Assessing visual search performance differences between Transportation Security Administration Officers and non-professional visual searchers

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

Some visual searches depend upon accuracy (e.g., radiology, airport security screening), and it is important for both theoretical and applied reasons to understand what factors best predict performance. The current study administered a visual search task to both professional (Transportation Security Administration Officers) and non-professional (members of Duke university) searchers to examine group differences in which factors predict accuracy. Search speed—time taken to terminate search—was the primary predictor for non-professional searchers (accounting for 59% of their accuracy variability) and for the least experienced professional searchers (37% of variability). In contrast, consistency—how similarly (in terms of search speed) an individual spent searching from trial-to-trial—was the primary predictor for the most experienced professional visual searchers (39% of variability). These results inform cognitive theory by illuminating factors that differentially affect search performance between participants, and real-world issues by identifying search behaviors (consistency in particular) important to experienced professional searchers.

Visual search, the process of finding target items amongst distractors, is an everyday task that can be as mundane as locating car keys on a cluttered desk, or as complicated as finding explosives purposefully hidden in airport luggage. Most people conduct visual searches constantly throughout the course of a day (e.g., finding your place while reading, detecting a new email in your inbox, or finding the right pair of socks in the morning), and these searches are conducted so frequently that most adult humans have almost immeasurable experience with visual search. Beyond such everyday tasks, some individuals conduct visual searches for a living. Radiologists, airport baggage screeners, border patrol agents, lifeguards, and others spend much of their professional careers performing specialized visual searches. Accuracy in these real-world visual searches is often very important, as success can often have life-or-death implications. Consequently, professional searchers receive directed and organized training unique to their fields on what to search for and how to search for it. This directed training provides professional searchers with very different visual search experiences than those of non-professionals despite both groups having a lifetime of exposure to visual search experiences. As such, an important question is whether professional searchers' unique experiences alter their visual search behaviors.

Professional searchers' directed training can clearly have an impact on their within-domain performance (e.g., Biederman & Shiffrar, 1987), and this advantage partially stems from their domain-specific knowledge (i.e., radiologists are better able to find a tumor because they are better able to identify a tumor). However, do domain-specific performance benefits generalize to search abilities more broadly? Any such generalized benefits would suggest that professional training impacts cognitive functioning beyond the task for which the professional searcher is trained and could inform the nature of generalized learning.

The goal of the current study was to directly compare professional and non-professional searchers on an identical visual search task to explore if, and how, they might differ. To achieve this goal, we recruited Transportation Security Administration (TSA) Officers who regularly conduct searches through X-ray images of luggage as professional searchers and members of the Duke University community as non-professional searchers. All individuals completed a simple visual search task wherein they looked for a particular target shape among similarly-shaped distractors (see Figure 1). This simple task provided a level playing field for evaluating performance between professional and non-professional searchers; for example, if participants searched for guns and bombs hidden in luggage X-ray images, then the professional searchers would have an overwhelming advantage given their previous experiences with such stimuli and their domain-specific knowledge of potential targets. The *a priori* hypothesis was that the professional searchers would be more accurate than the non-professional searchers, but the critical question was *why* the professional searchers might be more accurate. Simple group comparisons have the potential to reveal enhanced performance in professional searchers, but more nuanced and targeted comparisons are needed to address *why* professionals may exhibit better performance. By simultaneously focusing on multiple levels of analyses, the current goal was to examine how professional and non-professional searchers differed from one another to inform both cognitive theories of visual search and to improve real-world search performance.

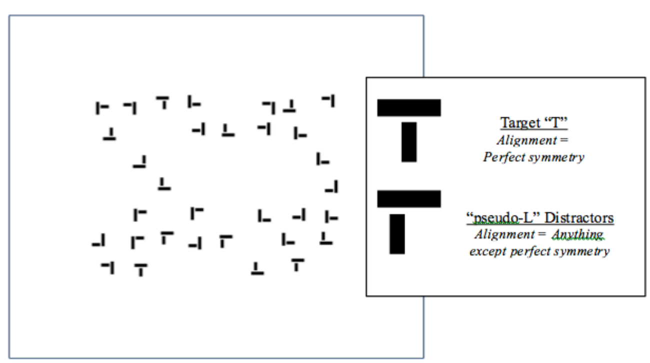

Figure 1. Sample trial (set size 32; target "T" present) with larger examples of a target "T" and a distractor "pseudo-L".

Potential Implications for Cognitive Theory

Previous research has broadly categorized the various factors influencing attention into two categories: "stimulus-driven" or "goal-driven" effects (e.g., Theeuwes, 2010; Yantis, 2000). Stimulus-driven effects refer to how the physical characteristics of the display influence attentional control. For example, salient, but irrelevant, items can automatically attract an observer's attention, such as when an irrelevant red item in a field of green items distracts an observer during visual search (Theeuwes, 1991; 1992; 2004). In contrast, goal-driven effects refer to when attention is influenced by the behavioral goals of the observer. For example, a person looking for a red item will be distracted by other red items (e.g., Folk, Remington, & Johnston, 1992). This stimulus-driven versus goal-driven dichotomy provides a powerful means to account for many of the potential influences on attentional allocation, but goal-driven effects, which have been utilized frequently to investigate top-down attention, cannot encompass all possible top-down influences (see Awh, Belopolsky, & Theeuwes, 2012). Specifically, a simple stimulus-driven versus goal-driven dichotomy does not account for another highly powerful influence that is relatively understudied as a driving force in attention: the observer.

The observer contributes a unique influence to attentional control via his or her life experience, personal preferences, wants, and needs. For example, smokers pay more attention to smoking-related items than non-smokers (Bradley, Mogg, Wright, & Field, 2003; Field, Mogg, & Bradley, 2004; Waters & Feyerabend, 2000), and drug users pay more attention to drug-related items than non-users (Field, Eastwood, Bradley, & Mogg, 2006; Jones, Jones, Blundell, & Bruce, 2002; Jones, Jones, Smith, & Copley, 2003). Moreover, a wide variety of research using the Implicit Associations Test (e.g., Greenwald, McGhee, & Schwartz, 1998) has demonstrated that an individual's unique implicit beliefs and associations can subtly affect their cognitive processing. With regard to visual search specifically, individuals will perform differently if they have prior knowledge about an otherwise ambiguous stimulus and if they have personal attachments to known symbols, such as sports team logos (Biggs, Kreager, Gibson, Villano, & Crowell, 2012). Collectively, this evidence suggests that individual differences can influence both bottom-up and top-down processes in attentional allocation. Here we examine how a specific individual difference factor—experience with real-world visual search tasks— can affect visual search accuracy.

Potential Implications for Real-World Searches

The present study aims to inform cognitive theories of visual search but also has the potential to inform real-world visual search. Many real-world visual searches, such as those in radiology and airport security screening, rely upon accurate assessments to save lives, and decades of research have focused on improving performance in such visual searches (e.g., Krupinski, 2010; McCarley, Kramer, Wickens, Vidoni, & Boot, 2004; Smith, 1967; Tuddenham, 1962). For example, previous research has focused on how recognizing threat items in X-ray images is influenced by image-based effects (e.g., Schwaninger, 2003; Schwaninger, Hardmeier, & Hofer, 2004) or how superimposed objects common to X-ray images can impair target detection (Schwaninger, 2005). Other researchers have suggested altering visual search goals as a means of improving performance, such as introducing a division of labor—having some individuals search for one target type and other individuals search for a different target type (Menneer, Barrett, Phillips, Donnelly, & Cave, 2007).

The research to date has been highly productive, but due to the nature of real-world visual search tasks, there are often constraints on what can and cannot be improved. Specifically, in many real-world settings it is not always possible to alter the stimuli themselves or the visual searcher's goals (i.e., the TSA cannot pre-determine stimuli since they cannot dictate the contents of passengers' luggage). These limitations can hinder efforts to improve visual search accuracy, so it is important to explore all possible ways to increase performance. In particular, if neither the visual search array nor the situation can be improved (i.e., the stimulus-driven and goal-driven factors), then the best alternative is to improve the abilities of those individuals performing the visual search (i.e., the observer). The human element is arguably the weakest link in airport screening (Schwaninger, 2005), which further emphasizes the need to examine top-down influences on visual search.

Differences between Professional and Non-professional Visual Searchers

When comparing search performance between professional and non-professional visual searchers, the obvious first question was whether the professional visual searchers are more accurate than the non-professional visual searchers. However, our central goal extended beyond assessing group differences between professional and non-professionals to also determine *how* professional visual searchers differed from non-professional visual searchers. Specifically, we focused on three factors that might reveal group differences: search speed, search consistency, and level of experience. By performing regression analyses, in addition to assessing group differences, we can offer insight into what factors are most influential in determining search accuracy. Here we briefly discuss the three primary factors to be examined.

Search Speed. We predicted that search speed, how long an observer takes to terminate a search, would be an important predictor of accuracy. A classic cognitive phenomenon is that of a speed-accuracy trade-off—where individuals perform a task faster at the cost of lower accuracy, or vice versa. Accuracy is often the critical outcome aspect for professional searchers, but, while professional visual searchers might be more accurate, do professional searchers also differ in the speed with which they conduct a search? Years of professional training and experience could lead to greater diligence in performing any visual search, which could produce slower response times. Alternatively, professional training and experience could enhance visual search efficiency, allowing professional searchers to accurately complete searches faster than non-professional visual searchers.

Search Consistency. We hypothesized that *consistency*, or how similarly an individual conducts a visual search from trial-to-trial, would be a powerful predictor of accuracy. Unless a target is immediately detectable (i.e., a "pop-out" search, Treisman & Gormican, 1988), the observer must dedicate attentional resources to the search process. These attentional resources would be engaged in a variety of activities, including maintaining a target set in memory, object recognition, and knowing what items have and have not already been searched. Previous research has shown that such cognitive burdens can affect accuracy (e.g., Cain & Mitroff, 2012), and so alleviating any cognitive burdens of the searcher may offer a means to improve accuracy. Executing a visual search in a consistent manner each time should reduce the need to remember which parts of the display have or have not been searched; and in turn, this should allow the visual searcher to allocate more cognitive resources to other aspects of the search process. Thus, we predicted that participants would perform more accurately when they were more consistent visual searchers.

For the current study, we operationalized consistency in terms of response times. A consistent visual searcher should take approximately the same amount of time to complete each search (e.g., five seconds one trial, four seconds the next, and five seconds after that), whereas an inconsistent searcher should take very different amounts of time from trial to trial (e.g., five seconds one trial, two seconds the next, and ten seconds after that). Defining consistency as a temporal construct is not as precise as a spatial measure (e.g., eye-tracking metrics), but has two key benefits; it is an easy measure to collect from common behavioral metrics, and it provides a comparable measure across individuals, regardless of what particular search strategy (or lack thereof) that they employed. This later point is especially important—by focusing on a temporal measure of consistency we can be agnostic as to what particular spatial strategy might be employed. As a result, we are able to assess a general measure of consistency that applies equally to all participants. Our specific consistency calculation is discussed further in the Methods section.

Level of Experience. While adults with normal vision have accumulated innumerable visual search experiences during their lifetimes, professional searchers have received directed visual search training and on-job experience with feedback. Is it possible that professional search experiences, above and beyond actual training, can significantly impact visual search behaviors? A key aspect of the current study is that we compare non-professional to professional searchers *and* we compare two groups of professional searchers—those with relatively little professional experience ("early-career"; less than 3 years at the TSA) to those with relatively a lot of professional experience ("experienced"; more than 6 years at the TSA).

A broad comparison between non-professional and professional searchers can compare the impact of training and experience on general search abilities, but a directed comparison between early-career and experienced professional searchers can more directly speak to the role of experience. By comparing two levels of experience within the same professional population, we can assuage generic concerns about alternative explanations. Specifically, there is always a concern when comparing two populations that one group may be more inherently motivated for the given task. Might the TSA Officers be more motivated to try hard in the current study than the non-professional searchers? Professional searchers performing a task associated with their profession could very well have greater incentive to perform well, whereas non-professional searchers volunteering to participate in an experiment have no such additional motivations. Any differences observed between the early-career and experienced professional searchers in the current study would diminish motivational explanations of the findings given that these two groups are matched on general motivation.

The Visual Search Task

To best explore our present goals—assessing search differences between non-professional and professional searchers and examining what factors best predict search accuracy—we employed a visual search task with several specific features. First, we wanted a task that could be equally well executed by non-professional and professional searchers alike. A search task with domain-specific aspects from the professional searchers' realm (e.g., searching for hidden bomb parts in a luggage X-ray) would place the non-professional participants at a significant disadvantage. Second, since we wished to examine what factors might predict variability in accuracy within population groups, the task needed to be able to engender reasonable variability in both response timing and accuracy. To put it simply, to assess what causes variance in accuracy, there must be variance in accuracy. Therefore, we employed a visual search task that non-professional and professional participants could both perform, and that was more difficult than typical "Ts and Ls" cognitive psychology paradigms. Specifically, participants searched for target "T" shapes amongst "pseudo-L" distractors. The distractors were identical to the targets except that the two bars were not perfectly symmetrical (see Figure 1). Having distractors that can be very similar to the targets adds difficulty to the task (e.g., Duncan & Humphreys, 1989) without requiring any specific knowledge or training.

Method

Participants

Non-professional Participants. Our non-professional searchers were 93 members of the Duke University community (Mean age = 20.3 years, SD = 2.81, 57 female) who participated for partial completion of a course requirement or $10.

Professional Participants. Our professional searchers were Transportation Security Administration (TSA) Officers employed to perform security screenings at airport checkpoints. These individuals participated as a part of the ongoing Duke-RDU Study at Raleigh-Durham International Airport. Because they participated while at work, several steps were taken to ensure their rights of voluntary participation and confidentiality. When scheduled to participate, the TSA Officers reported to a dedicated testing lab located at the airport and were given the option to participate in the experiment or to review training materials. Supervisors were unaware of which TSA Officers participated and which ones opted to review training materials. TSA Officers who participated were also given the option to allow their data to be used only for TSA research purposes or for both TSA and Duke University research purposes. However, participants were informed that their individual data would never be revealed to the TSA in connection to them, ensuring anonymity in their participation. As a part of this ongoing project, 310 TSA Officers were scheduled to participate; 7 opted not to participate in this particular experiment, and 14 designated their data for TSA purposes only.

For the current report, we further limited the data from the remaining 289 TSA Officers to only those who reported that they regularly conduct X-ray security searches as part of their duties. Sixty-seven did not report regularly conducting X-ray searches. These individuals included supervisors and others whose normal rotation through the various job duties (e.g., ID checks, pat-downs) did not involve regular X-ray searching at the time of testing. Data from an additional 2 TSA Officers were removed due to a computer error and from another 9 due to a failure to follow instructions (e.g., confusion of response key assignments or not providing necessary information). Finally, five TSA Officers were over the age of 65, and their data were removed from analyses as there were not enough participants in this uppermost age range for sufficient statistical power. After accounting for these various exclusion criteria, data from 206 professional visual searchers remained for analysis. TSA Officers were then divided into two groups: "early-career" TSA Officers with less than three years of TSA employment (N = 70, Mean age ≈ 42 years[1], 24 female), and "experienced" TSA Officers with six years or more of TSA employment (N = 96, Mean age ≈ 47 years, 28 female).

Design

Apparatus. Non-professional searchers from the Duke community were tested at the Duke Visual Cognition Laboratory on the Duke University campus with Dell Inspiron computers with 20-inch CRT monitors. Professional searchers recruited from the TSA were tested in a private room at Raleigh-Durham International Airport (RDU) on Dell Vostro 260 computers and 23.6-inch widescreen LCD monitors. The six testing stations were separated by dividers, and the room was dimly lit for testing. The computer displays were adjusted so that both the RDU lab and Duke testing stations presented the same physical display sizes to participants. All testing stations used Matlab software (The MathWorks, Natick, MA) and the Psychophysics Toolbox version 3.0.8 (Brainard, 1997; Pelli, 1997; Kleiner, Brainard, & Pelli, 2007) for experimental presentation and data collection.

Stimuli. Each search display was comprised of multiple pseudo-"L"s as distractors, and half of the displays contained one target "T." Each item was comprised of two perpendicular black lines (stroke width = 0.3°, subtending 1.3° x 1.3° total). Target "T"s had a crossbar directly in the middle, whereas distractor "L"s had a crossbar slid to variable distances away from center (see Figure 1). The distractor stimuli were variable in shape with some very close to the target "T"s. This was done intentionally to make the task sufficiently difficult so that participants would not perform perfectly and would require time to find the target. Each item was placed with a slight spatial jitter within randomly selected cells of an invisible 8 x 7 grid that subtended 25.4° x 19.1° at an approximate viewing distance of 60 cm. None of the cells overlapped, and the display items were presented against a white background. Both targets and distractors were rotated randomly in one of the four cardinal directions (0°, 90°, 180°, and 270°). All items were distractors for the target absent condition, and in the target present condition, all items were distractors except for one target "T." Displays consisted of four possible set sizes: 8, 16, 24, and 32.

Procedure. Each trial began with a black fixation cross presented in the center of the display. After 100 ms, the cross disappeared and the search array was presented and remained on screen until response. Participants were to make a presence/absence judgment about a target "T" by pressing one of two assigned keys on the keyboard; keys used for responses ("z" and "/") were counterbalanced across participants. Sixteen practice trials preceded 256 experimental trials. Both practice and experimental trials were equally divided among the four set sizes with equal numbers of target present and target absent trials at each of the four set sizes. Accuracy feedback was provided for the practice trials but not for the experimental trials. After a key response, the search array disappeared, and the next trial began automatically. Participants were provided with the opportunity to rest every 25 experimental trials.

Pilot testing revealed that some professional visual searchers engaged in exceptionally long searches (e.g., a search slope of over 1 second per item) if not provided with a time limit. The experimental goal was to understand the processes involved in comparable searches between the groups, and so a 30-second time limit was introduced for the professional visual searchers. Trials on which the time limit was exceeded were excluded from the analyses, which resulted in only 0.4% of trials being excluded for the professional visual searchers. Although the non-

[1] Ages are represented as approximations since TSA Officers reported their age via ranges in a questionnaire (e.g., between 18 and 25 years old).

professional visual searchers had no time limit, 0.06% of trials exceeded thirty seconds, and these trials were also excluded from analyses.

Planned Analyses

The visual search task employed here was relatively difficult, which resulted in participants making a sufficient number of errors and having a sufficient amount of variability in their response timing so that we could assess both accuracy and response time metrics. As such, we compared basic measures of accuracy and response times between the professional and non-professional visual searchers to examine group differences. In addition, we narrowed in on more nuanced analyses that could reveal what factors were most likely to influence search accuracy. Our primary focus involved two factors that could affect accuracy (search speed and search consistency), and we briefly describe each here.

Assessment of search speed: Correct rejection response time slope. An informative metric in visual search is the *search slope*—the additional time taken to terminate a search for each additional item present in the display. Search slopes provide a metric of search difficulty in terms of how much longer a participant spends searching per each additional display item; the higher the slope, the more difficult the search (see Wolfe, 1998). Search slopes were calculated for each participant in the current study based on the best fit line across the average response times for each display size. We specifically focused on correct rejection trials (target absent trials in which the participant correctly report no target was present), and this factor will represent how much the participant sacrificed speed for accuracy.

Assessment of Search Consistency: Response time consistency. Beyond search speed, the current experiment also assessed search consistency—the variability in response timing. We calculated a consistency measure for each participant through a three-step process. First, we calculated the variability of response times at each of the four set sizes for trials in which a participant voluntarily terminated a trial by reporting that no target was present (i.e., correct rejection and miss trials). We defined this variability as the standard deviation of each participant's response times for these trials at each set size. Second, to account for the differences in average response time between the different set sizes (e.g., participants were overall slower at set size 32 than set size 8), we divided each standard deviation by the average response time for the given participant at each set size. Third, we averaged these values across the four set sizes to produce a single measure of consistency for each participant. This measure indicates whether a participant took close to the same amount of time to terminate search (e.g., taking about five seconds on each trial) or showed substantial variability in the time taken before terminating search (e.g., taking five seconds one trial, three seconds on another, ten seconds on the trial after that, and so forth). Actual values for this measure will inherently depend upon the response times of the given task, but in general, it can range from 0 to less than 1 with lower values representing more consistent search. Mathematically,

$$\text{Consistency} = \left(\sum_{i=8}^{32} \frac{\sigma_i}{RT_i} \right) / 4, \qquad (1)$$

where i indicates the set size (8, 16, 24, or 32).

Results

Accuracy

Accuracy data were submitted to a 3x4 mixed-model ANOVA with a between-subject factor of group (non-professionals, early-career professionals, and experienced professionals) and a within-subject factor of set size (8, 16, 24, 32). See Figure 2 for results. There was a main effect of group with non-professional searchers completing the task with lower accuracy (82.3%) than professional searchers (early-career professionals: 88.8%, experienced professionals: 87.2%) $F(2, 256) = 19.29$, $p < 0.001$, $\eta_p^2 = 0.13$, and a main effect of set size with lower accuracy at higher set sizes (91.40%, 86.90%, 84.35%, 81.68% for set sizes 8, 16, 24, and 32 respectively), $F(3, 768) = 233.09$, $p<0.001$, $\eta_p^2 = 0.48$. There was a significant interaction between group and set size, where the decline in accuracy across set sizes was greater for non-professional searchers (–0.45% per item) than for early-career (–0.33% per item) or experienced professional searchers (–0.41% per item), $F(6, 768) = 3.07$, $p < 0.01$, $\eta_p^2 = 0.02$.

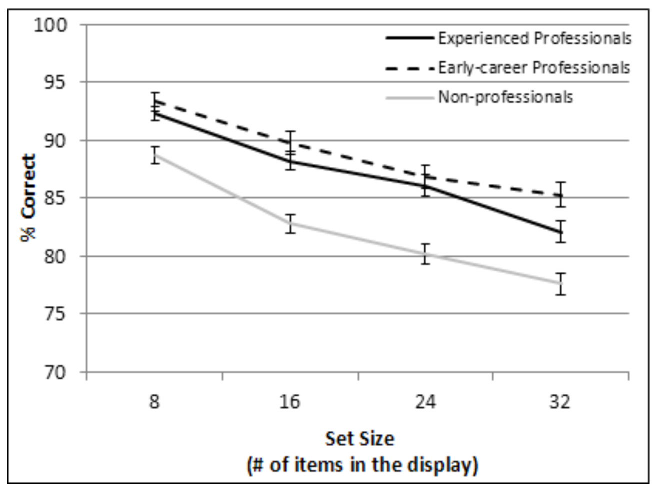

Figure 2. Accuracy by set size for professional (early-career and experienced TSA Officers) and non-professional visual searchers. Error bars represent standard error values.

Non-professional searchers had a significantly lower hit rate (M=69.13%, SE = 1.25%) than professional searchers (early-career professionals: M = 79.04%, SE = 1.44%; experienced professionals: M = 77.29%, SE = 1.23%), $F(2, 256) = 16.73$, $p < 0.001$, $\eta_p^2 = 0.12$, and non-professional searchers also had significantly higher false alarm rates (M = 4.50%, SE = 0.61%) than professional searchers (early-career professionals: M = 0.96%, SE = 0.70%; experienced professionals: M = 2.29%, SE = 0.60), $F(2, 256) = 7.05$, $p < 0.001$, $\eta_p^2 = 0.05$. Notably, there were no significant differences when directly comparing early-career and experienced professional searchers on overall accuracy, hit rate, or false alarm rate (all $p > 0.1$).

Response Time

Response time data were submitted to two 3x4 mixed-model ANOVAs with a between-subject factor of group (non-professionals, early-career professionals, and experienced professionals) and a within-subject factor of set size (8, 16, 24, 32); one ANOVA for hit trials and another for correct rejection trials (see Figure 3). Only correct trials were included for the response time analyses.

Hit trial response times. There was a main effect of group with non-professional searchers taking significantly less time to locate a target (3.86 s/trial) than professional searchers (early-career professionals: 6.01 s/trial, experienced professionals: 6.12 s/trial), $F(2, 256) = 94.73$, $p < 0.001$, $\eta_p^2 = 0.43$. There was also a main effect of set size with slower response times to find a target at larger display sizes (2.98 s, 4.64 s, 6.15 s, & 7.56 s for set sizes 8, 16, 24, & 32, respectively), $F(3, 768) = 1395.88$, $p < 0.001$, $\eta_p^2 = 0.85$. Finally, there was a significant interaction between group and set size in the time to locate a target, where search slopes were smaller for non-professional searchers (134 ms/item) than for professional searchers (early-career professionals: 216 ms/item, experienced professional searchers: 221 ms/item), $F(6. 768) = 33.58$, $p < 0.001$, $\eta_p^2 = 0.21$. The early-career and experienced professional searchers did not significantly differ in hit trial response times or hit trial search slopes (F's < 1).

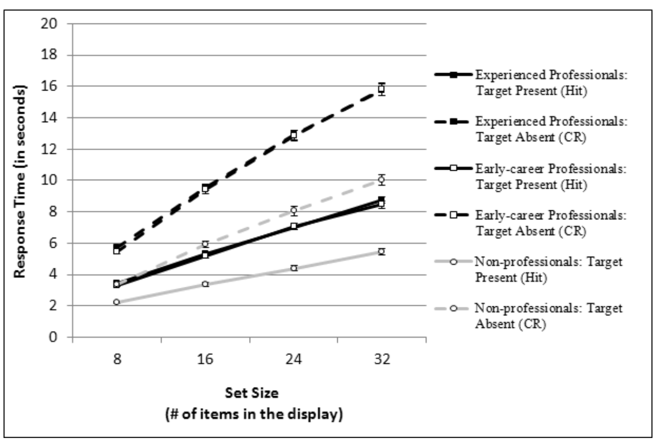

Figure 3. Response times for correct rejection (CR) and hit trials by set size for professional (early-career and experienced TSA Officers) and non-professional visual searchers. Error bars represent standard error values.

Correct rejection trial response times. There was a main effect of group with non-professional searchers taking significantly less time before terminating search (6.85 s/trial) than professional searchers (early-career professionals: 10.89 s/trial, experienced professionals: 10.99 s/trial), $F(2, 256) = 93.29$, $p < 0.001$, $\eta_p^2 = 0.42$. There was also a main effect of set size with longer response times to terminate search at larger display sizes (4.87 s, 8.30 s, 11.26 s, 13.88 s for set sizes 8, 16, 24, & 32, respectively), $F(3, 768) = 2597.67$, $p < 0.001$, $\eta_p^2 = 0.91$. Finally, there was a significant interaction between group and set size in correctly terminating search, where the increase in response time across set sizes was smaller for non-professional searchers (276 ms/item) than professional searchers (early-career professionals: 431 ms/item, experienced professionals: 419 ms/item), $F(6, 768) = 48.08$, $p < 0.001$, $\eta_p^2 = 0.27$. The early-career and experienced professional searchers did not significantly differ in correct rejection trial response times or correct rejection search slopes (F's < 1).

Consistency

A consistency value was calculated for each participant using Equation 1. This consistency measure ranged from 0.13 to 0.82 (M = 0.27, SE = 0.01), and lower values represent more consistent visual search response timing. There was a significant difference between non-professional and professional visual searchers, with non-professional searchers performing less consistently (M = 0.30, SE = 0.01) than professional searchers (early-career: M = 0.26, SE = 0.01; experienced: M = 0.25, SE = 0.01) $F(2, 256) = 8.85$, $p < 0.001$, $\eta_p^2 = 0.07$. Note that the calculation of consistency (Equation 1) uses the standard deviation, which is susceptible to trial count differences across set sizes and across participants. To control for any potential influences from differences in trial count, we also calculated consistency based upon standard error instead of standard deviation as it inherently takes into account trial count variance. The same data pattern emerged—there was a significant difference between non-professional and professional searchers, with non-professional searcher performing less consistently (M = 0.048, SE = 0.001) than professional searchers (early-career: M = 0.042, SE = 0.002; experienced: M = 0.041, SE = 0.001) than non-professional visual searchers, $F(2, 256) = 6.45$, $p < 0.001$, $\eta_p^2 = 0.05$.

Age

Compared to the relatively homogenous ages of the non-professional visual searchers (17 to 32 years; Mean age = 20.3 years, SD = 2.81), the professional visual searchers were more heterogeneous. The professional visual searchers' ages were collected as ranges for confidentiality reasons, which included 18–25 (N = 7), 26–34 (N = 34), 35–49 (N = 57), and 50–65 (N = 68). Separate 4x4x2 mixed model ANOVAs were run on the professional searchers' age ranges for accuracy, hit trial response times, and correct rejection trial response times, with set size as a 4-level within-subjects variable, and age and level of experience (early-career or experienced) as a 4-level and 2-level between-subjects factor, respectively. Age did not significantly affect accuracy ($F < 1$), but it did affect hit trial response times and correct rejection response times. Older participants were both slower to find a target (hit trial response times, 18–25: 3.94s, 26–34: 6.21s, 35–49: 5.97s, 50–65: 6.40s), $F(3, 158) = 4.08$, $p < 0.01$, $\eta_p^2 = 0.07$, and slower to correctly report that no target was present (correct rejection response times, 18–25: 5.19s, 26–34: 8.87s, 35–49: 12.11s, 50–65: 14.83s), $F(3, 158) = 2.81$, $p < 0.05$, η_p^2

= 0.05. Older participants were also more consistent (18–25: .27, 26–34: 0.28, 35–49: 0.27, 50–65: 0.23), $F(3, 162) = 5.15$, $p < 0.01$, $\eta_p^2 = 0.09$. Age did not interact with level of experience for any of these effects (all *p's* > 0.1).

There was a significant difference in age between the early-career and experienced professional searchers, $t(164) = 2.86$, $p < .01$. As such it is important to determine whether age or experience underlies any of the found differences. There are a few factors that suggest experience is the key driving force and not age. First, the early-career and experienced professional searchers did not differ in accuracy or response time results (see above). Thus they have a significant age difference, but this did not manifest into differences in basic search performance. Second, age was not a significant factor in the regression analyses that assess what factors predict performance (see below section). Third, the experienced professional visual searcher group had enough participants in two different age ranges such that we could compare performance from different ages within an experienced-matched sample. Experienced professional visual searchers aged 35–49 (N = 41) were compared to those aged 50–65 (N = 42), and there was no significant difference in accuracy, $t(81) = 1.38$, $p = 0.17$. Collectively, these data suggest that age is not a primary source of variability for our participant groups.

Predicting Accuracy

The large sample sizes of the current experiment provided an opportunity to not only compare group differences but also to use regression models to explore the variance within groups. A stepwise linear regression model was conducted for each of the three participant groups (non-professionals, early-career professionals, and experienced professionals) to examine what best predicted overall accuracy. This approach allows us to assess both explanatory value of the regression model as a whole and which individual factor explained the most variance within a given model. Five factors were included in an initial model for each group: (1) age, (2) hit response time slope, (3) overall response time, (4) correct rejection slope, and (5) response time consistency. Age was reported in ranges for the professional visual searchers (18–25, 26–34, 35–49, and 50–65), hit response time slope was how much longer a participant took to locate a target for each additional item in the display, and overall response time was the average response time across all trials. These three factors (age, hit response time slope, and overall response time) did not significantly contribute to the regression models and were removed from any additional analyses.

A stepwise linear regression was subsequently run on the two remaining factors: correct rejection response time slope and response time consistency (see planned analyses in the Methods section). Outliers were assessed based upon a Cook's D equal to or greater than 1 (Cook & Weisberg, 1982), but no data points were trimmed due to this criterion. Additionally, because speed and consistency are calculated from similar response time metrics, we assessed collinearity diagnostics. No factor included in the following models had a variance inflation factor above 1.08, which indicates that collinearity is not an issue for our regression models.

Non-professional searchers. As seen in Figure 4, speed and consistency explained a significant, and remarkably high, amount of the accuracy variance for the non-professional searchers (Adj. $R^2 = 0.69$, $F(2,90) = 102.80$, $p < 0.001$). Both speed ($\beta = 0.686$, $t(92) = 11.36$, $p < 0.001$) and consistency ($\beta = -0.326$, $t(92) = 5.41$, $p < 0.001$) contributed significantly to the model, with speed as the primary contributor ($\Delta R^2 = 0.59$) and consistency[2] as secondary ($\Delta R^2 = 0.10$).

Early-career professional searchers. Speed and consistency explained a significant amount of the accuracy variance for the early-career professional searchers (Adj. $R^2 = 0.42$, $F(2,67) = 25.56$, $p < 0.001$). Both speed ($\beta = 0.585$, $t(69) = 6.29$, $p < 0.001$) and consistency ($\beta = -0.227$, $t(69) = 2.44$, $p < 0.025$) contributed significantly to the model, with speed as the primary contributor ($\Delta R^2 = 0.37$) and consistency as secondary ($\Delta R^2 = 0.05$).

Experienced professional searchers. Speed and consistency explained a significant amount of the accuracy variance for the experienced professional searchers (Adj. $R^2 = 0.60$, $F(2, 93) = 70.64$, $p < 0.001$). Both speed ($\beta = 0.474$, $t(95) = 7.00$, $p < 0.001$) and consistency ($\beta = -0.503$, $t(95) = 7.43$, $p < 0.001$) contributed significantly to the model. Unlike for the non-professional and early-career professional searchers, however, speed was the secondary contributor ($\Delta R^2 = 0.21$) and consistency was primary ($\Delta R^2 = 0.39$).

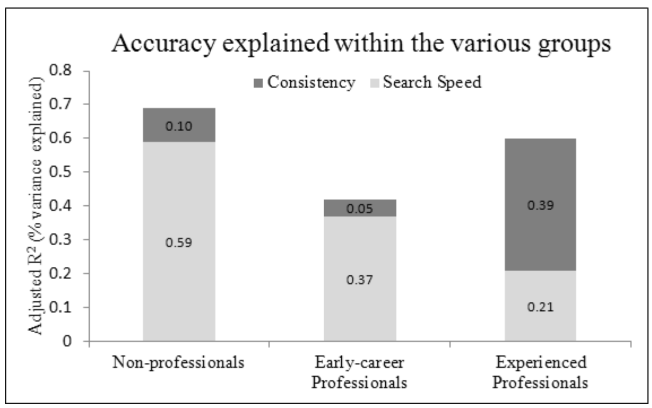

Figure 4. Amount of variance in accuracy accounted for by correct rejection response time consistency (see Equation 1 in text) and speed.

General Discussion

The primary aims of the current study were to explore if and how professional and non-professional searchers differed in performance on a simplified visual search task. For the professional searchers, we recruited Transportation Security Administration Officers who regularly conduct

[2] As noted before, the consistency measure (Equation 1) incorporates the standard deviation of response times for each set size for each participant. While this is potentially susceptible to differences in trial counts, the differences were minimal. Each of the values reported in this section were also calculated with standard error used in place of standard deviation in Equation 1, which reduces the impact of uneven trial counts, and the outcomes were the same.

visual searches as part of their employment, and for the non-professional searchers, we recruited members of the Duke University community. All participants completed a visual search task where they were to find target "T"s amongst distractor "L"s. Our results confirmed the *a priori* prediction that professional searchers would be more accurate than non-professional searchers. Interestingly though, the professional searchers were also significantly slower, both in the time taken to locate a target and the time taken to terminate a search. This response pattern fits a classic speed-accuracy trade-off explanation where the professional searchers are more accurate and slower, suggesting they are performing the task more diligently than non-professional searchers.

An important aspect of the current study is that it allowed for group analyses above and beyond basic search parameter differences. Specifically, we explored what factors best predicted accuracy for professional and non-professional searchers with a focus on search speed and search consistency. Search speed and consistency accounted for a remarkably large percent of the accuracy variability (69%, 42%, and 54% for the non-professionals, early-career professionals, and experienced professionals, respectively), but which factor explained more of the variability differed between groups. The non-professional searchers and early-career professional searchers (those who at the time of testing had been employed by the TSA for three years or less) both exhibited large speed-accuracy trade-offs, where search speed better predicted accuracy than search consistency. In contrast, among experienced professional searchers (those who at the time of testing had been employed by the TSA for six or more years), consistency better predicted accuracy than did search speed.

The search consistency aspects of the current study contribute to interpretations of the current findings and also inform aspects of visual search more broadly. For interpreting the current results, there are two key contributions that come from the consistency data. First, because professional search experience interacted with the factors most predictive of accuracy, a speed-accuracy trade-off explanation cannot fully account for the general accuracy and response time differences between the professional and non-professional searchers. Search consistency best predicted accuracy among experienced professional searchers, whereas speed best predicted accuracy among the early-career professional searchers, demonstrating a more nuanced determination of accuracy performance than just how fast one responds. Second, differences in performance between the early-career and experienced professional searchers reduce concerns over potential motivational or demographic differences between our professional and non-professional searchers. For example, one could be concerned about motivational differences given that the professional searchers performed the task while at work, whereas the non-professional searchers were volunteers who came to our laboratory on the Duke campus. However, this concern, and others like it, is lessened by finding differences *within* the professional population.

Another potential area for concern might be the age differences between our three groups. The non-professional searchers were significantly younger than the two professional groups, and the early-career professional participants were significant younger than the experienced professional participants. However, there are several factors that strongly suggest that experience is a driving force for the effects found here, and not age *per se*. First, age was not a significant factor in the regression models that assess what factors account for variability in accuracy. Age was included as a variable, but did not significantly contribute to the models. Second, accuracy did not differ between age groups among the professional searchers. Third, experienced-match participants from two different age groups (experienced professional searchers that were 35–49 years old compared to 50–65 years old) did not differ in accuracy. Previous research has shown that elderly participants tend to show strong speed-accuracy trade-offs wherein they are significantly slower and more accurate. (e.g., Rabbitt, 1979; Ratcliff, Thapar, & McKoon, 2007). Yet, the current study revealed a speed difference but not an accompanying accuracy difference within our professional population—for the professional searchers, age correlated with speed, but not accuracy. Obviously age can be a strong contributor to visual search performance, but experience appears to be the primary factor at play here. It is possible that our professional participants were too young to possess significant age-related decline. As well, it is possible that their profession has provided a buffer that helps stave off age-related decline (Shimamura, Berry, Mangels, Rusting, & Jurica, 1995). This later interpretation would be exciting, but more work is needed to support any substantial conclusions.

The relationship between consistency and accuracy (e.g., that consistency alone accounted for 35% of the variability in accuracy for the experienced professional searchers) emphasizes the high attentional and memory demands of visual searches. Specifically, a consistent visual search (orthogonal to which specific pattern is employed) can unburden the observer by requiring less effort in recalling previously searched locations. For example, if airport X-ray operators searched bags the same way every time, they could redistribute their cognitive resources from focusing on the search pattern to instead focusing on object recognition. This finding is consistent with previous research that has shown an important role for memory in visual search (e.g., Beck, Peterson, & Vomela, 2006; Cain & Mitroff, 2012; Dickinson & Zelinsky, 2007).

It is interesting that search consistency was highly predictive for the experienced professional searchers but much less so for the early-career professional searchers. There are several possible interpretations of this result. First, this group difference may result from the early-career professional searchers attempting to use a variety of search strategies. Not having adapted to a particular strategy would undoubtedly create substantial differences in how and when particular strategies are employed, thus making it difficult to be truly consistent. Second, consistency itself is likely an experience-dependent mechanism; there should be a learning curve wherein visual searchers are actively trying to maintain consistency, and this active burden could tax available resources much in the way that consistent visual search would unburden them. This possibility could explain why consistency aided the most experienced professional searchers and had much less impact for the early-career professional searchers. Finally, the benefits for consistency for the experienced professional searchers could represent a form of cognitive compensation—to overcome age-related declines in basic visual and attentional abilities, these individuals may rely on consistent visual search to maintain their level of performance. However, given that age was not found to be an important factor along several fronts in the current study, we do not feel this final interpretation is the most likely. This consistency effect is exciting no matter the cause, but regardless, we are currently exploring these hypotheses further with the professional searchers at the airport.

It is important to note that while we interpret the consistency data in terms of both temporal and spatial search factors, the consistency measure in the current study was determined by temporal factors alone. It is reasonable to assume a link between temporal and spatial consistency insofar as truly consistent spatial searches should take the same amount of time to complete on each instantiation, but we did not directly measure spatial search factors here. Our future plans include eye-tracking metrics with both professional and non-professional searchers, but the temporal analyses offered here serve as an important, and necessary, first step. Specifically, temporal consistency metrics can be collected across a large pool of participants regardless of which specific search strategy each participant adopted. Moreover, temporal consistency can be collected using simple behavioral metrics based upon response time, which offers an effective, predictive, and easy-to-collect performance metric. These characteristics make temporal consistency a valuable metric to use in professional settings.

Finally, it is interesting to consider the nature of the visual search task employed here. We utilized abstract, domain-general stimuli so that our non-professional and professional searchers would be equally capable of executing the task, and yet, we still observed differences in fundamental search behaviors among our participant groups. This evidence suggests that the effects of professional training and experience are at least somewhat transferrable to generalized visual search behaviors. Moreover, if specific search behaviors were important for the professional searchers on this abstract task, then they are likely equally important, if not more important, on their far more complicated professional task. Previous research, particularly as it relates to expertise, has shown performance can be enhanced through adaptation to task constraints (Ericsson & Lehmann, 1996). A simplified task, such as the one employed here, does not possess terribly demanding task constraints, yet we were still able to reveal benefits for the professional searchers. This suggests that the effects revealed here may be even more pronounced for professional searchers domain-specific search tasks.

Conclusions

In summary, the current study revealed differences between professional and non-professional searchers in fundamental aspects of visual search ability. This difference underscores the importance of top-down control and represents the most direct means of improving visual search performance on real-world tasks. Furthermore, this study also informs real-world visual searches by providing a better understanding of generalized search behaviors that are most associated with professional training. Search consistency, in particular, is a clear means of distinguishing between the effectiveness of early-career and experienced professional searchers. Consistency is a valuable performance variable as it may be critical in allowing visual searchers to focus their cognitive resources on correctly identifying targets rather than expending resources on remembering where they have or have not already searched. Additionally, consistency provides an easily accessible mechanism for improving accuracy in visual search, even if it requires experience before becoming fully effective. It is recommended that consistency be emphasized during professional visual search training and evaluation.

References

Awh, E., Belopolsky, A.V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: a failed theoretical dichotomy. *Trends in Cognitive Sciences, 16,* 437-443.

Beck, M.R., Peterson, M.S., & Vomela, M. (2006). Memory for where, but not what, is used during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(2), 235-250.

Biederman, I., & Shiffrar, M. M. (1987). Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 640-645.

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.

Bradley, B.P., Mogg, K., Wright, T., & Field, M., (2003). Attentional bias in drug dependence: Vigilance for cigarette-related cues in smokers. *Psychology of Addictive Behaviors, 17,* 66-72.

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

Cain, M.S., & Mitroff, S.R. (2012). Memory for found targets interferes with subsequent performance in multiple-target visual search. *Journal of Experimental Psychology: Human Perception & Performance*. DOI: 10.1037/a0030726

Cook, R.D., & Weisberg, S. (1982). *Residuals and influence in regression* (Vol. 5). New York: Chapman and Hall.

Dickinson, C.A., & Zelinsky, G.J. (2007). Memory for the search path: Evidence for a high-capacity representation of search history. *Vision research*, *47*, 1745-1755.

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

Ericsson, K.A., & Lehmann, A.C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. Annual Review of Psychology, 47, 273-305.

Field, M., Mogg, K., & Bradley, B.P. (2004). Eye movements to smoking-related cues: Effects of nicotine deprivation. *Psychopharmacology*, *173*, 116-123.

Field, M., Eastwood, B., Bradley, B.P., & Mogg, K. (2006). Selective processing of cannabis cues in regular cannabis users. *Drug and Alcohol Dependence, 85,* 75-82.
Folk, C.M., Remington, R.W., & Johnston, J.C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1030-1044.
Greenwald, A.G., McGhee, D.E., & Schwartz, J.L.K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,* 1464-1480.
Jones, B.C., Jones, B.T., Blundell, L., Bruce, G. (2002). Social users of alcohol and cannabis who detect substance-related changes in a change blindness paradigm report higher levels of use than those detecting substance-neutral changes. *Psychopharmacology, 165,* 93-96.
Jones, B.T., Jones, B.C., Smith, H., Copley, N. (2003). A flicker paradigm for inducing change blindness reveals alcohol and cannabis information processing biases in social users. *Addiction, 98*, 235-244.
Kleiner, M., Brainard, D.H., & Pelli, D.G. (2007). What's new in Psychtoolbox-3? *Perception, 36,* ECVP Abstract Supplement.
Krupinski, E.A. (2010). Current perspectives in medical image perception. *Attention, Perception, & Psychophysics, 72,* 1205-1217.
McCarley, J.S., Kramer, A.F., Wickens, C.D., Vidoni, E.D., & Boot, W.R. (2004). Visual skills in airport security screening. *Psychological Science, 15,* 302-306.
Menneer, T., Barrett, D.J.K., 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.
Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437-442.
Rabbitt, P.M.A. (1979). How old and young subjects monitor and control responses for accuracy and speed. *British Journal of Psychology, 70,* 305-311.
Ratcliff, R., Thapar, A., & McKoon, G. (2007). Application of the diffusion model to two-choice tasks for adults 75-90 years old. *Psychology and Aging*, 22, 56–66.
Schwaninger, A. (2003). Evaluation and selection of airport security screeners. *AIRPORT*, 02/2003, 14-15.
Schwaninger, A. (2005). Increasing efficiency in airport security screening. *WIT Transactions on the Built Environment, 82,* 405-416.
Schwaninger, A., Hardmeier, D., & Hofer, F. (2004). Measuring visual abilities and visual knowledge of aviation security screeners. *IEEE ICCST Proceedings,* 258-264.
Shimamura, A.P., Berry, J.M., Mangels, J.A., Rusting, C.L., & Jurica, P.J. (1995). Memory and cognitive abilities in academic professors: Evidence for successful aging. *Psychological Science, 6,* 271-277.
Smith, M.J. (1967). *Error and variation in diagnostic radiology*. Springfield, IL: Thomas.
Theeuwes, J. (1991). Cross-dimensional perceptual selectivity. *Perception & Psychophysics, 50,* 184-193.
Theeuwes, J. (1992). Perceptual selectivity for color and form. *Perception & Psychophysics, 51,* 599-606.
Theeuwes, J. (2004). Top-down search strategies cannot override attentional capture. *Psychonomic Bulletin & Review, 11,* 65-70.
Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica, 135,* 77-99.
Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95,* 15-48.
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.
Waters, A.J., & Feyerabend, C. (2000). Determinants and effects of attentional bias in smokers. *Psychology of Addictive Behaviors, 14,* 111-120.
Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13-74). Hove, East Sussex, UK: Psychology Press Ltd.
Yantis, S. (2000). Goal directed and stimulus driven determinants of attentional control. In S. Monsell & J. Driver (Eds.), *Attention & performance XVIII* (pp. 73-104). Cambridge, MA: MIT Press.

Author Note

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 and Emma Wu Dowd for helpful comments, and Ricky Green 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 Human Factors 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 U.S. Government. The study is approved for public release. Matthew S. Cain is now at the Department of Cognitive, Linguistic & Psychological Sciences, Brown University, Providence, Rhode Island. Correspondence to: Adam T. Biggs; adam.biggs@duke.edu