British Journal of Psychology (2014)

The British Psychological Society

www.wileyonlinelibrary.com

Differences in multiple-target visual search performance between non-professional and professional searchers due to decision-making criteria

Adam T. Biggs* and Stephen R. Mitroff

Department of Psychology and Neuroscience, Center for Cognitive Neuroscience, Duke University, Durham, North Carolina, USA

Professional visual searches, such as those conducted by airport security personnel, often demand highly accurate performance. As many factors can hinder accuracy, it is critical to understand the potential influences. Here, we examined how explicit decision-making criteria might affect multiple-target search performance. Non-professional searchers (college undergraduates) and professional searchers (airport security officers) classified trials as 'safe' or 'dangerous', in one of two conditions. Those in the 'one = dangerous' condition classified trials as dangerous if they found one or two targets, and those in the 'one = safe' condition only classified trials as dangerous if they found two targets. The data suggest an important role of context that may be mediated by experience; non-professional searchers were more likely to miss a second target in the one = dangerous condition (i.e., when finding a second found target did not change the classification), whereas professional searchers were more likely to miss a second in the one = safe condition.

Visual search, the act of finding targets amongst distractors, is a common task that people perform numerous times each day. Searches can be mundane, such as a child finding her shoes on the way out of the house, or far more critical, such as an airport security screener looking for bomb parts concealed in luggage. Decades of laboratory-based research have focused on delineating the underlying nature of visual search, and the vast majority of these studies have implemented a single-target search scenario – a given search can have zero or one target present, but never more (for recent reviews, see Eckstein, 2011; Nakayama & Martini, 2011). However, many searches outside of the laboratory can contain multiple targets, and the searcher cannot simply quit looking after locating one target – the child looking for her shoes and the airport security screener must continue searching after having found a first target.

Because search must continue following a found target in a multiple-target visual search scenario, there are potential sources of error above and beyond what might be present in single-target searches (Berbaum, 2012). Specifically, it is a well-established phenomenon that finding one target can cause a searcher to miss additional targets in the display. Such errors, referred to as 'satisfaction of search' errors (SOS; Smith, 1967;

**Correspondence should be addressed to Adam T. Biggs, Department of Psychology and Neuroscience, Center for Cognitive Neuroscience, Duke University, B203 LSRC, Box 90999, Durham, NC 27708, USA (email: adam.biggs@duke.edu).*

DOI:10.1111/bjop.12096

Tuddenham, 1962), have been suggested to account for nearly a third of errors in radiology (Anbari, 1997; Berbaum, Franken, Caldwell, & Schartz, 2010) and a majority of errors in diagnostic medicine, where they are known as 'premature closure' (Kuhn, 2002; Voytovich, Rippey, & Suffredini, 1985). The original explanation – and the source of the name – suggested that searchers discontinued search because finding one target 'satisfied' their understanding about the meaning of the display. However, the original 'satisfaction' explanation cannot fully account for the phenomenon as searchers do continue searching after finding a target (Berbaum, Dorfman, Franken, & Caldwell, 2000; Berbaum *et al.*, 1991; Cain, Adamo, & Mitroff, 2013; Fleck, Samei, & Mitroff, 2010).

Research from both radiology and cognitive psychology has suggested that SOS errors are not due to any single underlying cause, and instead stem from a multi-faceted set of factors (e.g., Berbaum, 2012; Berbaum *et al.*, 2010; Cain *et al.*, 2013). For example, SOS errors have been linked to changes in attentional and/or perceptual focus (e.g., Berbaum *et al.*, 2010; Fleck *et al.*, 2010), and to limits on working memory (e.g., Cain & Mitroff, 2013). A recent paper used eye-tracking metrics to distinguish between several types of SOS errors, including scanning errors, where the second target was never fixated, and decision errors, where the target was fixated for an appropriate interval but never reported (Cain *et al.*, 2013). Given these varied explanations for the phenomenon, it is clear that a 'satisfaction' account does not sufficiently describe SOS errors, and as such, we support the more mechanistic-agnostic label of 'subsequent search misses' (SSM; Adamo, Cain, & Mitroff, 2013; Cain *et al.*, 2013). We will thus refer to these errors as 'SSM errors' for the remainder of this article.

One core goal of SSM research is to minimize these dangerous errors, and recent research has offered intriguing advances by demonstrating that situational factors influence SSM rates. For example, one study demonstrated that simply changing the searchers' global understanding of what it means to complete a task could impact performance (Clark, Cain, Adcock, & Mitroff, 2013). In that study, participants completed a multiple-target search task in one of two conditions – fixed goal or fixed duration. The fixed goal condition was meant to approximate radiological search parameters, and participants were required to perform a fixed number of visual searches to complete the study (akin to a set number of cases radiologists typically examine in a normal work day). Alternatively, the fixed duration condition was meant to approximate airport security search, and participants searched for a fixed time period (akin to timed shifts as the X-ray operator at an airport checkpoint). Despite identical tasks, the global framework differences led to fewer SSM errors in the fixed duration condition (Clark *et al.*, 2013). Thus, certain procedures – peripheral to how search is conducted during the actual search task – could be adopted to reduce SSM errors.

The goal of this study was to further explore efforts to minimize SSM errors while simultaneously informing cognitive theories of search. Specifically, when searchers inspect a display, they have an overarching framework that is specific to their goal at hand. For example, searchers tasked with finding each and every possible target – without any room for error – will have a different quitting threshold (e.g., Chun & Wolfe, 1996; Wolfe & Van Wert, 2010) than searchers tasked with finding as many targets as quickly as they can. Our hypothesis is that such global instructions and criteria may have an especially large impact on SSM errors because they may create a contextual influence on the search itself (cf., Clark *et al.*, 2013). If global instructions and decision-making criteria significantly impact search performance, then these instructions could provide both practical and theoretical insights into multiple-target search. Practically, finding that different decision criteria can influence SSM errors could impact how professional searchers are asked to perform a given task. Search performance could be improved simply by changing how search goals are labelled and presented. Theoretically, it has been shown that quitting threshold decisions are more complicated in multiple-target searches compared to single-target searches (Cain, Vul, Clark, & Mitroff, 2012; Wolfe, 2013). Understanding how global decision criteria can impact multiple-target search performance may offer a valuable window into the underlying mechanisms that drive SSM errors.

The present study employed a multiple-target visual search task to understand how global decision criteria differences could impact SSM errors. All participants were tasked with searching arrays that could contain zero, one, or two targets, and their goal was to find all targets present. Participants ended a trial by classifying the search display as either safe or dangerous; however, the criteria for judging a trial as safe or dangerous changed based upon the condition to which a participant was randomly assigned. Those in the *one = dangerous* condition were told to classify a trial as 'dangerous' if it contained one or two targets, and those in the *one = safe* condition were told to classify a trial as 'dangerous' only if it contained two targets (zero targets was considered a safe trial in both conditions). Thus, the basic goal and objectives were identical across conditions (i.e., locate all targets), with the only difference being the global criterion used to classify each trial after search was completed.

Importantly, one condition *did not* involve a change in classification upon finding a second target (one = dangerous condition; both one and two found targets indicated a dangerous trial), whereas the other condition *did* involve a change in classification upon finding a second target (one = safe condition; one found target indicated a safe trial and two found targets indicated a dangerous trial). We hypothesized that this criterion manipulation would primarily affect subsequent search (i.e., search after a target is found), with the effects manifesting in changes to accuracy and/or response time. For accuracy, participants may commit fewer SSM errors if finding a second target has the potential to alter their decision in classifying the trial. For search time, participants may spend longer searching after finding a first target if finding a second target changes their decision. Should the results indicate a difference in either accuracy or response timing, it would be evidence that global decision-making criteria can influence how search is conducted – and specifically, whether or not a criterion shift might alter the decision to terminate search.

Given our goals of informing both practical and theoretical aspects of visual search, we tested non-professional (university undergraduates) and professional (airport security officers) searchers. The professional searchers obtained specific and structured training that could impact their susceptibility to differences in decision criteria, so it is possible that they might be less influenced by – or might react differently to – the manipulation employed here. Furthermore, the safe versus dangerous decision utilized in our search task has important parallels to how professional security officers perform visual searches. For example, a prohibited item needs to be removed, but there is a critical difference between treating a found target as something likely to be dangerous (e.g., a gun; or, the one = dangerous condition) and something likely to be benign (e.g., a larger than allowable shampoo bottle; or the one = safe condition).

Methods

Non-professional participants

Forty-two university undergraduate students participated for credit as partial completion of a course requirement, and were randomly assigned to one of two conditions. Twenty-one were assigned to the one = dangerous condition (11 female; $M_{age} = 18.95$, $SD = 0.83$), and 21 were assigned to the one = safe condition (13 female; $M_{age} = 18.85$, $SD = 0.81$). One participant was excluded from analyses in the one = safe condition for failing to complete the experiment. Our data outlier criterion was that single-target accuracy must be within 3 *SD* of the group mean, and no participants' data were eliminated due to this criterion.

Professional participants

Thirty-eight Transportation Security Administration (TSA) Officers participated and were randomly assigned to either the one = dangerous or one = safe condition. The TSA Officers participated in the study in a quiet area away from the checkpoint while at work. Participants were given the option of allowing their anonymous data to be used only for TSA purposes or for both TSA and research purposes. One participant from the one = dangerous condition did not provide consent for the data to be used for research purposes. Data from two participants in the one = dangerous condition were excluded from analyses due to single-target accuracy more than 3 *SD* below the group mean. We also limited our analyses to TSA Officers who conducted X-ray image searches as a regular part of their daily activities (some TSA Officers have other duties that do not primarily involve X-ray searches). Six TSA Officers reported that they did not conduct X-ray searches as a part of their regular duties (five in the one = dangerous condition, one in the one = safe condition). The 29 participants remaining in the analysis (11 in the one = dangerous condition, 18 in the one = safe condition) reported their ages via age range options (18–25 years: $N = 2$; 26–34 years: $N = 6$; 35–49 years: $N = 6$, 50–65 years: $N = 10$; 66+: $N = 2$; three missing responses; nine female).

Stimuli

Stimuli were based upon prior studies (e.g., Biggs & Mitroff, 2014; Cain, Dunsmoor, LaBar, & Mitroff, 2011; Clark *et al.*, 2013; Fleck *et al.*, 2010). Displays included 25 total items arranged upon an invisible grid (8×7 positions) with each item randomly offset 0–10 pixels from perfect grid alignment (see Figure 1). Targets were perfect 'T' shapes and appeared in one of two salience levels (high salience: 57–65% black; low salience: 22–45% black). Distractors were non-T shapes drawn from the same salience ranges. Each item was composed of two rectangles slightly separated and oriented perpendicularly (approximate width of 0.3° of visual angle and length of 1.0° for each item, or $1.3° \times 1.3°$ at its widest point). Each item appeared in one of four possible rotations on a background of grey 'clouds' (4–37% black).

Apparatus

Non-professional participants were tested on either a Dell Inspiron computer with 20-inch CRT monitor or a Dell Vostro 260 computer with 23.6-inch widescreen LCD monitors. Professional participants were tested at the airport inside a small laboratory

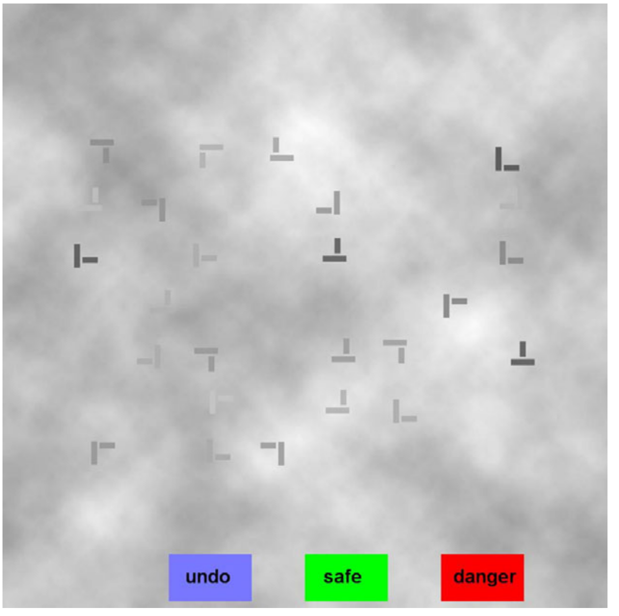

Figure 1. Sample stimulus display with one high-salience target on the right side of the display.

away from the security checkpoints, on Dell Vostro 260 computers with 23.6-inch widescreen LCD monitors. The computer displays presented the same physical display sizes for non-professional and professional participants, and all participants were seated approximately 57 cm from the screen without head restraint. All testing stations used Matlab software (The MathWorks, Natick, MA, USA) and the Psychophysics Toolbox version 3.0.8 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) for experimental presentation and data collection.

Procedure

Each participant completed a practice block of 25 trials before 255 experimental trials, which were presented in a unique random order for each participant. The experimental trials were comprised of 125 with a single, high-salience target; 40 with a single, low-salience target; 40 with both a high-salience and a low-salience target (dual-target trials); and 50 trials with no target present. The trial type distribution was based upon previous work (e.g., Fleck *et al.*, 2010) that demonstrated significant SSM effects. Experimental trials were separated into five blocks of 51 trials, and participants were allowed to take breaks between blocks as desired.

Participants were instructed to search each display for target 'T' shapes and to make a mouse click on each found target. A click within a 35-pixel radius (1.14°) of the centre of an item was considered a click on that item. To end the trial, participants clicked on one of two buttons located on the screen, but beneath the search stimuli (see Figure 1). The buttons were labelled either 'Safe' printed in black against a green button or 'Danger' printed in black against a red button. Zero found targets always indicated a 'Safe' trial, and two found targets always indicated a 'Danger' trial. In the one = dangerous condition, one found target indicated a 'Danger' trial. In the one = safe condition, one found target indicated a 'Safe' trial. Participants used the computer mouse to click on either the 'Safe' or the 'Danger' button and then the subsequent trial automatically began. If neither button was clicked within 15 s, the trial was terminated and a message appeared encouraging the participant to search faster. Feedback on misses and false alarms was given during the practice block, but no feedback was provided during the experimental trials. Displays also included an 'Undo' button at the bottom of the screen, which allowed participants to undo any previous mouse clicks made by accident during that particular trial. Any trials with an 'Undo' button click were excluded from the analyses.

Planned analyses

Subsequent search misses errors were operationally defined as the difference in accuracy on single-target trials compared to accuracy on dual-target trials *after* one target had already been found. That is, we calculated the difference score for low-salience target accuracy between single-target trials and dual-target trials in which the high-salience target was found first, and the difference score for high-salience target accuracy between single-target trials and dual-target trials in which the low-salience target was found first. The baseline accuracy is different between high- and low-salience targets (high-salience targets are found more often, and found first more frequently), so we combined these data by taking a weighted measure of SSM errors that compares accuracy after a found target on dual-target trials to the corresponding single-target baseline. The weights for this measure were determined for each searcher by that individual's percentage of dual-target trials where the high-salience target was found first versus dual-target trials where the low-salience target was found first. Significant SSM errors occur when this weighted accuracy measure is significantly greater than zero.

Results

While trials were excluded from the primary data analyses if they involved a timeout (i.e., the 15 s time limit was reached before a click was made on one of the two response buttons), it is still potentially informative to examine differences in timeout rates by condition and group. A 2 × 2 between-subjects ANOVA with group (non-professional vs. professional) and condition (one = dangerous vs. one = safe) as factors revealed a main effect of group on the number of timeouts; the non-professional searchers had fewer timeouts (M = 4.17 trials, SE = 1.09 trials) than the professional searchers (M = 8.41 trials, SE = 1.34 trials), $F(1, 66) = 6.05$, $p = .02$. There was also a main effect of condition with fewer timeouts in the one = dangerous condition (M = 4.34 trials, SE = 1.30 trials) than in the one = safe condition (M = 8.24 trials, SE = 1.13 trials), $F(1, 66) = 5.11$, $p = .03$. There was no significant interaction ($F < 1$).

SSM analyses

The primary dependent variable of interest – particularly as it pertains to the professional searchers – was search accuracy. Specifically, did our experimental manipulation affect search accuracy after a first target had already been found? To investigate whether the manipulation influenced subsequent search accuracy, we conducted a 2 × 2 between-subjects ANOVA with group (non-professional vs. professional) and condition (one = dangerous vs. one = safe) as the between-subject factors and SSM errors as the dependent variable. There was no significant main effect of group or condition for SSM errors (both *p*s > .30). The non-significant main effect of condition indicates that searchers performed similarly despite the difference in instructions between conditions. Searchers in the one = dangerous condition could have accurately classified the trial without conducting any subsequent search after a found target, although they still appear to have followed the general instruction to attempt to locate all targets present. Importantly, there was a significant interaction, $F(1, 66) = 4.05, p = .046, \eta_p^2 = .06$, such that the non-professional searchers had fewer SSM errors in the one = safe condition and the professional searchers had fewer SSM errors in the one = dangerous condition (see Table 1).

Follow-up tests were conducted to determine if each of the four separate experimental cohorts produced a SSM effect that was significantly greater than zero. As we made four separate comparisons, the significance level was adjusted for the *post-hoc* analyses per Bonferroni corrections (α=.05/4 = .0125). Non-professionals exhibited significant SSM errors in the one = dangerous condition, $t(20) = 4.41, p < .001$, but non-significant SSM errors in the one = safe condition, $t(19) = 2.64$, $p = .02$. In contrast, professional searchers exhibited non-significant SSM errors in the one = dangerous condition, $t(10) = 1.84, p = .10$, but significant SSM errors in the one = safe condition, $t(17) = 4.97$, $p < .001$. Non-professional searchers exhibited *more* SSM errors when one target indicated a dangerous trial versus when one target indicated a safe trial, whereas professional searchers exhibited *fewer* SSM errors when one target indicated a dangerous trial versus when one target indicated a safe trial.

Response time after a found target

Our second primary variable of interest was whether participants would spend more time searching after finding a first target due to the experimental manipulation. We conducted a 2 × 2 between-subjects ANOVA with group (non-professional vs. professional) and condition (one = dangerous vs. one = safe) as the between-subject factors and time spent searching after finding a first target as the dependent variable. There was no significant main effect of group ($p > .40$), and the main effect of condition only exhibited a trend towards significance, $F(1, 69) = 2.81$, $p = .10$. The interaction between group (non-professional vs. professional) and condition (one = dangerous vs. one = safe) was also non-significant, $F(1, 66) = 1.71$, $p = .20$, $\eta_p^2 = .03$ (see Table 2).

Table 1. SSM errors by search condition and group

Search condition	SSM error rate	Significant SSM effect
Non-professionals		
One = dangerous	10.98% (2.49%)	*
One = safe	7.24% (2.75%)	–
Professionals		
One = dangerous	7.85% (4.27%)	–
One = safe	16.98% (3.42%)	*

Note. SSM = subsequent search misses.
*Indicates $p < .0125$, significance altered by Bonferroni corrections ($\alpha = .05/4 = .0125$).

Table 2. Time spent searching after finding a target

Group	Condition		
	One = dangerous	One = safe	Comparison
Non-professionals	4.70 s (0.25 s)	5.50 s (0.29 s)	$t(39) = 2.06, p < .05$
Professionals	5.34 s (0.38 s)	5.34 s (0.29 s)	$t(27) = 0.01, p = .99$

Additional measures

Although our primary interests involved subsequent search, this robust search task yields several metrics that we evaluated against previous work, including additional accuracy metrics, response timing metrics, and general comparisons between professional and non-professional searchers. See Table 3 for full descriptive statistics.

For accuracy, we did not observe any significant main effects or interactions in single-target accuracy for either high-salience targets or low-salience targets (all ps > .10), which replicates a previous finding with this paradigm (e.g., Biggs & Mitroff, 2014). For response times, non-professional searchers were faster than professional searchers when no target was present, $t(68) = 4.04, p < .001$, and faster to find a first target, $t(68) = 4.29, p < .001$; however, there was no significant main effect of condition (one = dangerous vs. one = safe) or any significant interactions between group and condition for time to quit search when no target was present or for the time to find a first target (ps > .25). This evidence replicates a previous finding that non-professional visual searchers are faster than professional visual searchers (Biggs, Cain, Clark, Darling, & Mitroff, 2013; Biggs & Mitroff, 2014). More importantly, this evidence suggests that there was no global shift in performance *before* a target was found despite significant influences *after* a target was found.

Table 3. Descriptive statistics for non-professional and professional searchers

Group	Condition	
	One = dangerous	One = safe
Non-professionals		
Timeouts	2.14 (1.52)	6.20 (1.56)
Mean response time	8.13 s (0.32 s)	8.40 s (0.38 s)
1st Target response time	3.52 s (0.20 s)	3.62 s (0.21 s)
Single-target hit rate (high-salience)	88.77% (1.53%)	92.35% (1.12%)
Single-target hit rate (low salience)	48.70% (3.47%)	53.84% (2.86%)
Targets found on dual-target trials	1.39 (0.04)	1.46 (0.03)
False alarms	6.19 (1.49)	5.70 (1.53)
Professionals		
Timeouts	6.55 (2.10)	10.28 (1.65)
Mean response time	9.70 s (0.44 s)	9.77 s (0.35 s)
1st Target response time	4.42 s (0.32 s)	4.63 s (0.22 s)
Single-target hit rate (high salience)	89.96% (2.19%)	89.11% (1.61%)
Single-target hit rate (low salience)	52.04% (3.63%)	55.90% (4.06%)
Targets found on dual-target trials	1.45 (0.05)	1.39 (0.04)
False alarms	6.00 (2.06)	11.06 (1.61)

GENERAL DISCUSSION

The goal of the current manuscript was to investigate whether global decision-making criteria about a visual search task could influence search performance during the task. Previous evidence demonstrated that 'SSM' are affected by contextual conditions surrounding the search (Clark *et al.*, 2013), which provided the opportunity here to examine differences in performance due to a specific contextual influence – manipulations of the searchers' decision-making criteria. In this study, all participants were given the same primary task – find each and every target present – but the decision-making criterion about how to classify trials differed by condition. Those in the one = dangerous condition were instructed to label a trial as dangerous if they found one or two targets, and those in the one = safe condition were instructed to label a trial as dangerous only if they found two targets. These labels did not change the within-trial goals, but nevertheless changed performance between non-professional and professional searchers.

The findings suggest that the pattern of results depended upon whether the searchers had professional training or not. The non-professional searchers fit the hypothesized result pattern, demonstrating fewer second-target errors when finding a second target altered how a trial was classified. This fits with expectations that when additional data can alter a classification, participants will spend more time searching for that data to ensure accuracy. Professional searchers, however, showed the opposite pattern – increased second-target errors when a second target altered how a trial was classified. So, why would professional searchers react differently to the same decision-making criterion? It is possible that, upon deeming some of the contents to be safe (as a lone found target would be in the one = safe condition), they were likely to judge the rest of the display content in a similar fashion. Professional searchers could become more complacent in their subsequent search upon finding a 'safe' first target, which suggests that the mindset of a professional searcher might be more vulnerable to contextual conditions about the search than non-professional searchers. This interpretation is further supported by the evidence that professional searchers were less likely to commit SSM errors when a found target indicated danger – a scenario akin to finding explosives concealed in luggage.

Another issue at play between the professional and non-professional searchers might be their experience with maintaining an exhaustive search strategy – attending to and evaluating each item in a search. An exhaustive search requires a searcher to consider every item as a possible target and is in contrast to a non-exhaustive search – wherein a searcher decides to terminate the search after having evaluated 'enough' of the items as possible targets. The decision to terminate a search is complicated (e.g., Chun & Wolfe, 1996; Wolfe & Van Wert, 2010; Wolfe *et al.*, 2007), and many factors can influence whether an exhaustive search is performed or not. For example, in visual foraging (e.g., trying to pick the most berries from bushes in the shortest amount of time) the primary goal is to maximize yield (e.g., Cain *et al.*, 2012; Kristjánsson, Jóhannesson, & Thornton, 2014; Wolfe, 2013). This leads to a non-exhaustive search strategy, where it is most beneficial to move on to a new berry bush when there are not 'enough' ripe berries left to pick. In contrast, security screening requires an exhaustive search of all possible locations where contraband might be located.

The professional searchers in this study were likely accustomed to conducting exhaustive searches as they regularly performed security screenings as a part of their daily work duties, yet the non-professionals might not have been so accustomed to these types of search. As a result, the non-professional searchers might have developed a tendency to intermittently use a non-exhaustive search strategy (i.e., on some trials they might have chosen to quit prematurely rather than conducting an exhaustive search). A 100% exhaustive search strategy would be immune from variability in a searcher's quitting threshold as each trial would end with the same self-termination rule – quit when the entire display has been searched. Alternatively, an intermittent non-exhaustive search strategy would be vulnerable to various contextual conditions given that some trials would be terminated prior to the searcher having examined every item in the display. That is, searchers who sometimes employ an exhaustive strategy and sometimes employ a non-exhaustive strategy might employ one strategy more so in one context than in another.

There is some evidence in the current manuscript to support the supposition that non-professional searchers intermittently used a non-exhaustive search strategy. Specifically, the non-professionals searched longer after having found a first target when finding a second target could change how they classified a trial (i.e., they spent longer searching after finding a first target for the one = safe condition than the one = dangerous condition). While this is just one theory about how the decision-making criteria manifested different behavioural reactions in the non-professionals and professionals, it does suggest two conceptually different influences at play. Non-professional searchers might have been affected by a change to their quitting threshold, which was evident via differences in time spent searching after finding a target. However, professional searchers might have been affected by the classification of the first found target, which made them complacent during subsequent search if the first found target was 'safe'.

In a practical sense, professional searchers could be highly influenced by the common scenarios and professional practices that they encounter on a regular basis. For example, the first found target could be something perceived as relatively harmless, such as an oversized water bottle that presumably only contains water. Water bottles are far more common objects than weapons or explosives – which increase their likelihood of being seen – and the person carrying a water bottle filled with water is not likely a significant threat. This example could generate greater complacency from a professional screener when search resumes. Additionally, some airport security uses the threat image projection (TIP) system at the checkpoint (Hofer & Schwaninger, 2005) wherein prohibited items are randomly projected into real bags, and professional screeners are instructed to identify every such image. These items are not real threats, and the screener is informed of that upon identifying the item, which makes a TIP image akin to a 'safe' found target. However, there remains a need to continue searching after identifying a TIP image, and there is the potential that the searcher could become complacent during subsequent search. The practical application of this evidence is that professional searchers should treat all target items with the same level of priority to avoid such contextual influences during subsequent search.

This study also replicated previous findings regarding non-professional versus professional searchers (e.g., Biggs & Mitroff, 2014; Biggs *et al.*, 2013). In particular, non-professional searchers were faster overall than professional searchers. Training is designed to improve performance, and visual search training appears to enhance diligence rather than speed – that is, searchers enhance strategy by slowing down and using more consistent approaches as opposed to finishing the same search in a shorter period of time. The same evidence has been observed with radiologists (Clark, Samei, Baker, & Mitroff, 2011) and orthodontists (Jackson, Clark, & Mitroff, 2013), which further supports the idea that visual search training enhances diligence by slowing the searcher down as opposed to speeding them up.

Ultimately, decision-making criteria represent an important contextual influence upon multiple-target visual search. Professional searchers appear to benefit from a mindset that all prohibited items share the same level of priority, which represents another of many situational or procedural factors that could be adjusted to minimize SSM errors. For example, more SSM errors occur when the searcher is highly anxious (Cain *et al.*, 2011) or when stricter time pressures are imposed (Fleck *et al.*, 2010) – both of which are factors that could be altered through procedure without changing the search task. Recent research has also described other procedural measures that could be addressed to limit the likelihood of SSM errors, such as re-run procedures aimed at turning multiple-target search into multiple single-target searches (for more detail, see Cain, Biggs, Darling, & Mitroff, 2014). Contextual conditions represent another means of affecting search accuracy without altering the actual search task, whether those contextual conditions involve the decision-making criteria, as manipulated here, or the searcher's broader understanding of when they are finished conducting visual searches, as with a fixed objective versus fixed duration framework (Clark *et al.*, 2013). Decision-making criteria are potent factors to include when considering what might improve search accuracy, and, especially for professional searchers, it is vital that every potential target be treated with the same level of caution to avoid making further errors.

Acknowledgements

We thank Michael Zunk, Hung Nguyen, Bob Kingan, Kellis Turner, Ken Macdonald, Richard Menard, and Everett Vaughn for making it possible to conduct research at RDU airport. We especially thank Dave Evans and Stuart Campbell for their extraordinary efforts to support the research at RDU. We also thank Stephen Adamo Emma Wu Dowd, Michael Hout, and Jeremy Wolfe for helpful comments, and Elise Darling and undergraduate research assistants for assistance with data collection at Duke University. This work was partially supported by the Army Research Office (#54528LS) and partially through a subcontract with the Institute for Homeland Security Solutions, a research consortium sponsored by the Resilient Systems Division in the Department of Homeland Security (DHS). This material is based upon work supported by the DHS under Contract No. HSHQDC-08-C-00100. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the official policy or position of DHS or of the US Government. The study is approved for public release.

References

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

Anbari, M. M. (1997). Cervical spine trauma radiography: Sources of false-negative diagnoses. *Emergency Radiology*, *4*, 218–224. doi:10.1007/BF01508174

Berbaum, K. S. (2012). Satisfaction of search experiments in advanced imaging. In: *Proceedings of SPIE 8291*. Human Vision and Electronic Imaging XVII, 82910V. doi:10.1117/12.916461

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. doi:10.1016/S1076-6332(00)80176-2

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.

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

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. doi:10.1080/13506285.2013.790329

Biggs, A. T., & Mitroff, S. R. (2014). Different predictors of multiple-target search accuracy between non-professional and professional visual searchers. *Quarterly Journal of Experimental Psychology*, *67*, 1335–1348. doi:10.1080/17470218.2013.859715

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

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

Cain, M. S., 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. doi:10.1037/xap0000014

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

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

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

Chun, M. M., & Wolfe, J. M. (1996). Just say no: How are visual searches terminated when there is no target present? *Cognitive Psychology*, *30*, 39–78. doi:10.1006/cogp.1996.0002

Clark, K., Cain, M. S., Adcock, R. A., & Mitroff, S. R. (2013). Context matters: The structure of task goals affects accuracy in multiple-target visual search. *Applied Ergonomics*, *45*, 528–533. doi:10.1016/j.apergo.2013.07.008

Clark, K., Samei, E., Baker, J., & Mitroff, S. R. (2011). *Expertise in radiological screening and satisfaction of search*. Poster presented at the annual Object Perception, Attention, and Memory meeting. Seattle, WA.

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, *11*, 1–36. doi:10.1167/11.5.14

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

Hofer, F., & Schwaninger, A. (2005). Using threat image projection data for assessing individual screener performance. *WIT Transactions on the Built Environment*, *82*, 417–426.

Jackson, T. H., Clark, K., & Mitroff, S. R. (2013). Enhanced facial symmetry assessment in orthodontists. *Visual Cognition*, *21*, 838–852. doi:10.1080/13506285.2013.832450

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

Kristjánsson, Á., Jóhannesson, Ó. I., & Thornton, I. M. (2014). Common attentional constraints in visual foraging. *PLoS ONE*, *9*, e100752. doi:10.1371/journal.pone.0100752

Kuhn, G. J. (2002). Diagnostic errors. *Academic Emergency Medicine*, *9*, 740–750. doi:10.1197/aemj.9.7.740

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

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. doi:10.1163/156856897X00366

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

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

Voytovich, A. E., Rippey, R. M., & Suffredini, A. (1985). Premature conclusions in diagnostic reasoning. *Academic Medicine*, *60*, 302–307.

Wolfe, J. M. (2013). When is it time to move to the next raspberry bush? Foraging rules in human visual search. *Journal of Vision*, *13*, 1–17. doi:10.1167/13.3.10

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. doi:10.1037/0096-3445.136.4.623

Wolfe, J. M., & Van Wert, M. J. (2010). Varying target prevalence reveals two dissociable decision criteria in visual search. *Current Biology*, *20*, 121–124. doi:10.1016/j.cub.2009.11.066

Received 17 April 2014; revised version received 12 August 2014