possible target items and 94 possible distractor items. These analyses focused only on certain targets that appeared alongside distractor stimuli and could be viewed without in-game upgrades—leaving a final target set of 79 different possible target items. Both targets and distractors appeared in a semitransparent state so that they could overlap without occluding one another. Identity, color,

and size of individual targets were assigned by Kedlin Co. for gameplay purposes.

For the original study, the data included approximately 1.1 billion total trials. Significant filters were used to trim the data down to only the most reliable information (for further details on data trimming procedures in Airport Scanner, see Biggs, Adamo, & Mitroff, 2014; Mitroff & Biggs, 2014; Mitroff et al., 2015). Single-target accuracy data were collected from those trials in which only one target was present, and included 1,795,907 total trials after filters were applied. Accuracy values for individual targets were calculated based upon the particular target type (Range = 703 to 114,390 appearances, $M = 22,733$ appearances, $SE = 3,724$ appearances). Dual-target trials included 126,579 valid trials after all data trimming (for more details, see Biggs et al., 2015), including at least 136 appearances for each target item with a maximum of 15,086 appearances ($M = 3,205$ appearances, $SE = 440$ appearances). The analyses presented here focus primarily on the identical target analyses (see also Mitroff et al., 2015). Dual-target trials rarely contained two identical targets (dual-target trials with identical targets: 4.34 %, dual-target trials with different targets: 95.66 %), and so the dataset was limited only to those target types contributing at least 20 valid cases to the dataset after all filters were applied. This cutoff point (i.e., 20 valid cases) is based on previous work involving rarely appearing targets where 20 valid cases was considered sufficient for meaningful analyses (see Fleck & Mitroff, 2007; Russell & Kunar, 2012; Wolfe, Horowitz, & Kenner, 2005; Wolfe et al., 2007). This approach left 33 target types available for the SSM analyses.

SSM Calculations

For the baseline method (1), SSM errors are calculated as the difference score between single-target accuracy for that particular target type and dual-target accuracy for the same target type given that another target had been found first. For the weighted method (2), SSM errors are calculated by adjusting the observed SSM errors based upon the number of contributing cases. To do so, the proportion of cases where Target Type A had been found first is compared to the proportion of cases where Target Type B had been found first. SSM errors are then calculated for each target type and the overall SSM error rate is adjusted based upon the number of cases where each target type had been found first. For the adjusted method (3), the measurement is very similar to the baseline method. The difference is that the single-target accuracy is used to control for variance in single-target accuracy. For the dependent method (4), the measurement is designed to take into account accuracy differences in both the found target accuracy and second target accuracy.

These formulas are represented below with the following variables to illustrate each method: “A” represents the

single-target accuracy of Target Type A, “a” represents the dual-target accuracy of Target Type A, “B” represents the single-target accuracy of Target Type B, and “b” represents the dual-target accuracy of Target Type B. In these cases, assume that “A” and “B” are the only two possible target types, another target had been found first on the dual-target trials, and a significant difference from zero would represent significant SSM errors.

$$A - a \tag{1}$$

$$(A - a) * (\%b \text{ found first}) + (B - b) * (\%a \text{ found first}) \tag{2}$$

$$\frac{A-a}{A} \tag{3}$$

$$(A * B) - a \tag{4}$$

Results

Ts and Ls Search (Biggs & Mitroff, 2014)

The original study assessed how various performance metrics related to multiple-target search accuracy between professional visual searchers (members of the Transportation Security Administration) and nonprofessional visual searchers (members of the Duke community). With regard to SSM errors, there are two important aspects to consider: whether significant SSM errors were observed in general, and whether the observed SSM errors differed significantly between groups. The original results supported significant SSM errors in both conditions, but significantly more SSM errors among the professional searchers than the nonprofessional searchers. Notably, the original study presented dual-target trials (i.e., trials on which two targets appeared in a single display) that always contained one high-salience target and one low-salience target. The high-salience targets were found first far more often than the low-salience targets, but nonprofessional searchers were even more likely to find the high-salience target first ($M = 86.09$ %, $SE = 1.00$ %) than professional searchers ($M = 82.74$ %, $SE = 1.30$ %), $t(173) = 2.07$, $p = .04$. Because of the significant difference, the original study used a weighted SSM measure for analyses to control for which first-found target contributed more to the SSM error rate.

See Fig. 2 for the SSM errors as assessed by all four measurements, and see Table 1 for statistical comparisons between professional and nonprofessional searchers. The baseline method yielded significant SSM error rates for both conditions with a significant difference between the groups. In this case, the baseline method delivers a conclusion similar to the weighted method, albeit with smaller effect sizes. One possible explanation for the similarity

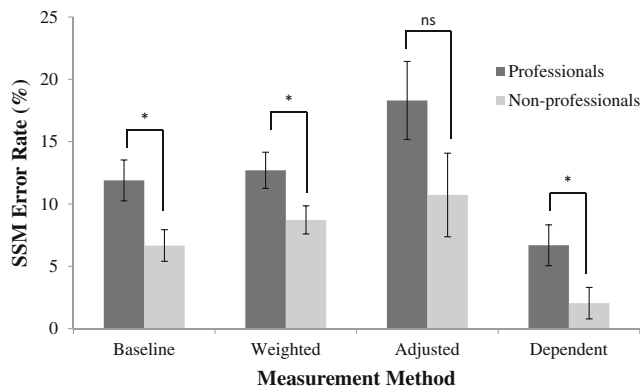


Fig. 2 SSM error rates (in percentage with standard error bars) as assessed by the four different measurement methods for the Biggs and Mitroff (2014) study. Asterisks indicate a significant difference between groups, whereas *ns* indicates a nonsignificant difference

is the large discrepancy in finding the high-salience target first. With one target type found so much more often by both groups, the baseline method will yield error rates similar to the weighted method. The weighted method still has the advantage of accounting for more of the available data (i.e., all dual-target trials with at least one target found first), although, for practical purposes, the weighted method is comparable to the baseline method when there is a large discrepancy in the type of target found first.

The adjusted method yielded significant SSM error rates for both professional and nonprofessional searchers, but not a significant difference between groups. This data would change the original conclusion of the study that SSM errors were present *and* there was a significant difference between groups. Notably, the adjusted method yielded the highest SSM error rates of any measurement method. However, the effect size and statistical difference analyses are very different due to the increase in variance. This method was specifically intended to account for individual differences in performance (e.g., a searcher with

a baseline accuracy rate of 65 % vs. a searcher with a baseline accuracy rate of 95 %), but the result appears to inflate the overall variance. It should also be noted that the adjusted method was calculated only for the low-salience targets, which does include the majority of the data as participants located the high-salience targets first far more often than the low-salience targets on dual-target trials. With such a large discrepancy in the target type found first, however, using the adjusted method to calculate SSM errors when the low-salience target is found first becomes problematic. Specifically, there are so few contributing trials to the measurement that without some additional accountability, as with the weighted method, the error inflation becomes extremely large. The adjusted method thus yields even larger SSM errors for the high-salience second target—professional searchers: $M = 23.81\%$, $SE = 3.51\%$; nonprofessional searchers: $M = 30.42\%$, $SE = 3.40\%$; $t(173) = 1.32$, $p = .19$, Cohen's $d = 0.21$.

The dependent method yielded significant SSM error rates for professional searchers, but nonsignificant SSM error rates for nonprofessional searchers. Although the comparison between groups remained statistically significant, these results would change some existing conclusions about this paradigm. In particular, this manipulation has yielded significant error rates in several published studies (Biggs & Mitroff, 2014, 2015b; Cain et al., 2011; Fleck et al., 2010), yet those previous studies used variants of the baseline and weighted methods. In turn, the discrepancy raises the question as to whether significant SSM errors would have been observed in these previous studies with the dependent method. The dependent method also yielded lower SSM errors in general than any of the other methods, which may indicate that the dependent method is avoiding the potential inflation of error variance by incorporating baseline accuracy for both the

Table 1 Significance analyses for SSM errors using all four of the different proposed methods to measure SSM errors from Biggs and Mitroff (2014). Data in the group columns provide the corresponding Cohen's d of SSM error rates for each individual assessment (e.g., for professional visual searchers using the baseline method). The comparisons columns are included to assess whether each measurement method would have demonstrated a significant difference in SSM error

	Group		Comparisons	
	Professionals	Nonprofessionals	<i>t</i> tests	Conclusion?
Baseline	1.72**	1.04**	$t(173) = 2.55, p = .01^*$	Same
Weighted	2.07**	1.53**	$t(173) = 2.18, p = .03^*$	-
Adjusted	1.39**	0.63**	$t(173) = 1.59, p = .12$	Different
Dependent	0.97**	0.32	$t(173) = 2.29, p = .02^*$	Different

* $p < .05$. ** $p < .01$.

rates between groups (*t* tests), and the conclusion column provides a qualitative comparison between the measurement method being demonstrated and the conclusion of the original study (*same* implies that the conclusion would have been the same as the original study, and *different* implies that the conclusion would have been different from the original study)

found target and second target into the formula for the expected difference.

Airport Scanner (Biggs, Adamo, Dowd, & Mitroff, 2015; Mitroff, Biggs, Adamo, Dowd, Winkle, & Clark, 2015)

In the original study, Biggs et al. (2015) and Mitroff et al. (2015) used data from Airport Scanner to assess possible influences of the first found target upon subsequent search. One particular comparison included any possible perceptual set and conceptual set biases that might develop. For example, if searchers first located bullets in an X-ray bag, would they become more biased to look for guns over water bottles during subsequent search? Airport Scanner provided an excellent medium for such as assessment as the target set included dozens of different items that varied in color and function (e.g., different types of explosives and different types of guns). This highly varied target set allows for more nuanced empirical examples when comparing how the type of found target might influence each SSM measurement. The first comparisons for this investigation will be primarily from the Mitroff et al. (2015) analysis concerning SSM error rates with two identical targets (see Fig.3), and the other comparisons will primarily follow from the Biggs et al. (2015) analyses regarding the impact of the first found target.

For the identical target analyses (see Fig. 4), the original results demonstrated significant reductions in SSM error rates when two targets were identical versus nonidentical using the baseline method for SSM error calculations. Additionally, the baseline method and the weighted method are identical for these analyses because all targets are identical. The baseline method yielded a significant SSM error rate ($M = 6.53 \%$, $SE = 1.62 \%$), $t(32) = 4.04$, $p < .001$, Cohen’s $d = 1.43$, and the adjusted method yielded a qualitatively similar SSM error rate ($M = 6.92 \%$, $SE = 2.47 \%$), $t(32) = 2.80$, $p < .001$, Cohen’s $d = 0.99$. These rates are still well below the SSM error rates for nonidentical targets observed in the original study ($M = 19.21 \%$, $SE = 1.36 \%$), and so either method would suggest a significant reduction, but not an elimination, of SSM errors.

However, the dependent method yielded a very different interpretation ($M = -4.24 \%$, $SE = 2.22 \%$), $t(32) = 1.91$, $p = .03$, Cohen’s $d = .67$. Not only was there a significant difference in SSM errors according to the dependent method, but the difference actually went in the opposite direction—searchers demonstrated significant improvements in search accuracy following a found target when the subsequent target matched the first identically. It is also worth noting that all methods yield moderate or larger effect sizes despite the interpretations being substantially different. One possible reason might be the influence of single-target accuracy, which factors differently into each formula. All three methods are positively related to single-target accuracy, but the correlations are smaller for both the baseline method, $r(31) = .37$, $p = .03$, and the

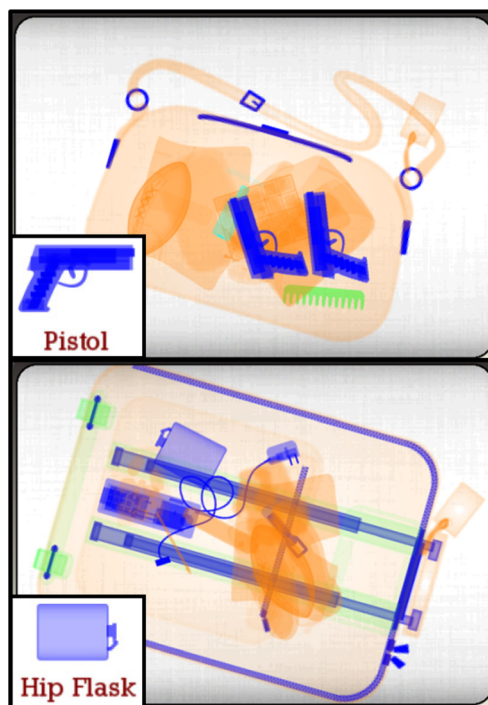


Fig. 3 Examples from the mobile application Airport Scanner used to collect the data presented in Biggs, Adamo, Dowd, and Mitroff (2015) and Mitroff et al. (2015). The top image presents a dual-target bag with two identical targets (both the *pistol* target type). The bottom image presents a single-target bag with a *hip flask* target

adjustment method, $r(31) = .31$, $p = .07$, than the dependent method, $r(31) = .69$, $p < .001$. These differences suggest an asynchrony between the measures and how much low single-target accuracy rates can influence the SSM error measurement.

This influence becomes more potent when the targets are not identical. For example, two targets can vary substantially in single-target accuracy, such as pistol ($M = 94.66 \%$) and hip flask ($M = 66.88 \%$). Both targets share the same color (blue),

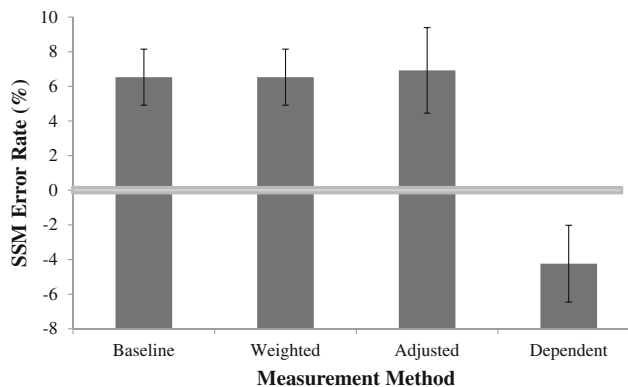


Fig. 4 SSM error rates (in percentage with standard error bars) as assessed by the four different measurement methods for the Mitroff et al. (2015) analysis involving identical targets. Please note that the baseline and weighted methods involve identical calculations in this particular circumstance, and the gray bar crossing the figure indicates a zero difference, or what would be considered insignificant SSM errors

and when eliminating the targets with a conceptual overlap (i.e., no guns or gun-related items to avoid a conceptual relationship with the pistol), the remaining data provides some insight into how found-target accuracy can influence the different SSM measures without the confounding influence of perceptual or conceptual set biases. Twenty-one second targets remain with no conceptual relationship to the target and enough trials to provide a valid SSM measure with both the hip flask found first and the pistol found first. See Table 2 for the results.

With the pistol as the first target, all three relevant methods yielded significant SSM error rates—baseline method: $M = 18.93\%$, $SE = 2.62\%$, $t(20) = 7.22$, $p < .001$, Cohen's $d = 3.23$; adjusted method: $M = 26.15\%$, $SE = 5.07\%$, $t(20) = 5.16$, $p < .001$, Cohen's $d = 2.31$; dependent method: $M = 15.08\%$, $SE = 2.45\%$, $t(20) = 6.15$, $p < .001$, Cohen's $d = 2.75$. With the hip flask as the first target, both the baseline method ($M = 17.35\%$, $SE = 2.79\%$, $t(20) = 6.21$, $p < .001$, Cohen's $d = 2.78$), and the adjusted method ($M = 28.32\%$, $SE = 4.13\%$, $t(20) = 6.86$, $p < .001$, Cohen's $d = 3.07$), yielded significant SSM errors. However, the dependent method yielded a different interpretation ($M = -6.51\%$, $SE = 3.12\%$, $t(20) = 2.09$, $p < .001$, Cohen's $d = 0.93$). Once again, the dependent method suggests a significant improvement in accuracy following the searcher finding a hip flask; only, unlike the identical targets analysis, this instance comes without any theoretical basis for the improvement. The weighted method, which accounts differently for SSM errors based on the number of contributing cases, also yielded significant SSM error rates ($M = 18.93\%$, $SE = 2.41\%$, $t(20) = 7.86$, $p < .001$, Cohen's $d = 3.52$).

The adjusted method is noteworthy here for its much higher variance. Specifically, the adjusted measures differed substantially based on the second target item. Some individual items elicited relatively small SSM error rates (crowbar: $M = 9.41\%$; wine bottle: $M = 8.91\%$), whereas others had extremely low SSM error rates (blow gun: $M = -54.65\%$). The blow gun in particular represents an important issue with the adjustment method because it has very low single-target accuracy (13.86%), which significantly inflates the 7.57% change into a 54.65% change. The formula thus preserves the size of the change for items with relatively high single-target accuracy, although it significantly inflates any differences for items with relatively low single-target accuracy.

General Discussion

Multiple-target search introduces a unique source of error not present in single-target searches. Specifically, searchers have to continue looking for additional targets after locating a first target, which creates the possibility for additional errors during subsequent search, or SSM errors (Adamo et al., 2013).

Measuring subsequent search errors introduces additional challenges above and beyond simple accuracy measurements in single-target search, and across several different studies, many different methods have been used. This investigation used four different measurement methods (baseline, weighted, adjusted, and dependent) to contrast and compare the advantages and disadvantages of each method.

The baseline method yielded several notable advantages. Foremost, it offers greater simplicity than the other methods, which provides numerous benefits in convenience alone. It is also relatively straightforward to account for differences in the found target type while using the baseline method—error calculations can be limited only to those trials with a specific type of target found first for the multiple-target search accuracy rate. The greatest disadvantage of the baseline method is that it carries an inherent ceiling effect. For example, single-target accuracy of 20% could have a maximum SSM error rate of 20%. Thus, although the baseline method controls for relative differences between target types due to single-target accuracy, the baseline method is not ideal for situations where extremely low accuracy would be expected from most target types.

The weighted method, meanwhile, holds two particular advantages over the baseline method. First, the weighted method captures more data by accounting for asynchronies in target types. For example, nonprofessional searchers were more likely to locate a high-salience target first than professional searchers (Biggs & Mitroff, 2014). Overall SSM error calculations could be different between groups because each group contributes a significantly different number of trials into the calculation. The weighted method addresses this discrepancy by limiting the influence of a particular target type to the relative portion it contributes into the overall SSM error calculation. Second, the weighted method captures a more complete view of the data while also accounting for baseline differences. More trials are factored into the SSM calculations, which can increase the reliability of the measurement. The primary disadvantage is that a target type commonly found first can dramatically bias the overall SSM results and corresponding conclusions in favor of that particular target type. The weighted approach also accounts for the broad number of trials contributing to the SSM error method, which could also limit any investigation into the influence of a particular found target type. For example, a conceptual set bias suggests that searchers look for additional targets with a similar function to the found target (Biggs et al., 2015). If a particular target type factors more trials into the measurement method than another target type, it would bias the SSM calculation in favor of the target type contributing more trials without any regard to the theoretical implications. Another disadvantage involves situations with many different target types. For example, a search with only two target types presents limited found target and second target combinations, whereas a search with dozens of different targets yields an exponentially larger number of

Table 2 SSM errors in Airport Scanner using each measurement method and divided by the first found target. Means are presented in the upper half of the table (with standard errors in parentheses), and effect sizes are presented in the lower half of the table. Effect sizes are the Cohen's *d* for the corresponding one-sample *t* test with a hypothesized mean of zero

	Found Target		
	Pistol	Hip flask	Combined
<i>Means</i>			
Baseline	18.93 % (2.62 %)	17.35 % (2.79 %)	-
Weighted	-	-	18.93 % (2.41 %)
Adjusted	26.15 % (5.07 %)	28.32 % (4.13 %)	-
Dependent	15.08 % (2.45 %)	-6.51 % (3.12 %)	-
<i>Effect sizes</i>			
Baseline	3.23	2.78	-
Weighted	-	-	3.52
Adjusted	2.31	3.07	-
Dependent	2.75	-0.93	-

possible target combinations. The end result is an unwieldy calculation that may overcomplicate the situation beyond what the researcher intended to examine. As such, the weighted method remains an ideal alternative to the baseline method when: the search includes only a small number of possible targets, one particular target type is found significantly more often than another target type, and the primary goal involves assessing overall SSM error rates without regard to a potential influence on subsequent search of any one particular target type.

The adjusted method offers a possible alternative to the baseline and the weighted method that could specifically apply to visual search scenarios with low accuracy rates. Specifically, it is designed to adjust the observed difference so that the SSM calculation better reflects the change in accuracy relative to the baseline, single-target observation. The greatest advantage of this method is that it could potentially account for relative variance in a way neither the baseline nor weighted methods could. Additionally, because the SSM calculation is determined relative to the single-target search accuracy, it does not have the inherent ceiling effect of the baseline method. Unfortunately, the largest disadvantage is that it may overinflate SSM error calculations. Across all analyses reported here, the adjusted method yielded the largest SSM error rates. Because standard deviations were also inflated, the effect sizes were not impacted as much as the overall means. Another consideration is that the inflated variance applies primarily to target types with low accuracy while having relatively no impact on target types with high accuracy. Thus, the adjusted method only seems appropriate when the search task only contains target types with low baseline accuracy rates to counter the ceiling effects inherent to the baseline method.

The fourth proposed measurement was the dependent method. This approach could theoretically address the problem of inflated error variance from a different direction than the adjusted method. Specifically, the dependent method takes the multiple-target aspect into account and establishes a

baseline rate that accounts for variability in both the found target and the second target. Its greatest advantage is that it may avoid inflated effect sizes and could provide a more viable means to control error variance as contributed by both targets rather than primarily calculating SSM errors off a single target type. However, this method appears to overadjust and could yield deceptively low SSM error rates due to the inflated influence of the first found target. If the found target has high baseline accuracy (e.g., the pistol form Airport Scanner), then the method appears to be a viable option. If the found target has a lower accuracy rate, then this method might create a baseline estimate so low that observing significant SSM errors becomes highly unlikely. Thus, this measurement method would best be reserved for special cases where any found target has relatively high accuracy and the goal is to control for inflated variance.

In addition to which particular method should be used, it is important to consider the situational-specific variables and determine whether it is even appropriate to measure SSM errors. Notably, multiple types of targets are neither a necessary nor sufficient condition for an investigation into subsequent search misses. Both datasets used here included multiple target types; albeit the “T”s and “L”s search (Biggs & Mitroff, 2014) only varied target salience, whereas Airport Scanner (Biggs et al., 2015; Mitroff et al., 2015) uses dozens of different target types. However, some investigations accurately assess SSM errors using only a single-target type (Adamo et al., 2016). Multiple target types can dramatically influence visual search, although insofar as SSM errors are concerned, multiple target types largely complicate the possible measurements.

The most important aspect of any SSM investigation is that the display can contain multiple targets within a single search display. Granted, the ratio of single-target trials to multiple-target trials will directly affect the likelihood of observing SSM errors (e.g., Fleck et al., 2010). The ratio can alter searchers' expectations, which in turn will affect search

behaviors. The more important delineation involving multiple-target search involves multiple-target search versus foraging. This distinction is due to the nature of quitting behaviors in multiple-target search versus foraging. In multiple-target search investigations where one might expect to observe SSM errors, the quitting rule needs to be exhaustive—that is, continue searching until all targets have been found. Participants thus terminate search under the impression that no more targets were present in the display. By comparison, foraging search commonly involves nonexhaustive search—that is, choosing to terminate search despite knowing that additional targets are present. Foraging termination rules often involve factors such as the current or recent yield rates for

targets (e.g., number of berries picked from a bush; Wolfe, 2012b), but participants will generally quit when return rates fall below an optimal level. This specific instance is substantially different from an exhaustive, multiple-target search such as airport security screenings or radiology. In these examples, it is critical that the searcher find all targets, and while various factors can impact when the searcher chooses to terminate search (Biggs & Mitroff, 2015a), the underlying principle remains that the searcher does not terminate search with the knowledge that more targets are present in the display.

To aid in visualizing these differences between the methods, consider the graphs in Fig. 5. Each demonstrates how much the measurement method can impact the exact

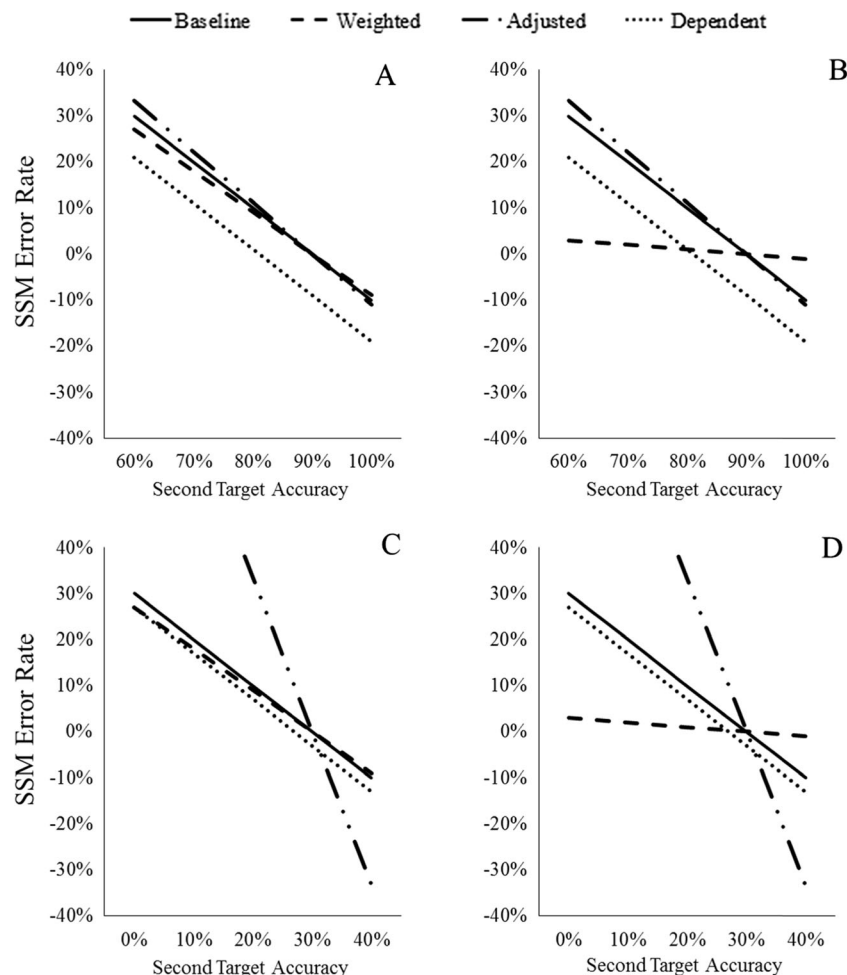


Fig. 5 Four hypothetical scenarios are depicted for each of the four measurement methods: baseline, weighted, adjusted, and dependent. For all scenarios, the *x*-axis represents second target accuracy (after a first target had been found), the *y*-axis represents SSM error rates, and the accuracy rate for the other target type not depicted is 90 %, whether found first or second during the search. The top two scenarios (A and B) presume a relatively high accuracy for the target type during single-target trials (90 %) and demonstrate how the SSM measurement methods change as the accuracy for that target type changes from 60 % to 100 % during subsequent search after a found target. The bottom two scenarios (C and D) presume a relatively low accuracy for the target during single-

target trials (30 %) and demonstrate how the SSM measurement methods change as the accuracy for that target type changes from 0 % to 40 % during subsequent search after a found target. The left two scenarios (A and C) demonstrate the change in SSM error rate when the other target type not depicted was found first on 90 % of multiple-target trials (i.e., the depicted second target accuracy contributes 90 % of the SSM calculations). The right two scenarios (B and D) demonstrate the change in SSM error rate when the depicted type was found first on 90 % of the multiple-target trials (i.e., the depicted second target accuracy contributes 10 % of the SSM calculations)

same hypothetical data. The baseline method appears to be the most consistent measurement across all cases, and the dependent method appears to be similarly consistent, but more conservative. The weighted method always takes the most data into account as it provides an overall SSM error rate based on multiple target types. However, it also biases the overall SSM error rate based upon which target type was found first more often. This caveat makes using the weighted method specific to certain situations where the researcher intends to collect an overall error rate across multiple target types despite any bias based upon target type. Finally, the adjusted method is easily the most liberal measurement method. This wider range could be very useful when dealing with a sample of very low single-target accuracy rates, but for most instances, it appears more likely that the adjusted measurement method will yield an overestimation of SSM error rates.

The best and most reliable approach for future interpretations may then be to use both the baseline method and the dependent method when appropriate and possible. They produce very similar results with the dependent method being the more conservative of the two. In this way, the converging evidence could help differentiate between a “passive” SSM error and an “active” SSM error. The baseline measurement might be best at determining the whether there is any performance decline during subsequent search by establishing a baseline comparison from single-target search. These SSM errors could be deemed “passive” as they could be due to theoretical differences or expected, less-than-perfect performance rates—that is, similar to an expected accuracy rate of less than 100 % in single-target search, subsequent search could experience the same decline. Thus, passive SSM errors could be the result of compounded error rates across multiple periods of search performance, and the baseline method is the best approach for a simple determination of whether SSM errors are present in a given scenario. Conversely, the dependent method could establish the line for active SSM errors—that is, the dependent method is so conservative that any significant change in SSM errors is almost certainly due to a theoretical and distinguishable impact upon the search scenario. This possibility is evident even in the Airport Scanner data depicted in Fig. 4. If the second target is identical to the first target, then there are likely to be strong priming effects (Huang, Holcombe, & Pashler, 2004; Kristjánsson & Campana, 2010; Maljkovic & Nakayama, 1994). This theoretical difference is reflected in the strong reduction of SSM errors for nonidentical versus identical targets as reported in Biggs et al. (2015). The baseline method thus establishes a reduction, and the dependent method further highlights a theoretical cause for the reduction in SSM errors. In cases where one of the two methods might not be feasible though, researchers should—as always—

consider differences in effect sizes when attempting to reach any conclusions.

In summary, there are several guidelines to consider when determining whether the task is subject to SSM errors, and what type of measurement to use:

1. It must be a multiple-target search. Multiple-category searches or hybrid searches, which use different types of targets in the target set, are neither necessary nor sufficient to investigate SSM errors.
2. In addition to multiple-target trials, single-target trials are necessary to establish expected accuracy rates for this particular target type.
3. While searching a particular display during multiple-target trials, at least one target must be found to initiate the subsequent search period.
4. The search must use an exhaustive search termination rule, where the goal is to find all targets present.
5. Different measurement methods can be used, although using a method other than the baseline method should have an empirically driven rationale. When possible, use both the baseline method and the dependent method to differentiate between passive and active SSM errors.

For example, a search could have only one target type and still qualify as a “multiple-target search” if a single display, not the target set, can contain multiple targets. Also, not every search display needs to contain multiple targets. Searchers’ expectations have already been shown to significantly alter the likelihood of observing SSM errors (Fleck et al., 2010), and the ratio of single-target to multiple-target trials remains an important manipulation to consider in multiple-target search. As for the specific measurement method, the baseline method appears to be the most direct and convenient approach to measuring SSM errors. However, as specific situations arise, there are alternative methods that could account for certain limitations of the baseline method.

Author’s Note This manuscript includes data that are published elsewhere (Biggs, Adamo, Dowd, & Mitroff, 2015; Biggs & Mitroff, 2014; Mitroff et al., 2015). These data are presented again here with new analyses to contrast and compare the results against the previously published results for illustrative purposes. I would like to thank Jeremy Wolfe, Hayward Godwin, and an anonymous reviewer for their comments on the original version. The author has no financial or nonfinancial competing interests in this manuscript. The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Department of the Navy, Department of Defense, nor the U.S. Government. I am a military service member (or employee of the U.S. Government). This work was prepared as part of my official duties. Title 17 U.S.C. §105 provides that “Copyright protection under this title is not available for any work of the United States Government.” Title 17 U.S.C. §101 defines a U.S. Government work as a work prepared by a military service member or employee of the U.S. Government as part of that person’s official duties.

References

- Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2013). Self-induced attentional blink: A cause of errors in multiple-target visual search. *Psychological Science, 24*, 2569–2574.
- Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2016). *Satisfaction at last: Evidence for the “satisfaction” hypothesis for multiple-target search errors*. Manuscript submitted for publication.
- Anbari, M. M. (1997). Cervical spine trauma radiography: Sources of false-negative diagnoses. *Emergency Radiology, 4*, 218–224.
- Berbaum, K. S., Dorfman, D. D., Franken, E. A., Jr., & Caldwell, R. T. (2000). Proper ROC analysis and joint ROC analysis of the satisfaction of search effect in chest radiography. *Academic Radiology, 7*, 945–958.
- Berbaum, K. S., Franken, E. A., Jr., Caldwell, R. T., & Scharzt, 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). Cambridge, UK: Cambridge University Press.
- Berbaum, K. S., Franken, E. A., Jr., Dorfman, D. D., Rooholamini, S. A., Coffman, C. E., Cornell, S. H.,... & Smith, T. P. (1991). Time course of satisfaction of search. *Investigative Radiology, 26*, 640–648.
- 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*, 844–855.
- Biggs, A. T., Adamo, S. H., & Mitroff, S. R. (2014). Rare, but obviously there: Effects of target frequency and salience on visual search accuracy. *Acta Psychologica, 152*, 158–165.
- 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 searchers. *Visual Cognition, 21*, 330–352.
- Biggs, A. T., Kreager, R. D., Gibson, B. S., Villano, M., & Crowell, C. R. (2012). Semantic and affective salience: The role of meaning and preference in attentional capture and disengagement. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 531–541.
- Biggs, A. T., & Mitroff, S. R. (2014). Different predictors of multiple-target search accuracy between nonprofessional and professional visual searchers. *The Quarterly Journal of Experimental Psychology, 67*(7), 1335–1348.
- Biggs, A. T., & Mitroff, S. R. (2015a). Differences in multiple-target visual search performance between non-professional and professional searchers due to decision-making criteria. *British Journal of Psychology, 106*, 551–563.
- Biggs, A. T., & Mitroff, S. R. (2015b). Improving the efficacy of security screening tasks: A review of visual search challenges and ways to mitigate their adverse effects. *Applied Cognitive Psychology, 29*, 142–148.
- Boettcher, S. E., Drew, T., & Wolfe, J. M. (2013). Hybrid search in context: How to search for vegetables in the produce section and cereal in the cereal aisle. *Visual Cognition, 21*, 678–682.
- Boettcher, S. E., & Wolfe, J. M. (2015). Searching for the right word: Hybrid visual and memory search for words. *Attention, Perception, & Psychophysics, 77*, 1132–1142.
- Cain, M. S., Adamo, S. H., & Mitroff, S. R. (2013). A taxonomy of errors in multiple-target visual search. *Visual Cognition, 21*, 899–921.
- Cain, M. S., Biggs, A. T., Darling, E. F., & Mitroff, S. R. (2014). A little bit of history repeating: Splitting up multiple-target visual searches decreases second-target miss errors. *Journal of Experimental Psychology: Applied, 20*, 112–125.
- 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.
- 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.
- 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.
- Charnov, E. (1976). Optimal foraging: The marginal value theorem. *Theoretical Population Biology, 9*, 129–136.
- 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.
- Cosman, J. D., & Vecera, S. P. (2009). Perceptual load modulates attentional capture by abrupt onsets. *Psychonomic Bulletin & Review, 16*, 404–410.
- Cosman, J. D., & Vecera, S. P. (2010a). Attentional capture under high perceptual load. *Psychonomic Bulletin & Review, 17*, 983–986.
- Cosman, J. D., & Vecera, S. P. (2010b). Attentional capture by motion onsets is modulated by perceptual load. *Attention, Perception, & Psychophysics, 72*, 2096–2105.
- Dougherty, M. R., Harbison, J. I., & Davelaar, E. J. (2014). Optional stopping and the termination of memory retrieval. *Current Directions in Psychological Science, 23*, 332–337.
- Drew, T., Boettcher, S. E., & Wolfe, J. M. (2015). Searching while loaded: Visual working memory does not interfere with hybrid search efficiency but hybrid search uses working memory capacity. *Psychonomic Bulletin & Review, 1–12*. doi:10.3758/s13423-015-0874-8
- Drew, T., & Wolfe, J. M. (2014). Hybrid search in the temporal domain: Evidence for rapid, serial logarithmic search through memory. *Attention, Perception, & Psychophysics, 76*, 296–303.
- Ehinger, K. A., & Wolfe, J. M. (2016). When is it time to move to the next map? Optimal foraging in guided visual search. *Attention, Perception, & Psychophysics*. doi:10.3758/s13414-016-1128-1
- Fleck, M. S., & Mitroff, S. R. (2007). Rare targets are rarely missed in correctable search. *Psychological Science, 18*, 943–947.
- 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.
- Fougnie, D., Cormiea, S. M., Zhang, J., Alvarez, G. A., & Wolfe, J. M. (2015). Winter is coming: How humans forage in a temporally structured environment. *Journal of Vision, 15*(11), 1. doi:10.1167/15.11.1
- Godwin, H. J., Menneer, T., Cave, K. R., & Donnelly, N. (2010). Dual-target search for high and low prevalence X-ray threat targets. *Visual Cognition, 18*, 1439–1463.
- Godwin, H. J., Menneer, T., Cave, K. R., Helman, S., Way, R. L., & Donnelly, N. (2010). The impact of relative prevalence on dual-target search for threat items from airport X-ray screening. *Acta Psychologica, 134*, 79–84.
- Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception, & Psychophysics, 72*, 1267–1282.
- Hout, M. C., & Goldinger, S. D. (2012). Incidental 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.
- Huang, L., Holcombe, A. O., & Pashler, H. (2004). Repetition priming in visual search: Episodic retrieval, not feature priming. *Memory & Cognition, 32*, 12–20.
- Irons, J. L., Folk, C. L., & Remington, R. W. (2012). All set! Evidence of simultaneous attentional control settings for multiple target colors. *Journal of Experimental Psychology: Human Perception and Performance, 38*, 758.
- Lavie, N., & Cox, S. (1997). On the efficiency of attentional selection: Efficient visual search results in inefficient distractor rejection. *Psychological Science, 8*, 395–398.

- Kristjánsson, Á., & Campana, G. (2010). Where perception meets memory: A review of repetition priming in visual search tasks. *Attention, Perception, & Psychophysics*, *72*, 5–18.
- Kristjánsson, Á., Jóhannesson, Ó. I., & Thornton, I. M. (2014). Common attentional constraints in visual foraging. *PLoS ONE*, *9*, e100752.
- Krupinski, E. A. (2010). Current perspectives in medical image perception. *Attention, Perception, & Psychophysics*, *72*, 1205–1217.
- Kuhn, G. J. (2002). Diagnostic errors. *Academic Emergency Medicine*, *9*, 740–750.
- Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: I. Role of features. *Memory & Cognition*, *22*, 657–672.
- Menner, T., Barrett, D. J., Phillips, L., Donnelly, N., & Cave, K. R. (2007). Costs in searching for two targets: Dividing search across target types could improve airport security screening. *Applied Cognitive Psychology*, *21*, 915–932.
- Menner, T., Cave, K. R., & Donnelly, N. (2009). The cost of search for multiple targets: Effects of practice and target similarity. *Journal of Experimental Psychology: Applied*, *15*, 125–139.
- Menner, T., Donnelly, N., Godwin, H. J., & Cave, K. R. (2010). High or low target prevalence increases the dual-target cost in visual search. *Journal of Experimental Psychology: Applied*, *16*, 122–144.
- Mitroff, S. R., & Biggs, A. T. (2014). The Ultra-Rare-Item effect: Visual search for exceedingly rare items is highly susceptible to error. *Psychological Science*, *25*, 284–289.
- Mitroff, S. R., Biggs, A. T., Adamo, S. H., Dowd, E. W., Winkle, J., & Clark, K. (2015). What can 1 billion trials tell us about visual search? *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 1–5.
- Nakashima, R., Kobayashi, K., Maeda, E., Yoshikawa, T., & Yokosawa, K. (2013). Visual search of experts in medical image reading: The effect of training, target prevalence, and expert knowledge. *Frontiers in Psychology*, *4*, 166. doi:10.3389/fpsyg.2013.00166
- Russell, N. C., & Kunar, M. A. (2012). Colour and spatial cueing in low-prevalence visual search. *The Quarterly Journal of Experimental Psychology*, *65*, 1327–1344.
- Smith, M. J. (1967). *Error and variation in diagnostic radiology*. Springfield, IL: Thomas.
- Stroud, M. J., Menner, T., Cave, K. R., & Donnelly, N. (2012). Using the dual-target cost to explore the nature of search target representations. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 113–122.
- Tuddenham, W. J. (1962). Visual search, image organization, and reader error in roentgen diagnosis. *Radiology*, *78*(5), 694–704.
- Voytovich, A. E., Rippey, R. M., & Suffredini, A. (1985). Premature conclusions in diagnostic reasoning. *Academic Medicine*, *60*, 302–307.
- Wolfe, J. M. (2012a). Saved by a log how do humans perform hybrid visual and memory search? *Psychological Science*, *23*, 698–703.
- Wolfe, J. M. (2012b). When is it time to move to the next raspberry bush? Foraging rules in human visual search. *Journal of Vision*, *13*(3), 10. doi:10.1167/13.3.10
- Wolfe, J. M., Aizenman, A. M., Boettcher, S. E., & Cain, M. S. (2016). Hybrid foraging search: Searching for multiple instances of multiple types of target. *Vision Research*, *119*, 50–59.
- Wolfe, J. M., Boettcher, S. E., Josephs, E. L., Cunningham, C. A., & Drew, T. (2015). You look familiar, but I don't care: Lure rejection in hybrid visual and memory search is not based on familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 1576–1587.
- Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature*, *435*, 439–440.
- Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General*, *136*, 623–638.
- Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, *20*, 121–124.
- Zhang, J., Gong, X., Fougny, D., & Wolfe, J. M. (2015). Using the past to anticipate the future in human foraging behavior. *Vision Research*, *111*, 66–74.