AN INDIVIDUAL DIFFERENCES APPROACH TO IMPROVING
LOW TARGET PREVALENCE VISUAL SEARCH PERFORMANCE

By

Chad Peltier

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Psychology – Doctor of Philosophy

2017

ABSTRACT

AN INDIVIDUAL DIFFERENCES APPROACH TO IMPROVING
LOW TARGET PREVALENCE VISUAL SEARCH PERFORMANCE

By

Chad Peltier

Critical real-world visual search tasks such as radiology and baggage screening rely on the detection of rare targets that may only be present on as few as .3% of searches (Gur et al., 2004). When targets are rare, observers search for a shorter amount of time and miss targets more often than when targets are common, a phenomenon known as the low prevalence effect (LPE). Given the real-world importance of the detection of low prevalence targets, researchers have attempted to improve search performance. There have been several experimental attempts to reduce the LPE, but none have been wholly successful, as even the best methods have increased hits at the cost of more false alarms. As an alternative to improving visual search performance through experimental manipulations, researchers have recently started using an individual differences approach to predict those who would be best at rare target detection. The individual differences approach has shown that it is possible to predict low prevalence target detection using working memory capacity (WMC) (Peltier & Becker, 2016b; Schwark et al., 2012) and moderate prevalence target detection using a personality assessment (Biggs, Clark, & Mitroff, 2017) and vigilance (Adamo, Cain, & Mitroff, 2016). Experiment 1 expands on the previous research by predicting low prevalence visual search performance using measures of WMC, near transfer high prevalence visual search accuracy, vigilance, attentional control, and introversion. The regression using these predictors accounts for 52% of the variance in accuracy. Experiment 2 addresses practical and theoretical limitations of Experiment 1 by replicating the original finding, including new potential predictors of low prevalence search performance (fluid

intelligence, task unrelated thought frequency, and far transfer search accuracy), using more realistic search stimuli to increase external validity, and using eye tracking to investigate how individual differences relate to specific components of performance. The results show that near transfer search, far transfer search, WMC, introversion, and fluid intelligence account for 53% of the variance in accuracy in a more realistic low prevalence search. Using the beta weights from Experiment 1's significant predictors and each observer's score on the corresponding measures in Experiment 2, I find that the old predictors account for 42% of the variance in a novel search task's accuracy. Finally, the eye-tracking results show that we can significantly predict quitting thresholds (the number of items inspected before terminating search), selection error rates (misses caused by never inspecting the target), identification error rates (misses caused by misidentifying an inspected target), item re-inspection rates, target decision times, and distractor decision times. I conclude that the individual differences approach has the potential to be a highly effective tool in selecting those who are most likely to perform at a high level in real-world searches.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

EMF   Eye Movement Feedback

K   Measure of Working Memory Capacity (Pashler, 1988)

LPE   Low Prevalence Effect

MDM Multiple Decision Model (Wolfe and Van Wert, 2010)

TSA   Transportation Security Administration

TUT   Task Unrelated Thought

WMC Working Memory Capacity

CHAPTER 1. INTRODUCTION

Visual search is a common task that we perform on a daily basis. Due to its ubiquity, there has been a great deal of research investigating visual search, such as how we guide our attention (Bays & Husain, 2012; Williams, 1966; Wolfe, 1994), how we recognize objects (Haxby et al., 2001; Treisman & Gelade, 1980), and how we decide to terminate search (M. Chun & Wolfe, 1996; Wolfe, 2014). Despite this long history of research, we have, until recently, ignored a critical characteristic of some real-world searches: low target prevalence. Most of the research on visual search uses a target prevalence rate of 50% or 100%, but this does not accurately represent some real-world searches, limiting the generalizability of past research.

As target prevalence in visual search tasks decreases from 50% to 10%, the chances of an observer failing to detect a target increases dramatically, with miss rates up to two to five times greater (Ishibashi, Kita, & Wolfe, 2012; Rich et al., 2008; Schwark, Sandry, Macdonald, & Dolgov, 2012; Wolfe, Horowitz, & Kenner, 2005; Wolfe et al., 2007). This effect is known as the low prevalence effect (LPE), and is a major concern because many critical, real-world search tasks have a target prevalence rate below 1%.

Radiology and baggage screening are both low prevalence search tasks. In mammography screening, it is estimated that only .3% of inspected scans will contain a tumor (Gur et al., 2004). Though information about the prevalence rate of banned items in baggage screening is classified or unknown, the fact that the Transportation Security Administration (TSA) funds research to improve rare target detection suggests that targets are rare in baggage screening (Wolfe et al., 2007). Consistent with lab results investigating low prevalence search tasks, targets are often missed in these real-life, low prevalence search tasks. Radiologists miss as many as 30% of cancers in their examinations (Berlin, 1994; Evans, Birdwell, & Wolfe, 2013)

and a recent investigation of TSA performance found that TSA agents missed over 95% of weapons that investigators attempted to smuggle through screening (Fishel, Levine, & Date, 2015).

There are several potential causes of the low prevalence effect, some of which have been studied directly. Potential causes include a shift in decision making criterion (Peltier & Becker, 2016a; Wolfe & Van Wert, 2010), a change in quitting threshold (Wolfe et al., 2007; Wolfe & Van Wert, 2010), motor bias (Fleck & Mitroff, 2007), or a lack of vigilance (Ariga & Lleras, 2011; McVay & Kane, 2012; Warm, Parasuraman, & Matthews, 2008).

1.1 Causes of the Low Prevalence Effect

1.1.1 Criterion Shift

Wolfe and Van Wert (2010) proposed the Multiple-Decision Model (MDM) which states that when an item is being inspected, the observer makes a two-alternative forced choice about its identity as a target or distractor. This choice is modelled using signal detection theory (Green & Swets, 1966), where choices are determined by sensitivity and criterion. The MDM states that sensitivity is unchanged by prevalence, but as prevalence decreases, the observer's item by item decision making criterion becomes more conservative. A conservative criterion predicts that observers will commit more identification errors (failure to recognize a fixated target), which increases the miss rate and contributes to the LPE.

The MDM has been directly tested using a combination of eye-tracking and a prevalence manipulation in order to analyze the number of times a target was fixated, but went unrecognized as a function of target prevalence. Several researchers found that as prevalence decreased, the probability of recognizing a fixated target decreased (Godwin, Menneer, Riggs, Cave, &

Donnelly, 2014; Godwin, Menneer, Riggs, et al., 2015; Hout, Walenchok, Goldinger, & Wolfe, 2015), supporting the theory of a prevalence based criterion shift. Though these researchers supported the use of signal detection theory to model the decision making process, one group (Peltier & Becker, 2016a) proposed that drift diffusion would better model the decision making process. The advantage of using drift diffusion (Ratcliff & McKoon, 2008) is that it makes explicit predictions about the time to reach a decision when criterion shifts in a two-alternative forced choice decision. Therefore, a conservative shift in criterion predicts slower decisions to the target boundary and faster decisions to the distractor boundary when compared to a higher prevalence search where target decisions are fast and distractor decisions are slow. These predictions were supported by Peltier & Becker (2016a), and fit with data from others (Godwin, Menneer, Cave, Thaibsyah, & Donnelly, 2015; Godwin, Menneer, Riggs, et al., 2015; Hout et al., 2015) who had only analyzed target decision making time and found that it increased as prevalence decreased.

The use of eye tracking has supported the theory that target prevalence changes an observer's decision-making criterion. As target prevalence decreases, criterion becomes more conservative resulting in an increase in target decision time. Critically, this shift in criterion also predicts an increase in target misidentifications, which contributes to the LPE.

1.1.2 Quitting Threshold

An observer's quitting threshold is the number of items they are willing to inspect before terminating search (Wolfe, 2014). Once the observer reaches their quitting threshold without finding a target, they respond target absent (Wolfe & Van Wert, 2010) (but see Peltier and Becker (2017b)). As target prevalence decreases, the observer's quitting threshold decreases, resulting in fewer item inspections (Hout et al., 2015; Rich et al., 2008). Fewer item inspections

3

lead to faster reaction times (Hout et al., 2015; Wolfe et al., 2005), and increases the probability of a selection error, where the observer fails to find a target, then incorrectly responds target absent after reaching their quitting threshold (Peltier & Becker, 2016a).

Decreasing target prevalence decreases observers' quitting thresholds. Once these thresholds are lowered, the observer inspects relatively few items before making a target presence judgement. This decreases both reaction time and hit rate.

1.1.3 Vigilance Decrement

Vigilance is the ability to maintain attention and alertness over long periods of time (Davies & Parasuraman, 1982). As the duration of a task requiring vigilance increases, performance decreases (Davies & Parasuraman, 1982). Vigilance is a common requirement in some professions, including radiology and baggage screening, where observers are trying to detect a rare signal in visually noisy displays (Wolfe et al., 2007).

In the current literature, the relationship between vigilance and low prevalence visual search performance is largely theoretical, though there are some suggestive correlations. In one study, Adamo et al. (2016) correlated vigilance performance with moderate prevalence visual search performance and found that vigilance scores were positively correlated with accuracy, suggesting that those with higher vigilance abilities are more accurate in visual searches. Though this theory is currently untested, it seems likely that vigilance, a measure of sustained attention, would be an even stronger predictor of low target prevalence than moderate target prevalence accuracy.

1.1.4 Motor Bias

In low prevalence search tasks, observers build up the expectation of responding target absent because that is the correct response on the vast majority of trials. Fleck and Mitroff (2007) claimed that the increased miss rate in low prevalence search may be due to increased motor errors where observers mistakenly execute the prepotent target absent response. If a motor bias was the only cause of the LPE, as they propose, then giving observers the opportunity to correct their responses should eliminate the high miss rates associated with low prevalence search. This is exactly what Fleck and Mitroff (2007) found; when observers were able to correct their initial responses, the LPE disappeared. These data suggest that the LPE, at least in part, is caused by an easily correctible motor bias, but as discussed below, this hypothesis is not always supported.

1.2 Experimental Methods to Reduce the Low Prevalence Effect

Given the critical nature of the LPE, researchers have also investigated ways to improve rare target detection. These investigations have included individual difference approaches (Peltier & Becker, 2016b; Schwark, Sandry, & Dolgov, 2013) and experimental manipulations (Navalpakkam, Koch, & Perona, 2009; Schwark et al., 2012; Wolfe et al., 2007). Given that we know potential causes of the LPE, successful methods to reduce the LPE should relate to one of the above causes.

1.2.1 Criterion Shift

Given a primary cause of the LPE is a conservative criterion, several attempts have been made to shift decision making criterion to become more liberal. These manipulations, if successful, would be helpful in reducing the LPE through the reduction in identification errors. One reason to be wary of this criterion adjusting approach is that any manipulation that shifts

criterion to be more liberal will also increase false alarms, which can be costly; a false cancer diagnosis would be extremely costly to a family, while a false alarm on a weapon in a security screening could be time consuming and embarrassing.

Two studies shifted decision making criterion by increasing the perceived target prevalence (Schwark et al, 2012, Wolfe et al., 2007). The researchers argue that observers use correct/incorrect feedback to adjust their criterion in visual search tasks, thus by giving feedback to indicate a missed target, observers should shift their criterion to be more liberal in order to detect future targets. Wolfe et al. (2007) accomplished this by introducing bursts of high prevalence trials in the middle of a low prevalence search task. In this experiment, observers performed a standard low prevalence search task, then would periodically perform a block of high prevalence trials during which they were given accuracy-based feedback. If observers maintained their conservative criterion from the low prevalence block of trials, they would often miss the targets in the higher prevalence blocks, and receive feedback that they missed the target, leading to a more liberal criterion that they would carry over into the next block of low prevalence trials. The blocks of high prevalence trials successfully shifted criterion to be more liberal and this effect carried over into the following low prevalence block. However, this method has two drawbacks. First, as previously stated, the liberal criterion increased hits, but it also increased false alarms. Second, the extra high prevalence blocks would need to be performed regularly to maintain a liberal criterion, which would be time consuming and potentially expensive as additional observers would need to work to fill in for the observer who is undergoing the retraining procedure. Though blocks of high prevalence trials can aid in increasing target detections, the cost of this manipulation may be too high.

Schwark et al. (2012) increased perceived target prevalence by giving false feedback on 20% of correct rejection trials to indicate that the observer had missed the target. The manipulation was successful, in that it increased hits in low prevalence searches, but it shared the same weakness as Wolfe et al.'s (2007) manipulation: a liberal criterion also increases potentially costly false alarms. Further, evidence from Peltier and Becker (2017b) suggests that the increase in hit rates may not be solely due to a decrease in identification errors. Peltier & Becker found that increasing target prevalence leads to an increase in target present guessing without actual target identifications, thus observers are responding to task demands, rather than improving search. This increased target present guessing would also explain the increase in false alarms in both Schwark et al. and Wolfe et al. Through both Schwark et al. (2012) and Wolfe et al.'s (2007) experiments we learn that one solution to the LPE is pushing an observer's decision making criterion to be more liberal, but doing so introduces new problems: extra search trials, which take time, cost money, and possibly fatigue observers, all the while increasing false alarms.

Rather than attempting to increase perceived target prevalence, Navalpakkam et al. (2009) attempted to increase performance through monetary rewards. Rewards were manipulated in a low prevalence visual search task to punish misses or reward hits. Neither reward manipulation improved performance until the monetary rewards were converted to points and the observers were told that the points would be used in a competition where the observer with the most points would win a large prize. Though the competition combined with a reward system effectively improved low prevalence hit rates, similar to other experimental manipulations, it also resulted in an increase in false alarms, indicating a criterion shift. Given the mixed results of these experiments, a lack of power (n=4 in each manipulation), and a lack of

a clear explanation as to why direct rewards and competition-based rewards would lead to different results, it is unclear if either experiment is a true result. This experiment also has a literal cost beyond that of Schwark et al. and Wolfe et al: the actual money used to motivate observers to perform at a higher rate. Additionally, it is not known how these monetary manipulations would affect a professional observer who is already paid, and presumably motivated without an additional incentive.

1.2.2 Change in Quitting Threshold

As misses caused by a low quitting threshold are the primary cause of the LPE (Peltier and Becker, 2016a), a method that successfully increases quitting thresholds should theoretically result in a greater benefit in performance than criterion shifts.

In their continuing effort to improve search performance, Wolfe et al. (2007) attempted to increase quitting thresholds by giving observers "speeding tickets" on trials where observers responded too quickly, which then required observers to view the original search display again and make an additional target present/absent decision. The goal was to slow observers in order to increase their quitting thresholds and reduce selection errors. The manipulation was not successful. Observers' reaction times slowed, but their hit rate did not increase. This result is interesting because quitting thresholds were seemingly increased without improving performance. Without eye-tracking, it is impossible to know definitively, but it is possible that once a "natural" quitting threshold is reached, the observer stops actively searching the display and simply waits a short period of time before responding target absent to satisfy the experimental condition. Unfortunately, there has not been any published data addressing this concern, but future research should attempt to determine what observers do with this extra time spent on a trial as well as ways to encourage "active search" during this time.

A recent investigation by Peltier and Becker (in press) tested the effectiveness of "eye movement feedback" (EMF) where observers were eye-tracked during visual search for a rare target. The eye-tracker was used to mark fixated areas by either removing an overlay as portions of the display were fixated or by adding the overlay once the eye left a segment of the image. One experiment provided automated EMF, such that a new region was uncovered at a constant rate. We hypothesized that feedback would guide attention to unsearched areas and increase the proportion of the display searched before observers made a target absent response, thereby increasing accuracy. Different experiments required observers to search for either Waldo, from "Where's Waldo?" children's books or Ts among Ls. We found some evidence that EMF may be effective in one experiment, but the remaining experiments either showed no effect of EMF, or that EMF actually reduced accuracy. We concluded that the one positive result we found was likely a Type I error and that the EMF method that we used is unlikely to improve visual search performance. At the theoretical level this suggests that the quitting thresholds and decision criterion that have been deemed responsible for the low-prevalence effect (Hout et al., 2015; Peltier & Becker, 2016a; Wolfe & Van Wert, 2010), while sensitive to target prevalence, may be insensitive to other types of manipulations.

1.2.3 Eliminate Motor Biases

Inspired by Fleck and Mitroff's (2007) proposal that misses are caused by a prepotent target absent motor responses, Wolfe et al. (2007) attempted to increase search performance by eliminating the target absent motor bias. In order to prevent a motor bias from building, at least one of three varying prevalence (1%, 10%, 44%) targets were set to appear on 50% of trials. Despite preventing a motor bias by maintaining equal target present/absent response probabilities, the results still showed a strong LPE such that the lower prevalence targets were

missed significantly more often than higher prevalence targets. While these data suggest that the Fleck and Mitroff finding was a false positive, two contradicting studies are inconclusive.

However, further research by Rich et al. (2008) clarified the contribution of motor biases to the LPE. Rich et al used both efficient (vertical and horizontal lines) and inefficient search stimuli (Ts among offset Ls) and enforced a delay before a response could be made (500ms minimum in efficient, 2000ms minimum in inefficient) to prevent early target absent responses that were assumed to be caused by low prevalence motor biases. The results showed that the delay before making a response prevented the LPE in the efficient, popout search condition, but did not prevent the LPE in the inefficient search. This suggests motor errors may occur when decisions are rapid, but motor biases likely cannot account for the LPE in inefficient searches. Indeed, current work by Peltier and Becker (in prep) that has manipulated the time it takes to reach a decision about a target's presence has found that motor biases no longer impact search performance when a search task takes 3 or more seconds. This finding, combined with the fact that TSA searches take 4 seconds on average (Schwaninger, Hardmeier, & Hofer, 2005), suggests that motor biases do not have an effect on serial low prevalence visual search performance and controlling or preventing motor biases does not improve search.

The purpose of describing the experimental methods that failed to minimize the low prevalence effect is to illustrate the variety of methods that have been attempted and highlight how difficult it has been for researchers to improve search for rare targets. Each experimental manipulation that succeeded in improving hit rates also increased false alarms, which can be costly. There were other costs in several of these manipulations as well; increased time (blocks of high prevalence trials), and increased pay for individuals (competitive reward manipulation).

Given the difficulty improving an individual's search performance, an alternative approach is to identify those individuals who are better at low prevalence search tasks.

CHAPTER 2. EXPERIMENT 1

2.1 Introduction

Thus far, experimental methods to improve low prevalence search have largely failed, or had associated costs that limit their applicability to real-world situations. An alternative approach is to find individual differences that predict visual search performance. This method measures performance in various tasks that predict low prevalence search performance, with the intention that employers will then use these predictor tasks to find those who will perform low prevalence search effectively. Though a relatively novel approach, the limited research available suggests this can be an effective method of increasing rare target detection.

Research by Schwark et al. (2013) provides data that suggests both of these requirements might be met. Schwark et al. (2013) found that there are large individual differences in low prevalence search performance, with hit rates ranging from 0% (3 subjects out of 40) to 100% (8 subjects). Given these large individual differences in performance, they investigated whether individual differences in working memory capacity (WMC) could predict low prevalence search performance. They found a significant relationship between WMC and low prevalence search performance, such that those with high WMC had higher hits rates and slower target absent search times. They attributed the high target detection among those with higher WMC to maintaining high quitting thresholds in low prevalence tasks, meaning those with higher WMC searched for a target longer before terminating search.

Encouraged by their work, here I attempt to identify additional predictors of rare target detection with the goal of developing a screener that can be used to identify people who would be particularly effective at detecting rare targets. Given that there has been limited research

investigating predictors of low prevalence visual search performance, this work is largely exploratory. A handful of cognitive that might be associated with rare search accuracy were chosen. These include measures of working memory capacity (WMC), vigilance, attentional control, and performance on a near transfer search that has identical stimuli with a higher prevalence rate. Also investigated was whether any of the big five personality factors (openness/intelligence, conscientiousness, extraversion, agreeableness, neuroticism) (Goldberg, 1990) could add additional ability to identify individuals who would be good at low prevalence search. Ultimately, the goal is to assemble a battery of tasks that can be used to reliably predict low prevalence search accuracy.

While the main goal was to find predictors of search accuracy, I also investigated how these factors were related to reaction time and false alarm rates. The reason I did so was to investigate the potential mechanisms by which the predictors influence target detection rates. If the predictors of good target detection are also predictors of slower search reaction times, the data would be consistent with the mechanism being the individual's quitting threshold; predictors of a high quitting threshold should be associated with both slower reaction times and better target detection. By contrast, if the mechanism was simply a change in the decision criterion for responding target present, then predictors of higher target detection rates should also be predictive of more false alarms. In short, investigating these two additional dependent variables provided preliminary evidence about the potential underlying mechanism by which the predictors influence low prevalence search performance.

2.2 Methods

2.2.1 Participants

One hundred fifty-eight undergraduates (109 females) from Michigan State University's human subjects pool gave consent to participate in the study for course credit. All subjects were between the ages of 18 to 24 with normal or corrected to normal vision. Fourteen subjects were excluded from further analysis for failing to complete all tasks.

2.2.2 Low Prevalence Visual Search Task

The task was to search for a rotated T among an array of 24 items and respond present or absent via button press (See Figure 1). Distractors were rotated, offset Ls. The use of these offset Ls makes the search task far more difficult and less efficient than the typical T among Ls search task. In target absent trials all 24 stimuli were distractors. In target present trials, one randomly chosen L was replaced with a rotated T. The orientation of each item was randomly assigned to be 0, 90, 180, or 270 degrees from vertical.
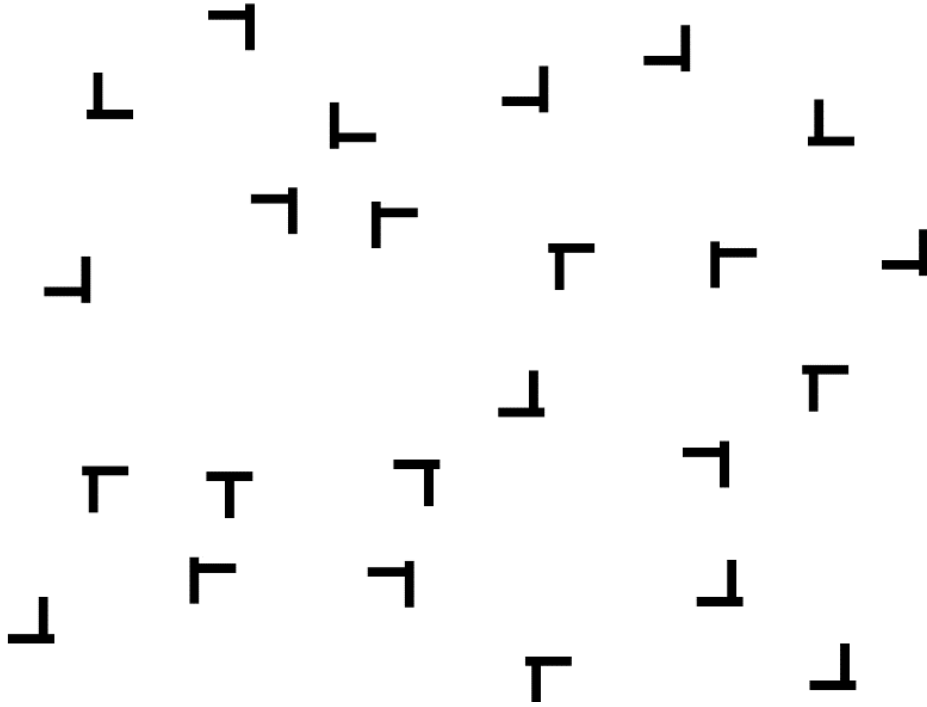
Figure 1. *Example image from Ts and Ls visual search task.* The target T is in the lower left quadrant.

Each item subtended 1.2° x 1.2° of visual angle. To create each array, the screen was divided into 24 (a 6 by 4 matrix) equal sized (6.4°X7.1°) regions. A single item was placed within each region, with random jitter that allowed the item to appear anywhere within the region. This jitter broke up the orderly organization of the matrix and resulted in the items appearing in different locations across trials.

In this low prevalence task there were 27 target present trials randomly interleaved with 243 target absent trials for a prevalence rate of 10%. The block of trials was preceded by 50 practice trials with a 10% target prevalence rate in order to allow search parameters (quitting threshold and decision criterion) to be set for a low prevalence task (Ishibashi et al., 2012; Wolfe & Van Wert, 2010).

2.2.3 Predictor Tasks

2.2.3.1 Near Transfer High Prevalence Visual Search Task

Finding that near transfer high prevalence search performance is a good predictor of low prevalence search performance would provide an efficient method for predicting who would be good at low prevalence search tasks. For instance, the target prevalence rate in cancer screening is ~.3% (Gur et al., 2004), thus it would take several thousand trials to gather reliable data about an observer's performance at these low prevalence levels. However, if near transfer high prevalence search performance is a good predictor of low prevalence search performance, then one could gather data about an individual's performance very quickly at a high prevalence rate and use that as a predictor of the unobserved low prevalence performance. The inclusion of this near transfer high prevalence block of trials also allowed the confirmation that subjects were demonstrating the traditional low-prevalence effect (Wolfe et al., 2005).

The near transfer high prevalence visual search task was identical to the low prevalence task, except that the target prevalence rate was set at 50%. Like the low prevalence task, there were 27 target present trials, for a total of 54 trials in the block. The block of trials was preceded by 50 practice trials with a 50% prevalence rate in order to allow search parameters (quitting threshold and decision criterion) to be set for a 50% prevalence task (Ishibashi et al., 2012; Wolfe & Van Wert, 2010). To calculate high prevalence performance, the predictor variable I use in the analyses, I subtracted false alarms from hits for each participant.

2.2.3.2 Change Detection (Working Memory Capacity)

Schwark et al. (2013) found a positive relationship between performance on the AOSPAN task (Unsworth, Heitz, Schrock, & Engle, 2005) and low prevalence hit rate. They

interpreted this as evidence that WMC was related to rare target detection. However, the AOSPAN task most likely measures both capacity and executive attention (Unsworth et al., 2005). To obtain a more pure measure of capacity, I used the change detection task popularized by Vogel et al. (2001) to measure WMC.

Observers viewed a display of 4, 6, or 8 colored squares for 100ms and tried to remember the color and location information during a 900ms retention interval. After the retention interval a single colored probe square appeared in one of the previously occupied locations. Participants had to indicate whether the color of the probe square matched the color at that location during the original display. The task consisted of 120 trials and took approximately 10 minutes to complete. I used the formula from Pashler (1988) to calculate each observer's capacity (K) from the observer's accuracy data.

2.2.3.3 AX Continuous Performance Task (Vigilance)

Vigilance is the ability to sustain attention over long periods of time (M. M. Chun, Golomb, & Turk-Browne, 2011). Given that low prevalence search tasks require observers to maintain the attentional goal of detecting the target over long periods of time, a task that can quickly assess vigilance may be a useful and valid predictor of low prevalence search.

We measured vigilance using a go/nogo continuous performance task (Covey, Shucard, Violanti, Lee, & Shucard, 2013). Four hundred ten letters were presented one at a time at fixation for 400ms each with a 1500ms interval between letters, for a total task duration of approximately 15 minutes. Observers were instructed to make a button press when they detected an X that was preceded by an A (go trial), which happened 40 times across the sequence of 410 letters. The letters A and X appeared an additional 40 times each without being paired together in the A then

X "go" combination. Each observer's vigilance score was calculated by subtracting the corrected hit rate (hits minus false alarms) in the first quarter of trials from the last quarter of trials.

2.2.3.4 Posner Cuing (Attentional Control)

Serial visual search tasks, such as the task Ts and Ls task used here, require a series of endogenous attentional shifts from item to item (Wolfe, 1994). This process involves disengagement from the currently attended item, a shift in spatial attention, then reengagement onto the new item. Those who are fast to perform these processes may be faster or more effective in visual search tasks.

We used a modified Posner cuing task (Posner, 1980) with a central endogenous cue to measure reaction to validly cued, neutral, and invalidly cued targets. The trial sequence consisted of a fixation point, followed by a central arrow cue that pointed to the left or the right, followed by a black square that could appear on the left or right of fixation. Observers were to report the location of the target (left/right) as quickly and accurately as possible. In half the trials the arrow cue was a neutral cue consisting of a two headed arrow that pointed to both potential target locations. In the remaining half of the trials, the cue was a unidirectional arrow that pointed to the eventual target location (valid cue) 75% of the time. The cue pointed to the wrong location (invalid cue) in the remaining 25% of unidirectional cue trials.

Cues appeared for 250ms and targets appeared for 100ms with a 550ms blank between cue and target. Time between target presentation and the start of the next trial randomly varied between 3000, 4000, and 5000ms. Observers responded if the target was on the right or left side using a button press. The task consisted of 108 trials and took approximately 15 minutes to

complete. In later text, this variable will be referred to as attentional control, and was calculated using the difference between invalid and valid trials' reaction times in correct trials.

2.2.3.5 Mini International Personality Item Pool (Personality)

Personality traits are a potential predictor of visual search performance. There is conflicting evidence over whether introverts are better at search (Sen & Goel, 1981), or perform equally well in comparison to extraverts (Newton, Slade, Butler, & Murphy, 1992). However, a meta-analysis of 53 studies (Koelega, 1992) has shown that introverts consistently perform better than extraverts on tasks requiring sustained attention as measured by hit rate. Introverts also show a smaller performance decrement as time on task increases. These results can potentially be attributed to the theory that introverts have a higher base level of arousal, which allows them to perform monotonous tasks (like low prevalence searches) at a high level (Eysenck, 1967).

Other personality factors may be of interest as well. A meta-analysis of 65 studies (Judge & Ilies, 2002) investigated the relationship between the Big Five personality traits and performance motivation. The study showed that neuroticism and conscientiousness were the best predictors of performance motivation. It is possible that performance motivation predicts effort and quitting thresholds in low prevalence searches and thus could be correlated with accuracy. In sum, personality factors are reasonable candidate predictors of visual search performance and thus were included in the design.

In accordance with my effort to build a battery of measurements of maximum utility in a real-world setting where speed of assessment is important, I measure the Big Five personality traits using the Mini IPIP (Donnellan, Oswald, Baird, & Lucas, 2006), which is only 20 questions. The Mini IPIP is similar in reliability, convergent validity, and criterion validity

measures to longer personality assessments, while still tapping nearly the same content, despite being an abbreviated assessment (Donnellan et al., 2006).

2.2.4 Procedure

Observers first completed the two blocks (high prevalence and low prevalence) of visual search trials. The order of these blocks was randomized for each observer. Observers then completed the following tasks in the same order: change detection, vigilance, Posner cuing, and Mini IPIP.

All tasks were programed in E-prime, and presented individually in sound attenuated booths, on PCs with 20-inch CRT monitors set at a resolution of 1024x768 with a 100Hz. refresh rate. Each task began with on-screen instructions about the upcoming task, and participants were able to take brief breaks between each task.

2.2.5 Data Preparation

Prior to conducting the analyses below, the data was filtered for outliers. This included eliminating subjects who had accuracy in the low prevalence visual search task below 3 standard deviations from the mean. This eliminated three subjects from further analyses, leaving a final sample size of 141. Visual search trials with a reaction time beyond three standard deviations from the mean for each subject at each prevalence rate were also discarded from further analysis, resulting in the exclusion of 1.1% of low prevalence trials and .6% of high prevalence trials.

To investigate the ability of individual difference factors to predict low prevalence search performance, I performed linear regression models. For all regression models, I ensured that the assumptions of the regressions models were met in the following manner. Linearity was assessed through visual inspection of partial regression plots and a plot of studentized residuals

against the predicted values. Homoscedasticity was assessed by visual inspection of a plot of studentized residuals versus unstandardized predicted values. Normality was assessed by visual inspection of the P-P Plot. Durbin-Watson statistics were all between 2.05 and 2.19, while tolerance values ranged from .77 to .93, thus showing no violations of independence of residuals or multicollinearity.

2.3 Results

2.3.1 The Low Prevalence Effect: Reaction Time and Accuracy

To verify that observers exhibited the traditional LPE, I analyzed the data from the visual search blocks using two 2 (target present/target absent) X 2 (low/high prevalence block) repeated-measures ANOVAs, one on the corrected hit rate (percentage hits minus percentage false alarms) data and one on the reaction time data. The ANOVAs were followed-up with planned paired sample t-tests to verifying the presence of the LPE.

We found standard prevalence effects on accuracy (see Figure 2); as target prevalence decreased the proportion of misses increased. This was confirmed by a significant interaction between target prevalence and target presence, $F(1, 140) = 165.87$, $p < .001$, $\eta_p^2 = .54$, such that hit rate was lower in low prevalence trials. Pairwise comparisons show that 10% prevalence hit rate (M .40, SEM .018) was significantly lower than 50% prevalence hit rate (M .58, SEM .016), $t(140) = 12.31$, $p < .001$, $d = 2.08$. Pairwise comparisons also show higher correct rejection rate in the 10% prevalence (M .996, SEM 0) than 50% prevalence (M .98, SEM .004) trials, $t(140) = 4.73$, $p < .001$, $d = .80$. Though the difference between correct rejection rate at the two prevalence rates was significant, the effect of prevalence on hit rate was greater.
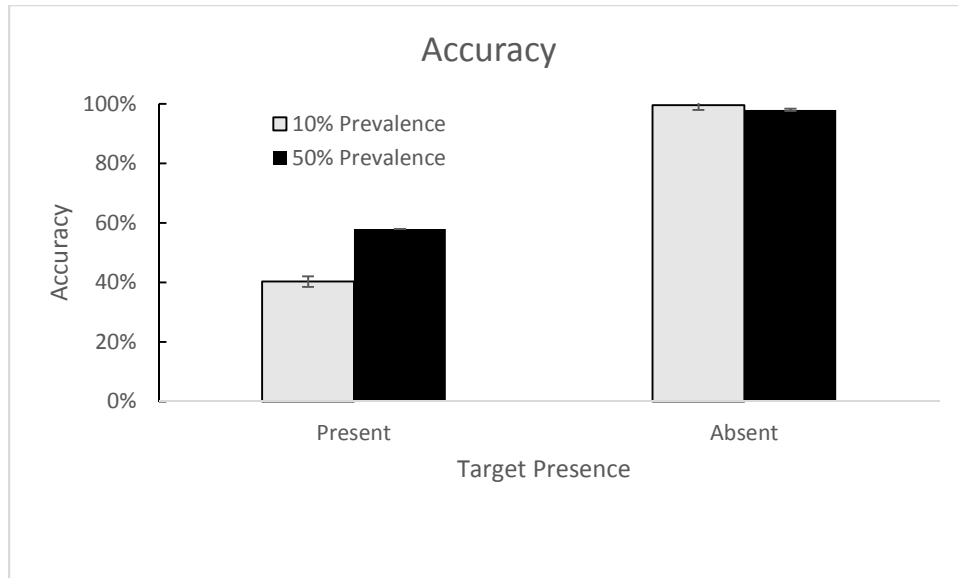
Figure 2. *Accuracy by Target Prevalence and Presence.* Error Bars represent the Standard errors of the means.

Similar to previous results, the reaction time data found that target prevalence had its primary effect on target absent trials (Ishibashi et al., 2012; Rich et al., 2008); as prevalence decreased, reaction times in target absent trials also decreased (see Figure 3). This was confirmed by a significant interaction between target prevalence and target presence $F(1, 140) = 103.23$, p $<.001$, $\eta_p^2 = .42$, driven by a much larger prevalence effect in target absent trials. Pairwise comparisons show that target absent reaction times in the 10% prevalence block (M = 4879.58, SEM = 168.50) were significantly faster than in the 50% prevalence block (M = 6540.76, SEM = 239.18), $t(140) = 7.80$, $p < .001$, d= 1.32. A pairwise comparison of target present trials showed no difference between 10% (M = 3725.84, SEM = 109.11) and 50% (M= 3749.21, SEM = 94.65), $t(140) = -.22$, $p = .83$, d = .037, confirming the interaction was caused by a drop in target absent reaction time as prevalence decreased.
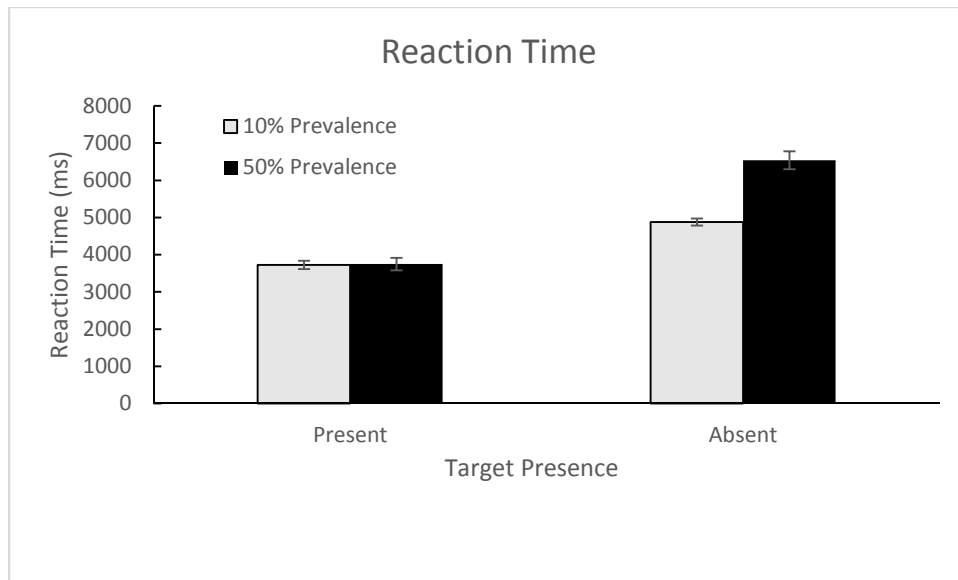
Figure 3. *Reaction time by Target Prevalence and Presence.* Error Bars represent the Standard errors of the means.

2.3.2 Accuracy Regression

Cognitive measures were used as predictors in a multiple regression model to predict low prevalence accuracy. The cognitive predictors that were entered into this model included near transfer high prevalence search performance, K, vigilance, and attentional control. The multiple regression model significantly predicted low prevalence accuracy, $F(4, 136) = 36.29$, $p < .001$, adjusted $R^2 = .502$. High prevalence performance, K, and attentional control all significantly contributed to the model, $ts > 2.2$, $ps < .03$, all betas $> .11$. Vigilance marginally contributed to the model, $t = 1.823$, $p = .07$, beta $= .11$, and was included in the overall model.

After validating the cognitive predictors of accuracy, I enter in the personality factors as measured by the Mini IPIP (Donnellan et al., 2006) into the second stage of a hierarchical linear regression. I separate the entry of personality and cognitive predictors into the regression model to show the benefit of including personality measures in predicting visual search performance.

The new multiple regression model, $F_{(9, 131)} = 17.94$, $p < .001$, adjusted $R^2 = .521$, marginally predicted low prevalence accuracy over and above the cognitive factors regression model, adjusted R^2 change $= .019$, F change $= 2.098$, significant F change $= .07$. Introversion was a marginally significant predictor of rare target detection accuracy, $t = -1.97$, $p = .051$, beta $= -.13$, with those higher rates of introversion predicting better target detection. All other personality measures did not approach significance, all ts<1.43, all ps>.15. Thus, for the final model, I reduced the model to include only the cognitive factors and the introversion personality factor.

A multiple regression model was used to predict low prevalence search accuracy[1] from the factors near transfer high prevalence search performance, K, vigilance, attentional control, and introversion. The overall accuracy regression model predicted low prevalence accuracy, $F_{(5, 135)} = 31.68$, $p < .001$, adjusted $R^2 = .523$. All factors significantly contributed to the model (see Table 1).

---

[1] A regression was also run using accuracy data transformed to A' (Zhang & Mueller, 2005). A similar pattern of results emerged; the same five factors were significant predictors. However, I report only the corrected hit analyses because A' values can be unreliable at the extremes (i.e. when performance is 0 or 100%) and over 90% of the subjects had either a hit rate of 100% or a false alarm rate of 0% in at least one condition.

	T	p value	beta
Near Transfer Search	9.99	<.001	.60
K	2.27	.03	.14
Vigilance	2.05	.04	.13
Attentional Control	2.59	.03	.14
Introversion	2.63	.009	.15

Table 1. *t, p, and beta values for each predictor in the final regression model for low prevalence accuracy.*

2.3.3 Reaction Time Regression

A multiple regression model was used to predict low prevalence reaction time from the factors near transfer high prevalence search performance, K, vigilance, and attentional control. The multiple regression model significantly predicted low prevalence target absent reaction time, $F(4, 136) = 19.97$, $p < .001$, adjusted $R^2 = .351$.

High prevalence performance was a significant predictor of low prevalence target absent reaction time, $t = 7.16$, $p < .001$, beta = .51. K and attentional control were marginally significant predictors of low prevalence reaction time, both $ts > 1.74$, both $ps < .084$, betas > .12. Vigilance was a non-significant predictor, $t = 1.48$, $p = .14$, beta = .10. I enter in the cognitive factors high prevalence performance, K, and attentional control into the model with personality factors.

Personality factors were entered into the second stage of a hierarchical linear regression after the cognitive factors. The new multiple regression model, $F (8, 132) = 10.85$, $p < .001$, adjusted $R^2 = .36$, did not predict low prevalence accuracy over and above the cognitive factors regression model, adjusted R^2 change $= .014$, F change $= 1.62$, significant F change $= .16$.

Of the personality factors, only introversion marginally predicted low prevalence target reaction time, $t = 1.94$, $p = .055$, beta $= .15$, such that introverts have greater reaction time, and thus is the only personality factor included in the final regression model.

A multiple regression model was used to predict low prevalence reaction from the factors high prevalence performance, K, attentional control, and introversion. The hierarchical reaction time regression model predicted low prevalence target absent reaction time, $F (4, 136) = 20.7$, $p < .001$, adjusted $R^2 = .36$, over and above the cognitive model alone, adjusted R^2 change $= .014$, F change $= 4.09$, sig F change $= .045$. High prevalence performance, attentional control, and introversion significantly contributed to the model and K was a marginal predictor (see Table 2).

	T	p value	beta
High Prevalence Performance	7.68	<.001	.53
K	1.89	.06	.13
Attentional Control	2.01	.04	.14
Introversion	2.02	.04	.14

Table 2. *t, p, and beta values for each predictor in the final regression model for low prevalence target absent reaction time.*

2.3.4 False Alarms

The goal of this research was to find those who have high accuracy, which is made up of both a high hit rate and a low false alarm rate. Due to the measure of low prevalence accuracy (hits minus false alarms), it is possible that observers high on the measures high prevalence performance, K, attentional control, vigilance, and introversion simply increased their hits more than their increase in false alarms, which could increase accuracy as a result of a shift in criterion. However, false alarms in cancer screening and security are costly, thus it would be ideal if these measures predicted an increase in hits without a corresponding increase in false alarms, which was the imperfect experimental solution to the low prevalence effect (Wolfe et al., 2007). To test this possibility, I correlated each significant predictor of low prevalence accuracy with false alarms. I found high prevalence performance and K are both significantly correlated with false alarms (see Table 3). Importantly, the direction of that relationship is such that

individuals who have higher scores on these measures make fewer false alarms, suggesting that they have higher sensitivity, rather than a liberal criterion.

	High Prevalence Performance	Attentional Control	K	Vigilance	Introversion
Pearson Correlation	-.321	.149	-.174	.05	-.095
p value	<.001	.082	.039	.555	.261

Table 3. *The correlations and p values of the relationships between significant predictors of low prevalence accuracy and low prevalence false alarms.*

2.3.5 Predictors of High Prevalence Accuracy

In finding predictors of low prevalence search performance, it is possible I did not identify tasks that predict who is uniquely suited for performing critical low prevalence search tasks, but perhaps found those who are generally good at search. If this alternative interpretation of the results is true, I should find that each variable that accounts for a significant proportion of low prevalence performance should also account for a significant proportion of high prevalence performance.

To test for this possibility, I performed a linear regression predicting high prevalence search accuracy using low prevalence search accuracy, K, Vigilance, Attentional Control, and

28

introversion.[2] I found that of these predictors, only low prevalence search performance, t = 9.99, p < .001, beta = .70, accounted for a significant portion of high prevalence performance, while introversion was a nearly significant predictor, t = 1.9, p = .06, beta = .12. All other predictors were not significant, ts < 1.3, ps > .19, betas <.08. From these data I conclude that these predictors of low prevalence performance are uniquely suited to account for low prevalence search performance, not visual search in general.

 2.4 Discussion

Using an individual differences approach, I found that better accuracy in a low (10%) target prevalence search task was predicted by higher accuracy on a near transfer high (50%) prevalence search task, higher WMC, more vigilance, and more rapid attentional shifting. When I then added personality measures to the regression model, only the introversion factor increased the model's predictive ability, with more introverted participants tending to perform better on the low prevalence search task.   A final regression model with these five factors accounted for more than half of the variance in rare target search performance.

Each of these predictors were chosen because of their relation to visual search. WMC has been shown to predict quitting thresholds and hit rate in low target prevalence (Schwark et al., 2013). Vigilance tasks and low target prevalence search share the need to maintain attention while trying to detect the rare target. Similarly, research has claimed that introverts maintain a higher level of baseline arousal (Eysenck, 1967), which allows them to perform monotonous tasks, such as a low prevalence search task, at a high level for prolonged periods. Visual search

---

[2] I also performed this regression leaving low prevalence search performance out of the model. I find that only vigilance, t = 2.15, p = .03, beta = .18, accounted for a significant portion of the variance. All other predictors were not significant, ts < 1.74, ps > .08, betas < .15.

and the measure of Attentional Control both require the shifting of attention between stimuli. Though I identified some valid predictors of low prevalence search here, later efforts should seek to expand on this battery with additional predictor tasks.

Importantly, both increased WMC and higher accuracy in the near transfer high prevalence search performance predicted fewer false alarms in the low prevalence search task, and the remaining predictor variables were not significant predictors of false alarms. This finding that the predictors that were associated with better low prevalence target detection were also associated with fewer false alarms suggests that these factors predict an increase in sensitivity, rather than only a shift in the decision criterion for responding "target present".

In addition, I found that each of the factors that predicted better target detection, with the exception of vigilance, also predicted *slower* low prevalence target absent reaction times. This pattern of results hints at a potential underlying mechanism that mediates the relationship between the predictors and rare target search performance, namely that the predictors may be systematically related to an individual's quitting threshold. Theoretical models designed to explain how target absent responses occur in visual search propose that, during the course of a trial, evidence accumulates toward a trial quitting threshold (Wolfe & Van Wert, 2010). If this accumulation of evidence reaches the quitting threshold prior to the identification of a target, a target absent response is made (Wolfe & Van Wert, 2010). In theory, as targets become rare the quitting threshold decreases, resulting in the need to examine less of the display before executing a target absent response, and leading to increased miss errors (Hout et al., 2015). In short, a lower quitting threshold is associated with both lower target absent reaction times and worse target detection accuracy. The fact that the predictors are associated with both of these outcomes provides a hint that the predictors may be related to an individual's low prevalence quitting

threshold, providing potential insight into the mechanism by which these predictors mediate low target search performance.

It is worth noting that the most powerful predictor of low prevalence search accuracy was near transfer high prevalence search performance. While this may not be surprising as these tasks represent near transfer, it has important practical implications. Real-world searches can have target prevalence rates as low as .3%, thus using a realistic target prevalence rate in a screener task would be time consuming and expensive as potential employees would have to complete several thousand trials to gather reliable data about their performance. Given the strong association between high and low target prevalence accuracy, using one's performance on a near transfer high prevalence search task as a proxy for their likely ability to detect low prevalence targets may be a much more economical approach. However, it is worth pointing out that the other factors, leaving out high prevalence search performance, constitute a significant model on their own. When removing the high prevalence search factor from the regression model, the cognitive factors are all still significant predictors (all ts > 2.28, all ps < .025), and introversion is a marginally significant predictor (t= 1.83, p = .07) of low prevalence search accuracy.

The current work demonstrates that approaching the problems associated with low target prevalence from an individual differences perspective has promise. These predictor tasks may have real-world significance in that they can be used to find those who are more suited for tasks where the goal is to find a rare target, such as in airport security checks. Each task is fast and easy to administer, maximizing its potential to be used in the workplace. The finding that these factors all predict higher accuracy without a significant increase in false alarms shows that an individual differences approach may be more suitable to increase accuracy in these situations

than the currently known experimental manipulations (Kunar, Rich, & Wolfe, 2010; Wolfe et al., 2007).

We have demonstrated the viability of an individual differences approach to improving visual search performance. I found that I can identify those who are likely to have a relatively high detection rate of rare targets without a corresponding increase in false alarms by measuring near transfer high prevalence search performance, WMC, vigilance, attentional control, and introversion. While this was an important first step in this research area, there are several limitations of the study that are critical to address. First, as this is a relatively novel and exploratory investigation, it is important to verify these findings with a new sample of observers. Second, the stimuli used here do not resemble those found in a realistic baggage screening task, which could limit the generalizability of these findings. Additionally, though this study shows that the individual differences approach has potential as an applied method, without eye-tracking measures, I can only speculate as to the mechanisms through which these measures relate to performance. In the following experiment, I address these weaknesses.

CHAPTER 3. EXPERIMENT 2

3.1 Introduction

Experiment 1 (published as Peltier and Becker (2017a)) was an exploratory, correlational study predicting low prevalence visual search performance, thus it is critical to attempt to replicate those findings. The first stage of the current study is to establish the basic LPE in a search task (an interaction between target prevalence and target presence in both the reaction time and accuracy data, such that as prevalence decreases, both hit rates and target absent reaction times decrease), then measure performance using the same tasks as Experiment 1 to predict search performance and verify the previous predictors. I will also use the beta values from the predictor tasks established in Experiment 1 and performance in the new sample to test how much of the variance in performance I can account for in the current low prevalence search task.

Beyond replicating my previous work, I aim to increase the external validity of this line of research by using more realistic stimuli. To increase external validity, the critical low prevalence search task stimuli will consist of colored, realistic search items, including guns, bombs, headphones, shoes, etc., instead of being limited to artificial Ts and Ls. In addition, the search task will have four possible targets, which is more similar to a real search environment where the observer has to search for multiple targets.

To help understand the mechanisms through which these individual differences relate to visual search performance, I will eye-track a sub-sample of observers as they perform the critical low prevalence search task to measure quitting thresholds (number of items inspected in a target absent search), identification errors, selection errors, target and distractor dwell times, and

refixations. With these eye-tracking measures, I will be the first study to link multiple individual difference measures (see Peltier and Becker, 2017a for an eye-tracking study using WMC as the only predictor) to the different components of search performance. I will then be able to answer which measures increase accuracy by reducing selection errors or identification errors, which measures reduce reaction time through faster item identification, and which measures influence reaction time through different levels of search efficiency (re-inspection rate).

Finally, I also aim to increase the predictive power of the individual differences approach to improving low prevalence visual search performance by adding new predictor tasks in addition to those established in Experiment 1. These include a far transfer high prevalence search task, Raven's Advanced Progressive Matrices, and a Task Unrelated Thoughts probe.

3.2 New Individual Differences Measures

3.2.1 Near and Far Transfer Visual Search Performance

Finding that high prevalence search performance is a good predictor of low prevalence search performance provides an efficient method for predicting who would be good at low-prevalence search tasks. In Experiment 1, the strongest predictor of performance in a low prevalence search task with Ts and Ls as search stimuli was a high prevalence search task with identical stimuli. This was a near transfer task, but neither the predictor (high prevalence search), nor criterion (low prevalence search) tasks accurately represented a real-world search due to the stimuli used. Here, I will test if a high prevalence search with far transfer stimuli (Ts and Ls) predicts a unique portion of variance in a more real-world search task. If far transfer high prevalence search with Ts and Ls accurately predicts performance with more real-world stimuli, it could show that much of the research done in the lab with unrealistic stimuli can generalize to

real-world situations. Alternatively, it is possible that a predictor search task must use similar stimuli to the criterion search task to be a significant predictor of performance, thus the far transfer high prevalence search task would not predict the baggage screening search task's performance. In this case, previous research investigating methods of improving visual search performance using unrealistic stimuli may be limited in their external validity.

In addition to using a far transfer high prevalence predictor task (Ts and Ls), I will also continue using near transfer search as a predictor, where the only difference between it and the critical task is the target prevalence, where the near transfer predictor task has a target prevalence rate of 50% and the critical task has a target prevalence of 10%. The use of two predictor search tasks (far transfer and near transfer) allows two critical analyses. First, I can attempt to replicate Experiment 1's findings using the near transfer search task in a regression predicting visual search performance for accuracy and reaction time. Second, I will be able to compare the predictive power of near transfer and far transfer search tasks by entering both into regressions predicting the critical variables. I expect that due to the near transfer nature of the high prevalence baggage screening search task, it will be a stronger predictor, possibly absorbing all the variance explained by the far transfer search task once they are entered in the same regression model. However, based on pilot data, I found that accuracy was higher and reaction time was faster in the baggage screening search task than the Ts and Ls search task, indicating that the near transfer search task is easier. It is then possible that the more difficult task will still have predictive value as it may relate to effort or other individual differences related to performance.

3.2.2 Task Unrelated Thoughts

Vigilance tasks requiring sustained attention over long periods of time are characterized by low arousal because of low task demands, which can lead to withdrawal from the task and task unrelated thoughts (TUTs) (Manly, Robertson, Galloway, & Hawkins, 1999). Experiment 1 found that a continuous performance vigilance task (Covey et al., 2013) was a valid predictor of low prevalence visual search performance. Vigilance predicted an increase in accuracy without a corresponding increase in reaction time, suggesting that high vigilance abilities may result in fewer identification errors, as a reduction in selection errors would have likely been associated with increased reaction time through greater quitting thresholds. This could be due to those with higher vigilance abilities having fewer task-unrelated thoughts (TUTs), which could occur at a critical moment when inspecting a target, resulting in the identification error. Both TUTs and identification errors were previously unmeasured, but I include a TUT probe (Levinson, Smallwood, & Davidson, 2012) in the high and low prevalence baggage screening search tasks, while identification errors will be measured using the eye-tracker.

TUTs occur when observers withdraw attention from the primary task and have self-generated thoughts unrelated to ongoing activities (Allen et al., 2013). TUTs have been shown to decrease performance in vigilance tasks (Smallwood et al., 2004) and this effect should theoretically carry over to the highly similar low prevalence search tasks. TUTs about a desired activity could prompt premature trial terminations, and thus selection errors, in order to move on to the other activity. TUTs could also automate behavior, such that an observer will continue searching a display, but when a target is fixated it is not perceived by the observer. This has been shown to occur in vigilance tasks (See, Warm, Dember, & Howe, 1997), and would show a

criterion shift. To test these predictions, a TUT probe can be used to monitor TUT frequency during task performance, then correlate the number of TUTs with accuracy, reaction time, and selection and identification errors.

3.2.3 Fluid Intelligence

Intelligence is the broad mental capacity that influences performance in cognitive tasks (Jensen, 1998). Processing speed, WMC, short term memory, and executive attention are commonly noted as components of general intelligence (Ackerman, Beier, & Boyle, 2005; Engle & Kane, 2003; Kane, Conway, Hambrick, & Engle, 2007). As intelligence is comprised of so many important and widely influential cognitive abilities, it is unsurprising that it is the best predictor of job performance (Schmidt & Hunter, 1993, 1998). Given that measures of WMC have already been shown to be valid predictors of visual search performance, it is easy to predict that visual search performance can be predicted by measures of intelligence. However, unlike its components, measures of intelligence do not make specific predictions regarding search performance; quitting thresholds, search efficiency, recognition time, identification errors, and selection errors could all correlate with g. Though this potential predictive power could be valuable in the work place, previous court rulings may make it difficult to use.

While intelligence may absorb a lot of the predictive power of other variables, it is important to note that from an applied perspective, it is rare to use intelligence tests to screen employees because it has been ruled as a discriminatory practice (Supreme Court ruling, Griggs vs Duke Power). However, it is unclear if using a non-language based intelligence test is considered discriminatory. Thus, the correlations between intelligence and visual search performance may be theoretically interesting, but it is possible that they are not useable in real-world applications where employers may not be able to use a measure of intelligence as a tool to

determine employee selection. Additionally, any variables that predict visual search performance over and above intelligence should be considered highly valuable in that they predict some unique ability related to visual search performance. To measure intelligence, we use Raven's Advanced Progressive Matrices, a measure of fluid intelligence (Raven, 1998).

3.3 Methods

3.3.1 Participants

One hundred fifty-eight undergraduates (sample size from Experiment 1) from Michigan State University's human subjects pool gave consent to participate in the study for course credit. 100 subjects were female. Observers were split between eye-tracked and non-eye-tracked conditions (79 for each condition). Subjects were between the ages of 18 to 24 with normal or corrected to normal vision. Thirteen (seven from the eye-tracking group) subjects were excluded from further analysis for failing to complete all tasks, leaving 145 subjects for behavioral analyses.

3.3.2 Low Prevalence Search Task

The task was to search for a randomly rotated target (0°, 90°, 180°, or 270°) among 16 items and respond target present or absent by using keys marked "P" for present or "A" for absent. There were four possible targets, a green gun, a blue gun, an orange crossbow, and a blue and orange bomb. All targets appeared during the instructions, so observers knew the exact target templates. A target was chosen from each color (and a conjunction) used in the search task so that observers could not use color as the sole guide of attention to find the target in parallel, as color is often the strongest target cue (Williams, 1966). Items were presented in a 4x4 grid with random jitter added to break up the orderly layout. Each item was matched in length, though not

stretched to match width, in order to maintain realistic size ratios. Items measured approximately 2.5 degrees of visual angle.

Preceding the critical trials were 40 practice trials to set an appropriate criterion (Ishibashi et al., 2012; Wolfe & Van Wert, 2010) for the upcoming low prevalence search trials. Each target appeared once in a random orientation during the practice trials. These practice trials were not considered in the data analysis. In the critical trial block, each target appeared twice at each orientation, giving a total of 32 target present trials. Targets were present on 10% of trials, giving a total of 320 critical search trials. Only a single target could appear on a given trial. Distractors were randomly chosen with replacement from 20 possible items, appearing equally often from each possible rotation (0°, 90°, 180°, or 270°).

3.3.3 Predictor Tasks

In addition to the predictor tasks below, all observers completed each predictor task used in Experiment 1, using the same methods.

3.3.3.1 High Prevalence Near Transfer Search

This portion of the task was identical to that of the low prevalence search task above, with one exception. Target prevalence was 50%, so there were only 64 total trials (32 target present). This block of trials was similarly preceded by 40 practice trials to set an appropriate criterion.

3.3.3.2 Raven's Advanced Progressive Matrices (Fluid Intelligence)

To assess fluid intelligence, observers completed 18 of the items in Raven's Advanced Progressive Matrices (Raven, Raven, & Court, 1998). Each trial was made up of a 3 X 3 grid

where all but the lower left right place was filled by a series of images that made a pattern. The observer's task was to choose the item that fit the pattern from 8 options. Each trial was untimed, but the task terminated after 10 minutes and the score was then calculated as the number correct out of 18, such that unanswered questions were automatically incorrect.

3.3.3.3 Task Unrelated Thoughts Probe

To measure the frequency of TUTs, observers were given a TUT probe using the procedure from Levinson et al. (2012). A TUT probe was given every 40 trials during the near transfer visual search task, during both high and low prevalence blocks and including practice trials. TUT probes required a single button keyboard press in response to the question, "What were you thinking just now? Press D for a task related thought, Press J for a task unrelated thought." The TUT score was the percentage of probes on which a participant indicated a TUT.

3.3.4 Procedure

All observers first completed the baggage screening search task trials. Observers were randomly assigned to complete the low or high prevalence block first. The eye-tracked observers had a brief calibration procedure using the EyeLink's 9-point calibration routine. Eye movements were recorded at 1000 Hz. All fixations less than 50ms in duration were discarded from analysis (Rich et al., 2008). A chin rest was used to stabilize the observer's head position. After every ten trials, a recalibration was performed on a single fixation point to maintain eye-tracking accuracy. Regardless of being eye-tracked or not, all observers completed this search task on a 20.5 (44cm x 27cm) in. monitor. All participants completed a TUT probe during the search task using a procedure described in the methods section.

Following the visual search trials, observers were moved to a noise and light isolated room with a computer monitor that measured 24 (53cm x 30cm) in. All observers completed the following tasks in the same order: personality inventory, Raven's Progressive Matrices, Change Detection, Posner Cuing, far transfer visual search, and AX Continuous Performance Task. Each task, including the eye-tracked baggage screening visual search task, was programmed using E-Prime, version 2.0. Between each task, observers had the opportunity to take a break. Prior to starting a new task, they would be instructed by an experimenter on how to complete the task.

3.3.5 Data Preparation

All tasks were completed by 145 participants. Data from 11 (three from the eye-tracking group) participants were eliminated from further analyses based on a corrected hit rate (hits minus false alarms) below three standard deviations from the mean in the critical low prevalence search task, leaving a final sample size of 134, with 69 eye-tracked.

In each visual search task (baggage screening low prevalence, near transfer high prevalence, far transfer high prevalence) trials with reaction times three standard deviations above the mean were discarded. This resulted in 1.65% of trials being discarded from the critical low prevalence task, 3.01% of trials being discarded from the high prevalence near transfer task, and .7% being discarded from the high prevalence far transfer task. In all visual search trials, incorrect trials were discarded from reaction time analyses.

For all regressions, I followed the same procedures as Experiment 1 to check for any violations of the assumption of a regression. I found no violations of linearity, homoscedasticity, normality, or independence of residuals, or multicollinearity.

3.4 Results

3.4.1 Reliability

All measures (except for the personality questionnaire) were tested for reliability. I used the
even-odd split method to split the data sets to analyze using Cronbach's alpha. An alpha value of
over .7 is considered reliable (Tavakol & Dennick, 2011). Each measure exceeds this threshold
for reliability.

	Odd	Even	Cronbach's Alpha
Low Prevalence Search	.68 (.20)	.68 (.20)	.78
Near Transfer Search	.82 (.16)	.82 (.15)	.86
Far Transfer Search	.55 (.21)	.54 (.21)	.81
Vigilance	.92 (.16)	.91 (.17)	.92
Change Detection	.69 (.09)	.68 (.08)	.73
Posner	323.23 (140.5)	324.93 (139.3)	.97
Raven's	.56 (.24)	.55 (.24)	.82

Table 4. *Reliability for tasks in Experiment 2.* Presented as means (and standard deviations) as a
function of an odd-even split to calculate Cronbach's alpha. Dependent variables for each
measure: low prevalence search, hit rate; near transfer search, hit rate; far transfer search, hit
rate; vigilance, hit rate; Change Detection, accuracy; Posner, reaction time collapsed across valid
and invalid trials; Raven's, accuracy for items that had a response.

3.4.2 Low Prevalence Effect: Reaction Time and Accuracy

To confirm that the prevalence manipulation resulted in the standard low prevalence effect, I conducted two 2 (target present/target absent) X 2 (low/high prevalence block) repeated-measures ANOVAs, one on the accuracy data and one on the reaction time data for near transfer search tasks. The ANOVAs were followed by paired sample t-tests to verify the presence of the low-prevalence effect.

The results show standard prevalence effects on accuracy (see Figure 4); as target prevalence decreased, hit rate decreased. This was confirmed by a significant interaction between target prevalence and target presence, $F(1, 133) = 94.80$, $p < .001$, $\eta_p^2 = .42$, showing the hit rate was lower in low prevalence trials. Pairwise comparisons show that hit rate was significantly lower in the 10% (M .68, SEM .015) than 50% (M .82, SEM .013) prevalence condition, $t(133) = 10.90$, $p < .001$, $d = 1.89$. Pairwise comparisons show no difference in correct rejection rates between 10% prevalence (M .99, SEM .004) and 50% prevalence (M .99, SEM .003) trials, $t(133) = .39$, $p = .70$, $d = .07$.
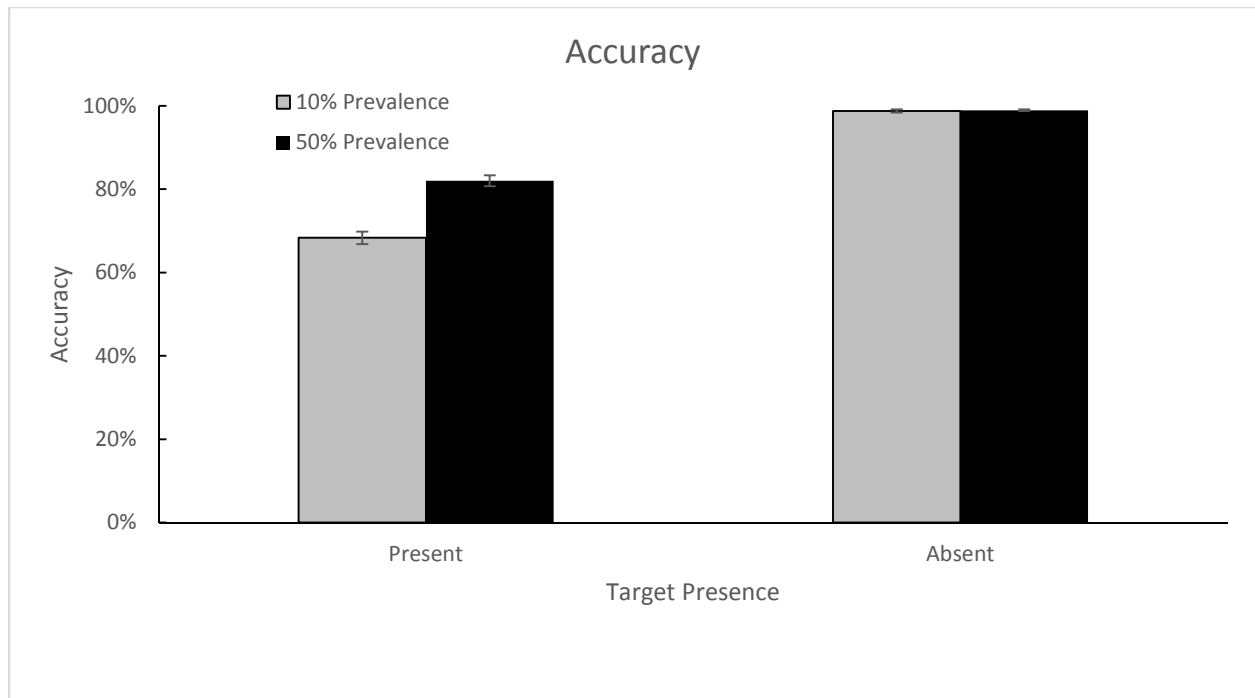
Figure 4. *Accuracy by Target Prevalence and Presence.* Error Bars represent the Standard errors of the means.

The reaction time data also shows the standard prevalence effect; as target prevalence decreased, reaction time in target absent trials decreased (see Figure 5). This was shown by a significant interaction between target prevalence and target presence $F(1, 133) = 159.11$, p $<.001$, $\eta_p^2 = .54$. Pairwise comparisons show that target absent reaction times in the 10% prevalence block (M = 2569.02, SEM = 82.35) were significantly faster than in the 50% prevalence block (M = 3151.04, SEM = 91.16), $t(133) = 7.89$, p < .001, d= 1.36. A pairwise comparison of target present trials showed no difference between 10% (M = 1722.46, SEM = 40.03) and 50% (M= 1666.34, SEM = 38.11), $t(133) = 1.92$, p = .24, d = .33, confirming the interaction was caused by the change in target absent reaction times.
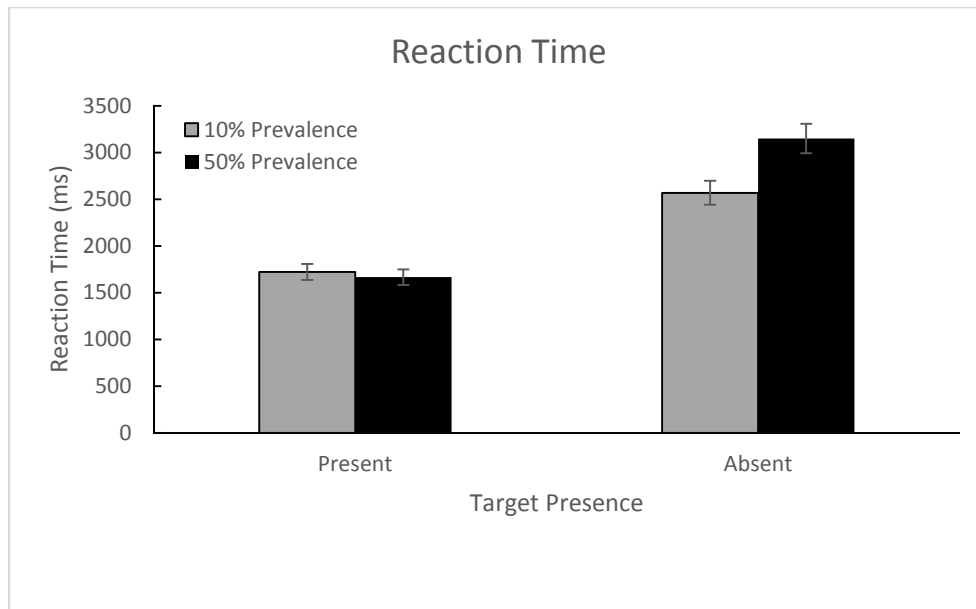
Figure 5. *Reaction time by Target Prevalence and Presence.* Error Bars represent the Standard errors of the means.

3.4.3 Accuracy Regression

In this section, I will use all predictors from Experiment 2 in regression models to predict low prevalence visual search accuracy (hits minus false alarms) and target absent reaction time. The predictors are the near transfer search accuracy (high prevalence baggage screening stimuli), far transfer search task (high prevalence Ts and Ls stimuli), WMC, attentional control, vigilance, personality inventory, fluid intelligence, and TUTs. Each assumption of a regression model was tested in the same manner as Experiment 1 and there were no violations.

In Experiment 1, I separated the entry of cognitive and non-cognitive predictors into the regression models to make it clear that adding a non-cognitive factor results in a significant increase in variance explained. Here, in the interest in brevity and having already demonstrated the value of using non-cognitive predictors, I use a single stepwise regression for each critical

measure in Experiment 2. This gives a single regression model for each measure that includes only the significant predictors.

I entered each predictor into a stepwise regression. The regression model significantly predicted low prevalence accuracy, $F(5, 128) = 30.76$, $p < .001$, adjusted $R^2 = .53$.[3] Near transfer high prevalence performance, far transfer high prevalence performance, WMC, and fluid intelligence all significantly contributed to the model, $ts > 2$, $ps < .05$, all betas $> .12$, such that higher performance in these tasks predicted higher accuracy. Introversion was a marginal contributor to the model, $t = 1.9$, $p = .05$, beta $= .12$, such that introverts had higher accuracy than extraverts, and is included in the model (see Table 5 to see the model)

	T	p value	beta
Near Transfer Search	8.82	<.001	.72
Far Transfer Search	2.20	.03	.13
K	2.63	.01	.17
Fluid Intelligence	2.01	.046	.13
Introversion	1.94	.05	.12

Table 5. *t, p, and beta values for each predictor in the final regression model for low prevalence accuracy.*

---

[3]The dependent variable for accuracy was corrected hits (hits minus false alarms). Corrected hit rate was lower in the 10% prevalence (M .67, SEM .017) than 50% prevalence (M .81, SEM .013) trials, $t(133) = 11.03$, $p < .001$, $d = 1.91$.

3.4.4 Reaction Time Regression

I then entered all predictors into a stepwise regression to predict low prevalence target absent reaction times. The multiple regression model significantly predicted low prevalence target absent reaction time, $F_{(2, 131)} = 24.35$, $p < .001$, adjusted $R^2 = .26$ (see Table 6 to see the full model). Near transfer search accuracy and far transfer search accuracy were significant predictors, $ts > 3.1$, $ps < .01$, betas $> .24$, such that as performance in these tasks increased, target absent reaction time increased.

	T	p value	beta
Near Transfer Search	4.93	<.001	.39
Far Transfer Search	3.16	.002	.25

Table 6. *t, p, and beta values for each predictor in the final regression model for low prevalence reaction time.*

3.4.5 Replicating Previous Individual Differences Results

Experiment 1 showed I was able to account for 52% of the variance in accuracy and 36% of the variance in reaction time in a Ts and Ls low prevalence search task using the predictors vigilance, WMC, near transfer search performance, attentional control, and introversion. Here, I use the same predictors, but to predict performance in the low prevalence baggage screening visual search task to test how well those predictors generalize and the degree to which the

47

Experiment 1 results replicate. This analysis does not use each predictor used in Experiment 2; I am not using the new predictors of far transfer search, fluid intelligence, or TUTs.

To predict accuracy, I entered each of the previously significant predictors, vigilance, WMC, near transfer search accuracy, attentional control, and introversion into a regression. The regression model was a significant predictor of low prevalence accuracy, $F (5, 128) = 26.34$, $P < .001$, adjusted $R^2 = .49$. WMC and near transfer search accuracy significantly contributed to the model, $ts > 3.5$, $Ps < .001$, betas $> .22$, such that those with higher WMC and near transfer search accuracy were predicted to have higher accuracy. Introversion marginally contributed to the model, $t = 1.9$, $p = .05$, beta $= .12$ (see Table 7 for the model. The table contains non-significant predictors for an easier comparison to Experiment 1 results). Attentional control and vigilance were non-significant predictors, $ps > .44$.

	T	p value	beta
Near Transfer Search	9.76	<.001	.64
K	3.51	.001	.23
Vigilance	.77	.44	.05
Attentional Control	.75	.45	.05
Introversion	1.93	.05	.12

Table 7. *t, p, and beta values for each predictor in the replicated regression model for low prevalence accuracy.*

The second variable to predict is low prevalence search target absent reaction time. The model accounts for a significant proportion of the variance in low prevalence visual search reaction time, $F(4, 129) = 4.45$, $p < .001$, adjusted $R^2 = .21$. Only the high prevalence near transfer search task was a significant predictor, $t = 5.65$, $p < .001$, beta $= .45$, such that as performance in the near transfer search task increased, target absent reaction time increased. All others failed to approach significance, $ps > .25$. (See Table 8 for the model. The Table contains non-significant predictors for an easier comparison to Experiment 1 results).

	T	p value	beta
Near Transfer Search	5.74	<.001	.47
K	1.12	.26	.26
Attentional Control	1.01	.28	.07
Introversion	.91	.37	.07

Table 8. *t, p, and beta values for each predictor in the replicated regression model for low prevalence reaction time.*

3.4.6 Betas from Experiment 1

Another critical question I can investigate by using the data set from Experiment 1 is how much variance in the current low prevalence search task I can account for by using the beta weights established from the Experiment 1 predictor tasks. If the beta weights derived from Experiment 1 can be used to predict performance in Experiment 2, we would have strong

evidence that the predictors are robust and meaningful. To examine this issue, I used only the predictors from Experiment 1 (i.e., did not include far transfer search, fluid intelligence, or TUTs) in the regression. To perform this calculation, I used the beta weights from Experiment 1's significant predictors (near transfer search, attentional control, vigilance, WMC, and introversion) and each observer's score on the corresponding measures in the current experiment to predict performance on the critical low prevalence search task. The results (plotted in Figure 6) show that the regression with the beta weights from Experiment 1 could successfully be applied to Experiment 2; it accounted for 42% of the variance in Experiment 2's low prevalence search performance (adjusted R^2 = .423, r = .66). Though this 42% of explained variance is lower than the 52% of the variance in performance accounted for in Experiment 1, this is still a highly significant proportion of variance explained.
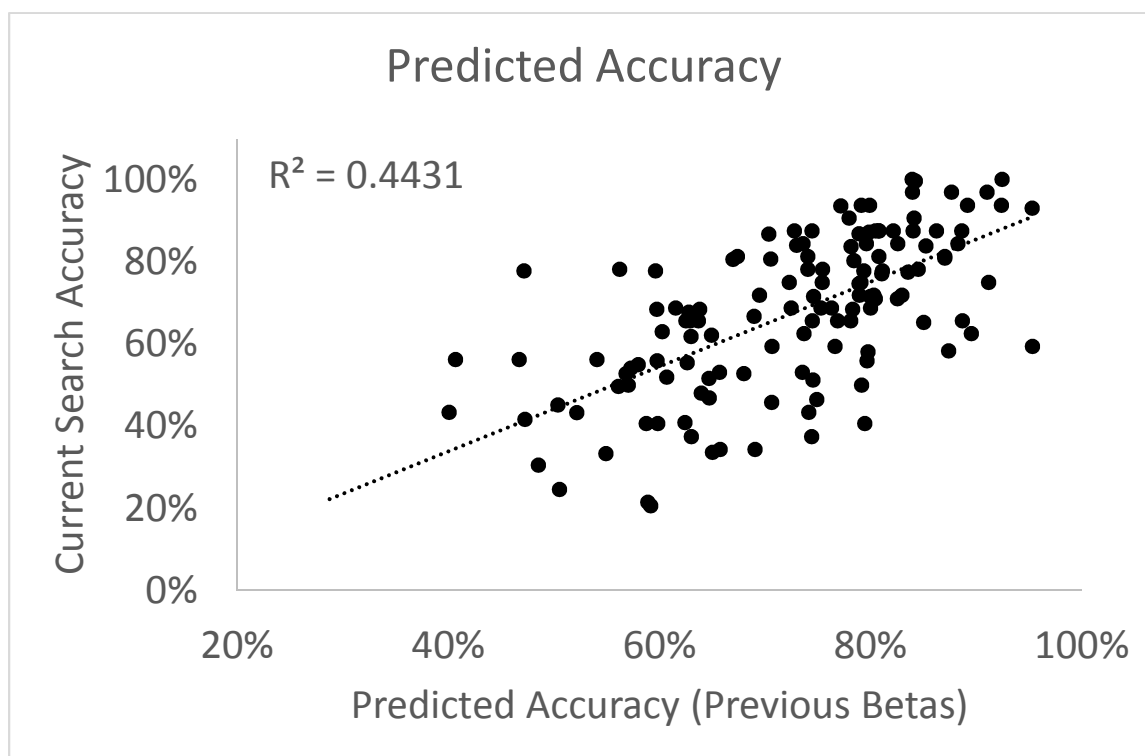
Figure 6. *Low Prevalence Visual Search Accuracy as a Function of Predicted Accuracy from Experiment 1 betas.*

3.4.7 Eye Tracking Measures

This experiment is novel in that it uses several individual differences to predict eye-tracking measures during a visual search task. The measures of interest are quitting thresholds (the number of items inspected in a target absent search), selection error rate (the percentage of misses caused by never fixating the target), identification error rate (the percentage of misses caused by failure to recognize a fixated target), re-inspection rate (the percentage of fixations that are to re-inspect an item), target decision time (the cumulative dwell time on an identified target), and distractor dwell time (the average cumulative dwell time on each distractor). I only use the significant predictors of accuracy in these regressions so that I can establish how these predictors are associated with increased accuracy. Each significant predictor of accuracy was entered into each stepwise regression to predict each eye-tracking measure.

The regression significantly predicted low prevalence quitting thresholds, $F(1, 67) = 5.64$, $p = .02$, adjusted $R^2 = .08$. High prevalence far transfer search was the only significant predictor, $t = 2.38$, $p = .02$, beta $= .12$.

The regression significantly predicted low prevalence selection error rate, $F(3, 65) = 13.65$, $p < .001$, adjusted $R^2 = .36$. WMC, $t = -2.18$, $p = .03$, beta $= -.22$, fluid intelligence, $t = -3.61$, $p = .001$, beta $= -.37$, and high prevalence near transfer search, $t = -3.07$, $p = .003$, beta $= -.31$, were significant predictors of selection error rate. Better performance for each significant predictor was associated with fewer selection errors.

The regression significantly predicted low prevalence identification error rate, $F(2, 66) = 14.86$, $p < .001$, adjusted $R^2 = .29$. Better performance in both WMC, $t = -2.79$, $p = .007$, beta $= -$

.29, and high prevalence near transfer visual search, t= -4.24, p < .001, beta = -.44, predicted

fewer identification errors.

The regression significantly predicted low prevalence reinspection rate, $F (1, 67) = 9.52$,

$p = .003$, adjusted $R^2 = .11$. High prevalence far transfer search was the only significant

predictor, $t = 3.65$, $p = .003$, $b = .35$.

The regression significantly predicted low prevalence target decision time, $F (1, 67) =$

4.06, $p = .04$, adjusted $R^2 = .04$. Higher fluid intelligence predicted faster target decision times, t

$= -2.02$, $p = .04$, beta $= -.24$.

The regression significantly predicted low prevalence distractor decision time, $F (1, 67) =$

12.18, $p = .001$, adjusted $R^2 = .14$. Better performance in the high prevalence far transfer search

task predicted slower distractor decision times, $t = 3.49$, $p = .001$, beta $= .39$

3.4.8 False alarms

My early criticism of the experimental approach to increasing low prevalence visual

search performance was that it could sometimes increases hits, but did so by causing a liberal

shift in criterion that led to a corresponding increase in false alarms. I also made the claim that an

individual differences approach to improving performance is more effective. I have shown that I

can predict low prevalence search performance using a battery of tasks, but to back up my claim

that the individual differences approach is superior, I need to rule out that the predictors of

increased accuracy do not also predict an increase in false alarms, just as I did in Experiment 1.

To do so, I correlated each significant predictor from the accuracy regression with low

prevalence false alarms. The results show the high prevalence near transfer and far transfer

searches, and WMC are all significantly negatively correlated with false alarms (all rs < -.24, all

ps < .005). Fluid intelligence is trending toward a significant negative correlation ($r = -.13$, $p = .15$) and introversion has a non-significant correlation of $r = -.05$ (see Table 9 for each correlation). Hence, I replicate my finding from Experiment 1 and stand by the claim that the individual differences approach is superior to the experimental approach of improving low prevalence visual search performance in that I can predict observers who will be superior at detecting rare targets without increasing false alarm rates.

	Near Transfer Search	Far Transfer Search	K	Fluid Intelligence	Introversion
Pearson Correlation	-.41	-.25	-.26	-.13	-.05
p value	<.001	.003	.004	.15	.52

Table 9. *The correlations and p values of the relationships between significant predictors of low prevalence accuracy and low prevalence false alarms.*

3.4.9 Predictors of High Prevalence Accuracy

To test for the possibility that the predictors of low prevalence search performance are general predictors of search, rather than unique to predicting low prevalence search, I performed a linear regression predicting high prevalence search performance using low prevalence search,

far transfer high prevalence search, WMC, fluid intelligence, and introversion.[4] I found that of

these predictors, only low prevalence search performance, t = 10.32, p < .001, beta = .67,

accounted for a significant portion of high prevalence performance. All other predictors were not

significant, ts < 1.4, ps > .19, betas <.09. From these data, I conclude that the predictors of low

prevalence performance are uniquely suited to account for low prevalence search performance,

not visual search in general.

3.5 Discussion

In Experiment 2, the goal was to replicate and extend the results of Experiment 1 by

using a more real-world representative search task, including three new predictors (far transfer

visual search, fluid intelligence, TUTs), and eye-tracking a sub-sample of observers. Using the

new predictors, high prevalence visual search, fluid intelligence, and TUTs, combined with the

replicated predictors, I conducted new regressions for accuracy and reaction time. I found that

TUTs did not predict accuracy or reaction time, nor was there a strong correlation between TUTs

and low prevalence search performance, thus I do not believe TUTs to be a candidate for future

individual differences research in relation to visual search. I did find, however, that fluid

intelligence and both near transfer and far transfer visual search were significant predictors of

accuracy. As fluid intelligence is highly correlated with WMC (a replicated predictor of visual

search performance (Peltier & Becker, 2016b; Schwark et al., 2013)) processing speed, and job

performance (Schmidt & Hunter, 1993), it is unsurprising that it is a significant predictor of low

prevalence visual search accuracy.

---

[4] I also performed this regression leaving low prevalence search performance out of the model. I find that only high prevalence far transfer search, t = 3.75, p < .001, beta = .31, accounted for a significant portion of the variance. All other predictors were not significant, ts < 1.7, ps > .1, betas < .15.

High prevalence near transfer and far transfer search were both significant predictors of low prevalence visual search accuracy. It is unsurprising that the near transfer search task was the strongest predictor of performance, but somewhat surprising that far transfer search also significantly contributed to the model after the variance explained by the near transfer search was accounted for. My hypothesis is that the far transfer search task, due to its greater difficulty (measured by increased reaction times and decreased accuracy) is likely predictive of effort, such that those who tried harder and performed better in the far transfer task were more likely to invest effort in the easier near transfer search task.

Near transfer and far transfer search were also both significant predictors of low prevalence visual search reaction time. Higher performance in both tasks predicted higher reaction times. As the primary contributor to errors is a decreased quitting threshold and selection errors (Peltier & Becker, 2016a), it then follows that those who maintained high quitting thresholds and accuracy in a near transfer high prevalence search task would also maintain high quitting thresholds and accuracy in a low prevalence search task.

I also sought to replicate the findings from Experiment 1 using only the previous predictors (near transfer search performance, WMC, vigilance, attentional control, introversion) in a regression to predict performance in the new search task. I replicated the finding that near transfer high prevalence search performance, WMC, and introversion were valid predictors, but found that vigilance and attentional control were not.

It is possible that vigilance was no longer a significant predictor due to the difference in the visual search tasks' difficulties and durations. Hit rate in the low prevalence Ts and Ls search from Experiment 1 was 40%, compared to 68% in Experiment 2, and low prevalence target absent reaction time was 4880ms in Experiment 1 compared to 2569ms in Experiment 2. This

shows that the search task in Experiment 1 was more difficult and required the observer to maintain attention for longer periods of time. Perhaps measures of vigilance abilities are stronger predictors of performance for harder tasks. This is important when considering its use as a predictor in a real-world scenario where an observer may be on a shift for hours (up to 12 hours in military satellite scanning), rather than approximately 25 minutes as in this task, as the comparison between Experiments 1 and 2 suggests that vigilance may be a more important characteristic when searching for longer. It is also worth noting that though vigilance was a non-significant predictor in the Experiment 2 accuracy regression, it was still significantly correlated with accuracy in the predicted direction (higher measures of vigilance predict higher accuracy), suggesting that the variance it explained may have been shared with a new predictor (the additional far transfer high prevalence search or fluid intelligence predictors). For these reasons, I believe that vigilance should remain a potential predictor in future individual differences research.

Attentional control was a significant predictor of performance in Experiment 1, but not Experiment 2. I originally predicted in Experiment 1 that attentional control would predict faster reaction times, but instead found that it predicted accuracy and not reaction time. By adding eye-tracking in Experiment 2, I hoped to clarify how attentional control related to accuracy, but I failed to replicate the original finding. I now believe that attentional control's validity as a predictor of accuracy was likely a Type 1 error as it did not predict the measure I expected it to in Experiment 1, nor did it replicate in Experiment 2.

As a further test of the Experiment 1 predictors' validities, I used the beta weights established from the Experiment 1 accuracy regression and the Experiment 2 predictor measures to predict performance in the low prevalence visual search task in Experiment 2. Despite a

different dependent variable (Ts and Ls search performance in Experiment 1, baggage screening search performance in Experiment 2) and lacking fluid intelligence as a predictor, I found that the beta weights from Experiment 1 accounted for a significant proportion in the variance ($R^2 =$ .42) in low prevalence visual search performance with baggage screening stimuli, demonstrating how robust these predictors are.

Through the use of eye-tracking, I investigated how the predictors related to visual search performance. I used each significant predictor of accuracy in regressions predicting quitting thresholds, selection errors, identification errors, re-inspection rate, target decision time, and distractor decision time. First, I found that only high prevalence far transfer search predicted quitting thresholds. It is somewhat surprising that low quitting thresholds, the primary cause of low accuracy in low prevalence search, are only weakly predicted by a single measure. This suggests some other factor is influencing this result.

A measure strongly related to quitting thresholds is selection error rate. Selection errors happen when a miss is caused by the failure to find a target before reaching an individual's quitting threshold. I found that higher levels of WMC, fluid intelligence, and high prevalence near transfer search all predicted fewer selection errors. Though only far transfer search was also a significant predictor of quitting thresholds, both fluid intelligence and WMC were positively, though not significantly, correlated with quitting thresholds (rs of .20 and .15, respectively). This suggests increased quitting thresholds are likely the mechanism through which selection errors are decreased.

WMC and high prevalence near transfer search were also significant predictors of identification error rates. Identification errors are an inherent part of the high prevalence near transfer search measure. Any identification errors that occurred in the high prevalence search

would lower accuracy in that task, thus predicting lower accuracy in the low prevalence search as well. Though I cannot isolate a mechanism through which those with high WMC commit fewer identification errors, this finding does fit with the initial prediction that those with higher WMC are better able to maintain and match target templates and replicates findings from Peltier and Becker (2016b). We also know that those with high WMC accumulate information more rapidly (Schmiedek, Oberauer, Wilhelm, Suss, & Wittmann, 2007), which may lead to a decreased chance of misidentifying an inspected object. Further, Engle, Tuholski, Laughlin, and Conway (1999) claim that those with higher WMC maintain representations in working memory better than those with low WMC, which may lead to a greater probability of accurately matching a template in working memory to an actual item. Faster evidence accumulation and greater template to item matching also fits with the finding that WMC is negatively correlated with false alarms.

I was interested in item re-inspection rate as a measure of search efficiency; if an observer is fixating the same objects more than once, this would be classified as inefficient search. High prevalence far transfer search was the only predictor of reinspection rate and higher performance predicted more re-inspections. I originally predicted that WMC would be negatively correlated with re-inspections, as those with higher WMC may be able remember previously inspected areas and avoid returning to them. Given the violation of my predictions, I took a closer look at the reinspection data. I found that four of the five significant predictors of accuracy (near and far high prevalence search accuracy, WMC, and fluid intelligence) were all positively correlated (though only the high prevalence searches were significantly correlated) with reinspection rate. This suggests that while re-inspections may be inefficient in terms of reaction time, they may be necessary to maintain high accuracy and prevent identification errors or false

alarms from misidentifying an inspected object. Reinforcing this interpretation is the significant negative correlation ($r = -.40$) between re-inspection rate and identification errors.

We also performed regressions predicting target and distractor decision times. I found that fluid intelligence predicted faster target decision times, which replicates previous research suggesting that higher intelligence is associated with faster item inspections (Deary & Stough, 1996). Though the negative correlation between fluid intelligence and identification errors did not reach significance, the faster target decision times without an increase in identification errors or false alarms suggests that those with higher fluid intelligence are not making a speed accuracy tradeoff, and are instead increasing efficiency. Though fluid intelligence predicted faster target decision times, it did not predict distractor decision times. Only high prevalence far transfer search performance was a significant predictor of distractor decision time, such that better performance was associated with slower distractor decision times.

Overall, this experiment strengthened the results of Experiment 1 in three ways. First, I replicated the major finding: an individual differences approach to reducing the LPE can be extremely effective. A short battery of tasks that takes approximately one hour can account for just over half of the variance in low prevalence visual search accuracy. Importantly, WMC, high prevalence visual search performance, and introversion all replicated as significant predictors of performance, and even though vigilance dropped out as a significant predictor in the regression model, it remained significantly correlated with low prevalence accuracy.

In Experiment 2, I found two new tasks that were significant predictors of low prevalence visual search performance. Fluid intelligence was a significant predictor of accuracy, selection errors, and target decision time, demonstrating its value as a highly predictive individual difference measure. By including both a near transfer and far transfer search task, I found that

59

both are predictors of several components of visual search, each accounting for a significant proportion of the variance. The finding that the far transfer search task was predictive of accuracy even after accounting for the variance explained by a near transfer task suggests that it would be useful to include both a near transfer task and a much more difficult task to predict performance. As the more difficult task was predictive of reaction time and thus quitting thresholds, I hypothesize that it may be tapping into the effort or patience that an individual invests in task performance, though it is not possible to say for certain.

Finally, the use of eye-tracking for allowed the first investigation into how several individual difference measures relate to the critical accuracy measure. Accuracy is made up of selection and identification errors. I found that WMC, fluid intelligence, and high prevalence near transfer search were all predictive of fewer selection errors, while WMC and high prevalence near transfer search were predictive of fewer identification errors. WMC predicting identification errors replicates a previous finding from our lab (Peltier & Becker, 2017a) and builds on the claims of Schwark et al. (2013) who found that WMC predicts accuracy, but claimed that it was predictive of only selection errors. The findings that high prevalence near transfer search predicts both selection errors and identification errors, while fluid intelligence predicts selection errors is novel.

CHAPTER 4. GENERAL DISCUSSION

In these experiments, I investigated the abilities of several individual differences to predict low prevalence visual search. In Experiment 1, I used measures of near transfer high prevalence visual search accuracy, WMC, vigilance, attentional control, and personality to predict visual search performance in a Ts and Ls task. After establishing the LPE in the search task, I entered all predictors into two regression models predicting accuracy and reaction time. The results showed that I can account for 52% of the variance in low prevalence visual search accuracy using near transfer search performance, WMC, vigilance, attentional control, and introversion. By then entering each of the predictors into a regression predicting reaction time, and finding that near transfer search, working memory capacity, attentional control, and introversion all predicted longer reaction times, consistent with the view that these predictors increase accuracy via an increase in quitting threshold. This exploratory study demonstrated that an individual differences approach to increasing low prevalence visual search performance is a viable one. There were, however, several limitations.

First, as this was an exploratory study, it is important to replicate the findings. As interventions that have attempted to improve low prevalence visual search performance have largely been failures, it is reasonable to view this broad success with skepticism. Second, without eye-tracking, I could say for certain that the predictors of increased reaction time were related to quitting thresholds. It is possible that those predictors were related to increased item inspection times, thus increasing overall reaction time without increasing quitting thresholds. Further, the lack of eye-tracking prevents us from determining which predictors are related to identification errors or selection errors. Third, despite the aim to generalize to a more real-world situation, the stimuli I used were not representative of a real-world search task. Finally, finding additional

predictors of visual search performance would help build a better selection tool for employers. Experiment 2 addressed these weaknesses.

In Experiment 2, I eye-tracked observers while they performed a more realistic baggage screening task before completing the same battery of tasks from Experiment 1, in addition to three new predictors, far transfer search, fluid intelligence, and task unrelated thoughts. Using each predictor in a regression predicting accuracy, I found near transfer search, far transfer search, WMC, fluid intelligence, and introversion were all significant predictors and accounted for 53% of the variance in low prevalence visual search accuracy. I found that high prevalence near and far transfer search accuracy were significant predictors of reaction time and accounted for 26% of the variance in low prevalence visual search reaction time. I also sought to replicate the findings from Experiment 1 and ran a regression using only the previously significant predictors. The previous predictors were significant, with the exception of attentional control and vigilance (though vigilance was significantly correlated), critically demonstrating that the predictors near transfer search, WMC, and introversion were unlikely due to a Type 1 error. Further, by using the significant predictors' betas established in the Experiment 1 accuracy regression, I was able to validate the replicated predictors and show that these predictors can account for 44% of the variance in low prevalence visual search accuracy in a novel task with a new sample of observers.

Experiment 2 was novel in that it was the first study that used eye-tracking to analyze how multiple individual difference measures relate to visual search performance. The primary results of interest were that WMC, fluid intelligence, and near transfer search were all predictive of selection errors, while WMC and near transfer search were predictive of identification errors. These analyses aided in understanding which components of accuracy the individual difference

62

measures related to. I also found that far transfer search was a significant predictor of re-inspection rate, fluid intelligence was a significant predictor of target decision time, and far transfer search was a significant predictor of distractor decision time. Despite these many findings, there are still questions researchers should aim to address.

Among the questions that future research should investigate is how critical a vigilance measure is to predict performance during longer search durations, the extent to which the prediction model I have identified would generalize to predicting expert performance, searches for critical real-world targets (e.g., cancers in radiological scans), and more real-world work environments (Clark, Cain, Adamo, & Mitroff, 2012). This experiment used 10% target prevalence as the low prevalence condition, but research has shown that the prevalence effect (i.e. increase in miss rates) becomes more pronounced as prevalence decreases further (Mitroff & Biggs, 2014; Wolfe et al., 2005), meaning that I may be underestimating the effects of extremely low target prevalence that are characteristic of real-world searches. However, my data suggests the strength of these predictors increases as prevalence decreases because these predictor tasks account for more variance in search performance as target prevalence decreases from 50% to 10%. If this trend continues to even lower prevalence rates, such as the .3% prevalence found in breast cancer screening (Gur et al., 2004), then the predictors may be even stronger. Alternatively, I could find that the predictors' strengths are overestimated when predicting expert performance because training may overcome basic individual differences; though in the expertise literature there is evidence suggesting that initial abilities, such as working memory capacity, predict performance over and above practice (Meinz & Hambrick, 2010). Future research addressing these concerns would be required to establish the value of implementing these types of screening tools in a real-world context.

Despite the need for future research, the current investigation has both theoretical and practical relevance. In terms of theoretical relevance, I was able to establish a relationship between predictors and the two different accuracy components, identification errors and selection errors. Further, the finding that far transfer search task performance was highly predictive of the more real-world representative search task indicates that visual search research done in the lab with artificial stimuli may be able to generalize to more realistic scenarios. In terms of practical relevance, a screener that uses the five factors of my model (WMC, fluid intelligence, high prevalence near transfer search, high prevalence far transfer search, and introversion) to identify people who would be likely to perform well in situations that require the detection of low prevalence targets, like baggage screeners, may significantly increase target detections.

APPENDICES

Appendix A. Stimuli for Tasks

The conditions required to use the baggage screening stimuli for this project included an agreement to not share or print the stimuli. We thank the Kedlin Company (airportscannergame.com) for sharing their stimuli.
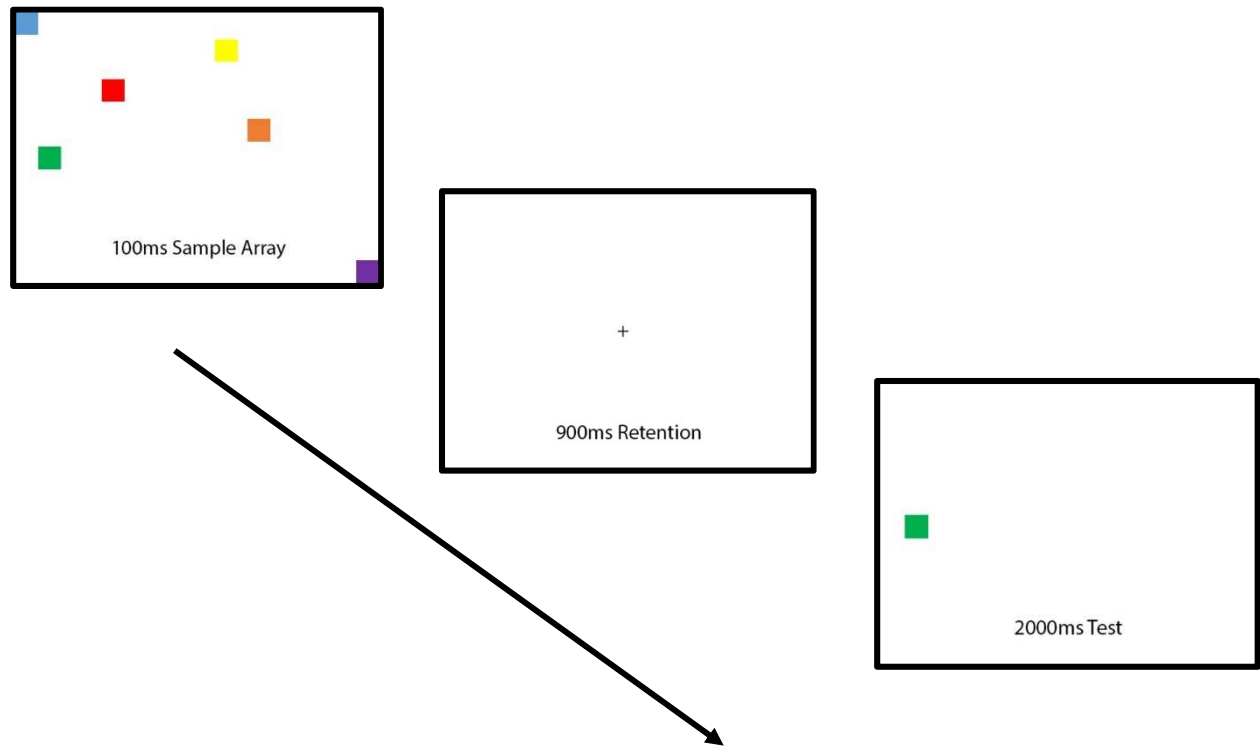
Figure 7. *Change Detection Task.* Report the probed item as same or different color from the item that was previously presented in the same location.
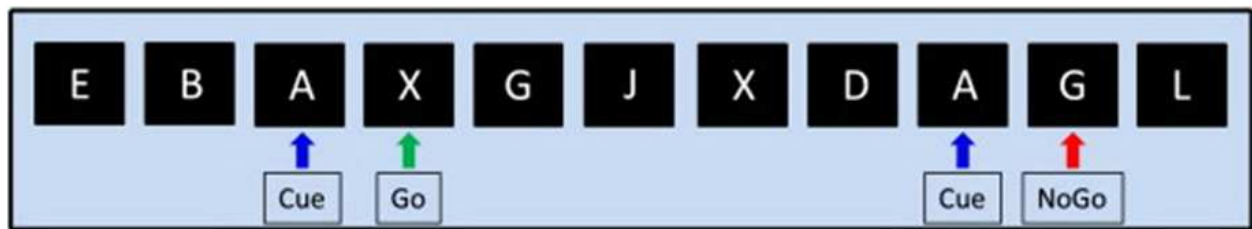
Figure 8. *AX Continuous Performance Task.* Press a button when you see an X preceded by an A. Withhold your response for all other letters.
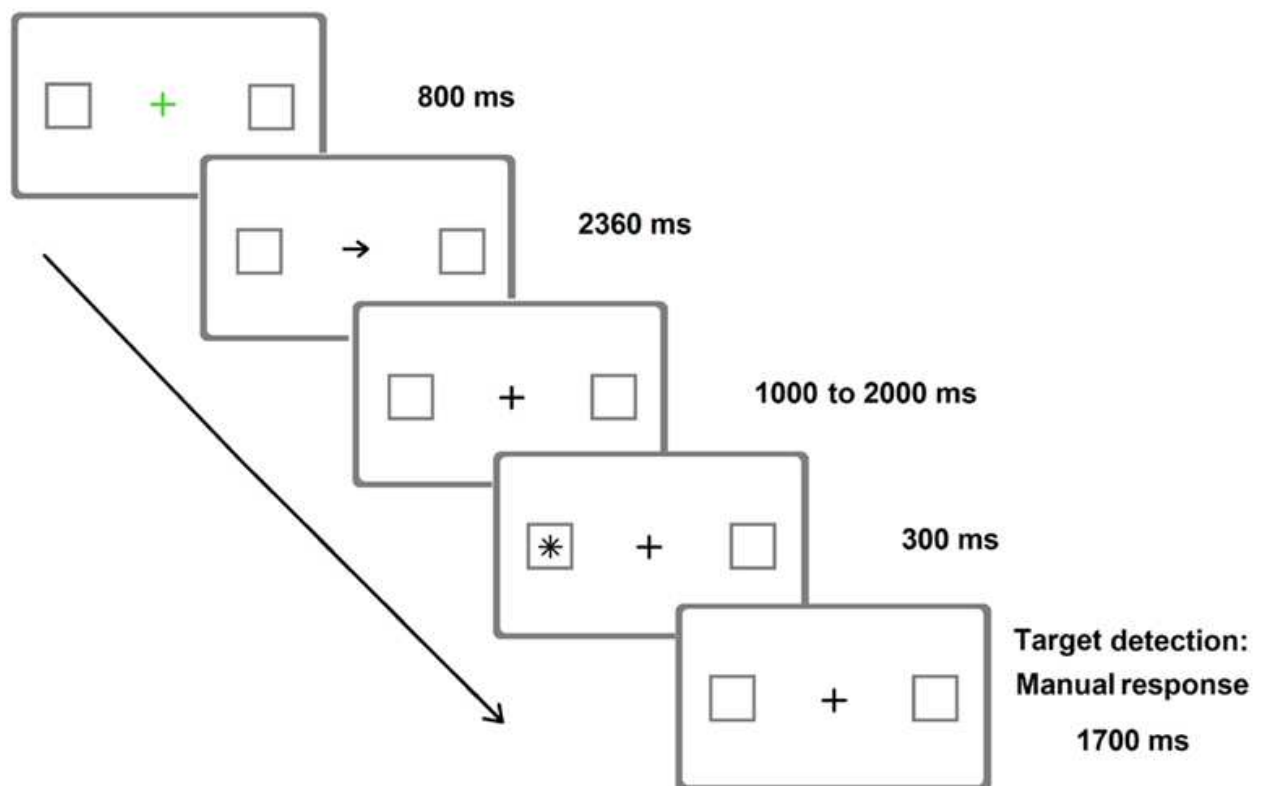
Figure 9. *Posner Cuing Task*. Identify which side of the fixation point a probe item appears on.

Raven's Progressive Matrices (odd items)

For each item in this test, you will see a box with 3 rows and 3 columns. The box should consist of 9 objects, but only 8 are provided. The object in the lower-right corner is always missing. Your job is to determine which of the 8 possible solutions logically fits in the missing box.

You can do this by looking for patterns. These patterns may occur across the rows, down the columns, both across the rows and down the columns, or throughout all of the spaces. Each of the possible solutions is numbered (1 to 8). When you have determined the correct answer, please select it from the options. There is one, and only one, correct answer for each item.

You will now see some practice problems to get an idea of what the problems will be like.

This is an example of the kind of problem you will solve during this test.

The top part of this problem is a pattern with a bit cut out of it. Look at the pattern, think what the piece needed to complete the pattern correctly both along and down must be like. Then find the right bit out of the eight bits shown below.

Only one of these pieces is perfectly correct. Number 1 completes the pattern correctly going downwards but is wrong going the other way. Number 4 is correct going along, but is wrong going downward.

Which piece is correct going both ways?

Number eight is the right bit.

Please press enter to continue.
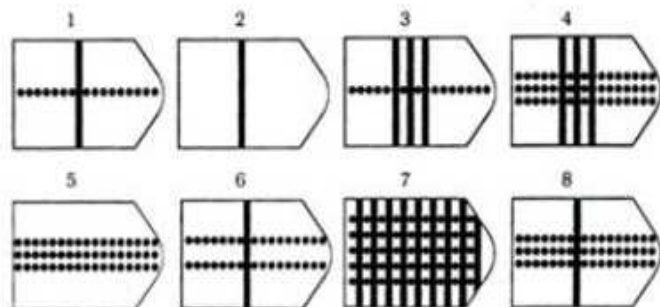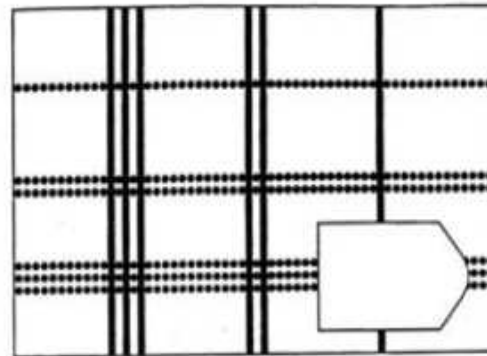
Figure 10. *Raven's Progressive Matrices practice problem.*

70
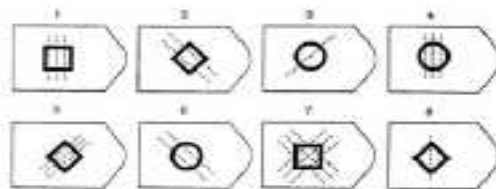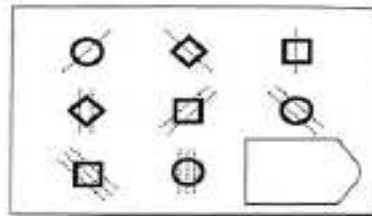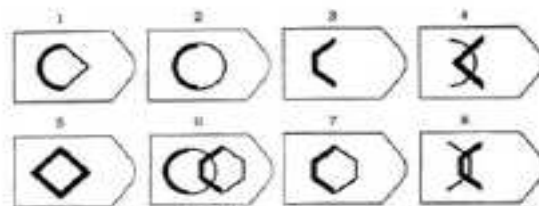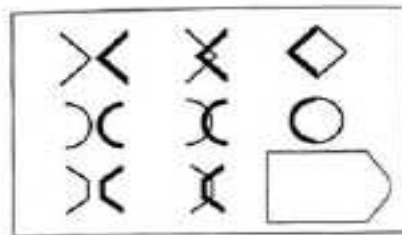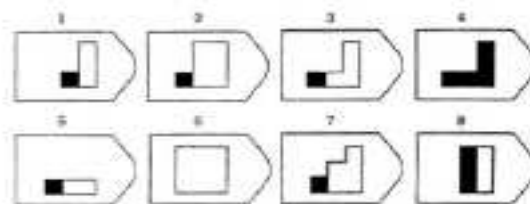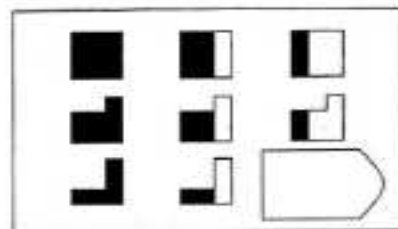
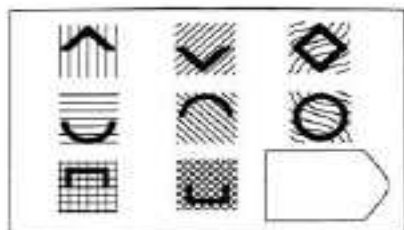**Figure 11.** *Raven's Progressive Matrices items 1-3.*
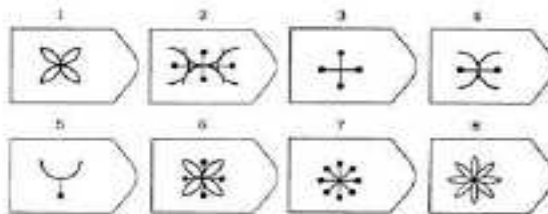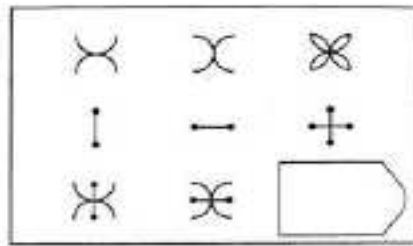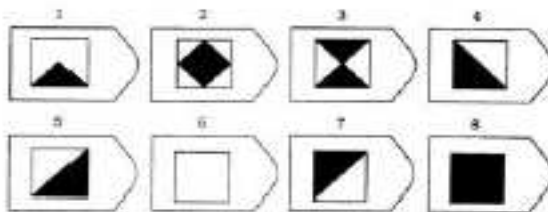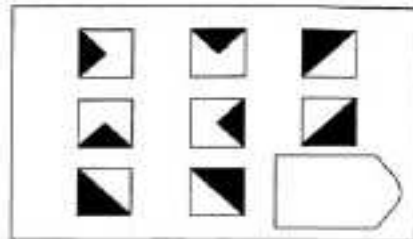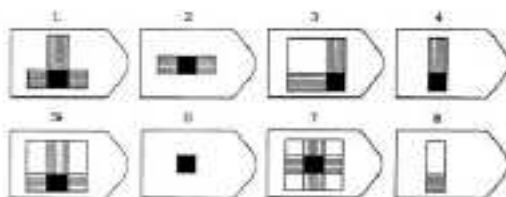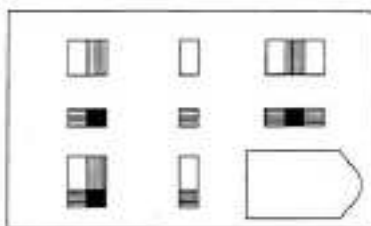
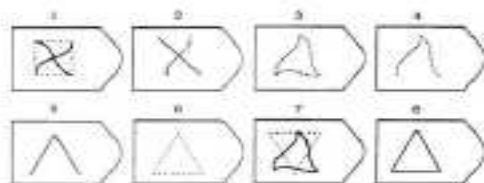**Figure 12.** *Raven's Progressive Matrices items 4-6.*

7.
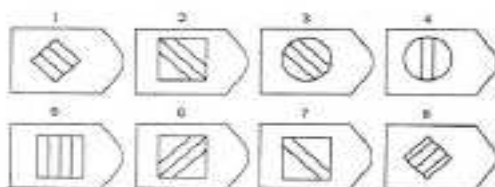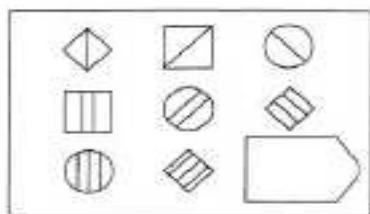
8.
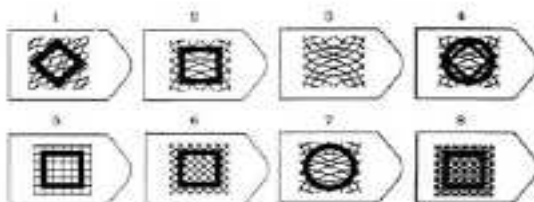
9.
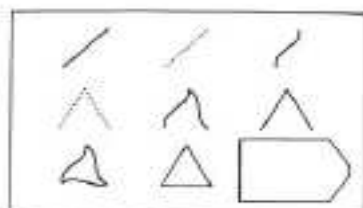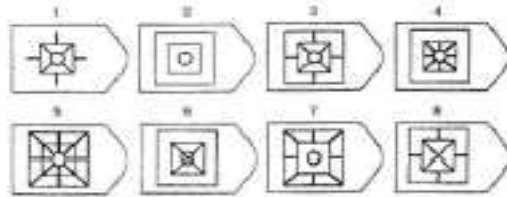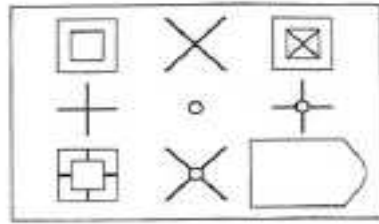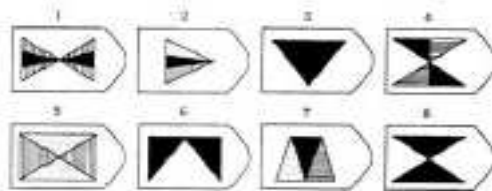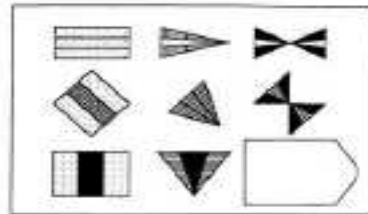

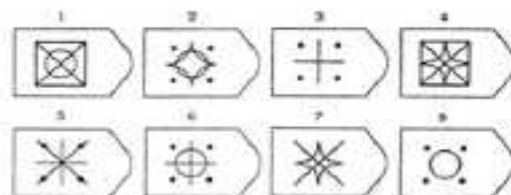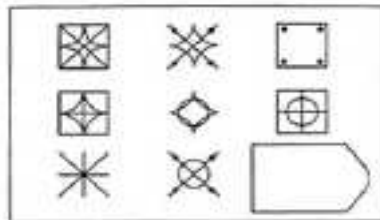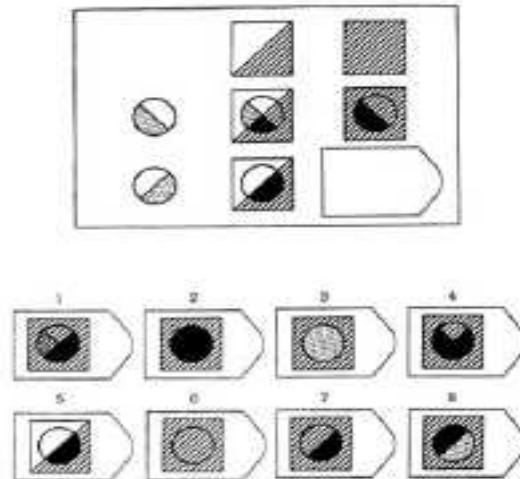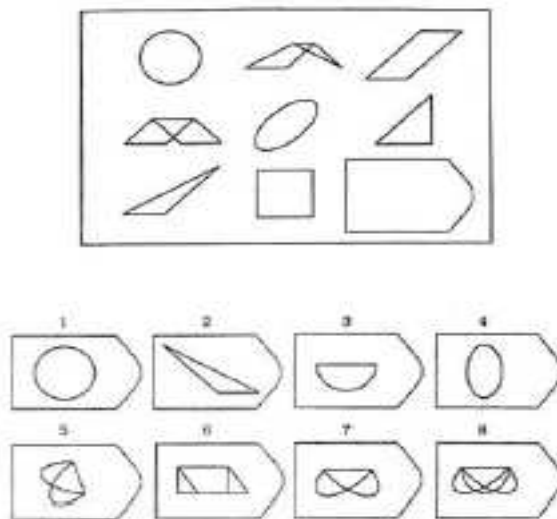**Figure 13.** *Raven's Progressive Matrices items 7-9.*

73

10.

11.

12.

Figure 14. *Raven's Progressive Matrices items 10-12.*

74

13.

14.

15.
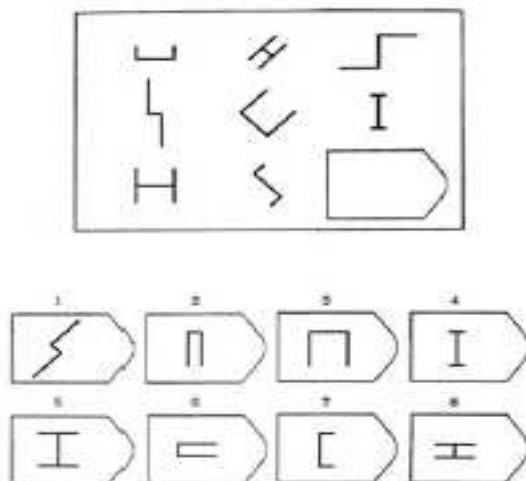
Figure 15. *Raven's Progressive Matrices items 13-15.*

16.

17.

18.

Figure 16. *Raven's Progressive Matrices items 15-18.*

Appendix B. Survey
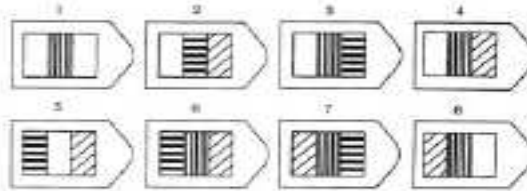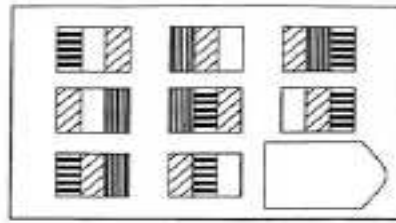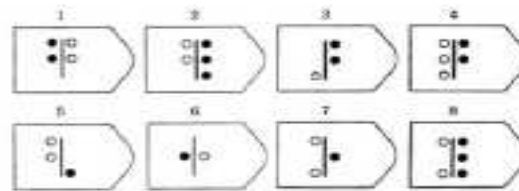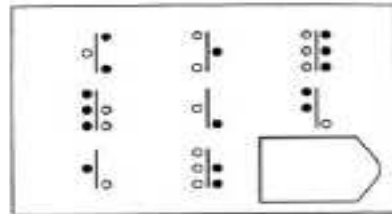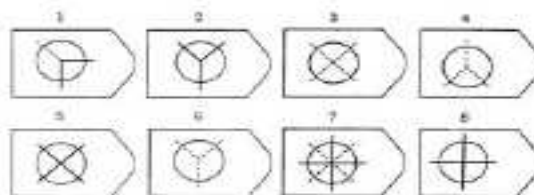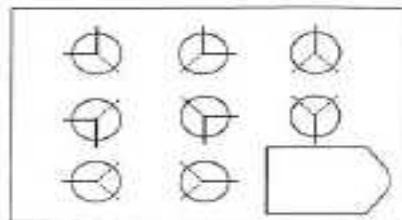
Item	Factor	Question
1	Introversion	Am the life of the party.
2	Agreeableness	Sympathize with others' feelings.
3	Conscientiousness	Get chores done right away.
4	Neuroticism	Have frequent mood swings.
5	Intellect (Openness)	Have a vivid imagination.
6	Introversion	Don't talk a lot. (R)
7	Agreeableness	Am not interested in other people's problems. (R)
8	Conscientiousness	Often forget to put things back in their proper place. (R)
9	Neuroticism	Am relaxed most of the time. (R)
10	Intellect (Openness)	Am not interested in abstract ideas. (R)
11	Introversion	Talk to a lot of different people at parties.
12	Agreeableness	Feel others' emotions.
13	Conscientiousness	Like order.
14	Neuroticism	Get upset easily.
15	Intellect (Openness)	Have difficulty understanding abstract ideas. (R)
16	Introversion	Keep in the background. (R)
17	Agreeableness	Am not really interested in others. (R)
18	Conscientiousness	Make a mess of things (R)
19	Neuroticism	Seldom feel blue. (R)
20	Intellect (Openness)	Do not have a good imagination. (R)

Note. (R) indicates a reverse scored item.

Table 10. *Mini IPIP Questions and Personality Factors.*

REFERENCES

REFERENCES

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: the same or different constructs? *Psychol Bull, 131*(1), 30-60. doi:10.1037/0033-2909.131.1.30

Adamo, S. H., Cain, M. S., & Mitroff, S. R. (2016). An individual differences approach to multiple-target visual search errors: How search errors relate to different characteristics of attention. *Vision Res*.

Allen, M., Smallwood, J., Christensen, J., Gramm, D., Rasmussen, B., Jensen, C. G., . . . Lutz, A. (2013). The balanced mind: the variability of task-unrelated thoughts predicts error monitoring. *Front Hum Neurosci, 7*, 743. doi:10.3389/fnhum.2013.00743

Ariga, A., & Lleras, A. (2011). Brief and rare mental "breaks" keep you focused: deactivation and reactivation of task goals preempt vigilance decrements. *Cognition, 118*(3), 439-443. doi:10.1016/j.cognition.2010.12.007

Bays, P. M., & Husain, M. (2012). Active inhibition and memory promote exploration and search of natural scenes. *J Vis, 12*(8). doi:10.1167/12.8.8

Berlin, L. (1994). Reporting the" missed" radiologic diagnosis: medicolegal and ethical considerations. *Radiology, 192*(1), 183-187.

Biggs, A. T., Clark, K., & Mitroff, S. R. (2017). Who should be searching? Differences in personality can affect visual search accuracy. *Personality and Individual Differences, 116*, 353-358. doi:10.1016/j.paid.2017.04.045

Chun, M., & Wolfe, J. M. (1996). Just Say No: How Are Visual Searches Terminated When There Is No Target Present? *Cognitive Psychology, 30*.

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A taxonomy of external and internal attention. *Annual review of psychology, 62*, 73-101. doi:10.1146/annurev.psych.093008.100427

Covey, T. J., Shucard, J. L., Violanti, J. M., Lee, J., & Shucard, D. W. (2013). The effects of exposure to traumatic stressors on inhibitory control in police officers: a dense electrode array study using a Go/NoGo continuous performance task. *International Journal of Psychophysiology, 87*(3), 363-375. doi:10.1016/j.ijpsycho.2013.03.009

Davies, D. R., & Parasuraman, R. (1982). The psychology of vigilance. *Academic Press*, 107-117.

Deary, I. J., & Stough, C. (1996). Intelligence and inspection time: Achievements, prospects, and problems. *American Psychologist, 51*(6), 599.

Donnellan, M. B., Oswald, F. L., Baird, B. M., & Lucas, R. E. (2006). The mini-IPIP scales: tiny-yet-effective measures of the Big Five factors of personality. *Psychological assessment, 18*(2), 192-203. doi:10.1037/1040-3590.18.2.192

Engle, R. W., & Kane, M. J. (2003). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of learning and motivation, 44*, 145-199.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of experimental psychology: General, 128*(3), 309.

Evans, K., Birdwell, R., & Wolfe, J. (2013). If you don't find it often, you often don't find it: Why some cancers are missed in breast cancer screening. *PLoS One*.

Eysenck, H. J. (1967). *The biological basis of personality* (Vol. 689): Transaction publishers.

Fishel, J., Levine, M., & Date, J. (2015). Undercover DHS Tests Find Security Failures at US Airports. *ABC News*.

Fleck, M., & Mitroff, S. (2007). Rare Targets Are Rarely Missed in Correctable Search. *Psychol Sci, 18*.

Godwin, H. J., Menneer, T., Cave, K. R., Thaibsyah, M., & Donnelly, N. (2015). The effects of increasing target prevalence on information processing during visual search. *Psychon Bull Rev, 22*(2), 469-475. doi:10.3758/s13423-014-0686-2

Godwin, H. J., Menneer, T., Riggs, C. A., Cave, K. R., & Donnelly, N. (2014). Perceptual failures in the selection and identification of low-prevalence targets in relative prevalence visual search. *Atten Percept Psychophys*. doi:10.3758/s13414-014-0762-8

Godwin, H. J., Menneer, T., Riggs, C. A., Taunton, D., Cave, K. R., & Donnel, N. (2015). Understanding the contribution of target repetition and target expectation to the emergence of the prevalence effect in visual search. *Psychon Bull Rev*. doi:10.3758/s13423-015-0970-9

Goldberg, L. R. (1990). An alternative" description of personality": the big-five factor structure. *J Pers Soc Psychol, 59*(6), 1216.

Green, D., & Swets, J. (1966). *Signal detection theory and psychophysics* (Vol. 1): Wiley New York.

Gur, D., Sumkin, J. H., Rockette, H. E., Ganott, M., Hakim, C., Hardesty, L., . . . Wallace, L. (2004). Changes in Breast Cancer Detection and Mammography Recall Rates After the Introduction of a Computer-Aided Detection System. *JNCI Journal of the National Cancer Institute, 96*(3), 185-190. doi:10.1093/jnci/djh067

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Hout, M. C., Walenchok, S. C., Goldinger, S. D., & Wolfe, J. M. (2015). Failures of perception in the low-prevalence effect: Evidence from active and passive visual search. *Journal of Experimental Psychology: Human Perception and Performance, 41*(4), 977-994. doi:10.1037/xhp0000053

Ishibashi, K., Kita, S., & Wolfe, J. M. (2012). The effects of local prevalence and explicit expectations on search termination times. *Attention Perception Psychophysics, 74*(1), 115-123. doi:10.3758/s13414-011-0225-4

Jensen, A. R. (1998). The g factor: The science of mental ability.

Judge, T. A., & Ilies, R. (2002). Relationship of personality to performance motivation: A meta-analytic review. *Journal of Applied Psychology, 87*(4), 797-807. doi:10.1037/0021-9010.87.4.797

Kane, M. J., Conway, A. R., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. *Variation in working memory, 1*, 21-48.

Koelega, H. S. (1992). Extraversion and vigilance performance: 30 years of inconsistencies. *Psychological bulletin, 112*(2), 239.

Kunar, M. A., Rich, A. N., & Wolfe, J. M. (2010). Spatial and temporal separation fails to counteract the effects of low prevalence in visual search. *Visual Cognition, 18*(6), 881-897. doi:10.1080/13506280903361988

Levinson, D. B., Smallwood, J., & Davidson, R. J. (2012). The Persistence of Thought. *Psychol Sci, 23*(4), 375-380. doi:doi:10.1177/0956797611431465

Manly, T., Robertson, I. H., Galloway, M., & Hawkins, K. (1999). The absent mind:: further investigations of sustained attention to response. *Neuropsychologia, 37*(6), 661-670.

McVay, J. C., & Kane, M. J. (2012). Drifting from slow to "D'oh!": working memory capacity and mind wandering predict extreme reaction times and executive control errors. *J Exp Psychol Learn Mem Cogn, 38*(3), 525-549. doi:10.1037/a0025896

Mitroff, S. R., & Biggs, A. T. (2014). The ultra-rare-item effect: visual search for exceedingly rare items is highly susceptible to error. *Psychol Sci, 25*(1), 284-289. doi:10.1177/0956797613504221

Navalpakkam, V., Koch, C., & Perona, P. (2009). Homo economicus in visual search. *J Vis, 9*(1), 31 31-16. doi:10.1167/9.1.31

Newton, T., Slade, P., Butler, N., & Murphy, P. (1992). Personality and performance on a simple visual search task. *Personality and Individual Differences, 13*(3), 381-382.

Pashler, H. (1988). Familiarity and visual change detection. *Perception and Psychophysics, 44*(4).

Peltier, C., & Becker, M. W. (2016a). Decision Processes in Visual Search as a Function of Target Prevalence. *Journal of Experimental Psychology: Human Perception and Performance*. doi:10.1037/xhp0000248

Peltier, C., & Becker, M. W. (2016b). Working Memory Capacity Predicts Selection and Identification Errors in Visual Search. *Perception*, 0301006616678421.

Peltier, C., & Becker, M. W. (2017a). Individual differences predict low prevalence visual search performance. *Cogn Res Princ Implic, 2*(1), 5. doi:10.1186/s41235-016-0042-3

Peltier, C., & Becker, M. W. (2017b). Target-present guessing as a function of target prevalence and accumulated information in visual search. *Atten Percept Psychophys, 79*(4), 1064-1069. doi:10.3758/s13414-017-1297-6

Posner, M. I. (1980). Orienting of attention. *Q J Exp Psychol (Hove), 32*(1), 3-25. doi:10.1080/00335558008248231

Ratcliff, R., & McKoon, G. (2008). The Diffusion Decision Model: Theory and Data for Two-Choise Decision Tasks. *Neural Computation, 20*(4).

Raven, J. C. (1998). *Raven's progressive matrices*: Oxford Psychologists Press Oxford.

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

Schmidt, F. L., & Hunter, J. E. (1993). Tacit Knowledge, Practical Intelligence, General Mental Ability, and Job Knowledge. *Current Directions in Psychological Science, 2*(1), 8-9. doi:10.1111/1467-8721.ep10770456

Schmidt, F. L., & Hunter, J. E. (1998). The Validity and Utility of Selection Methods in Personnel Psychology Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological bulletin, 124*(2).

Schmiedek, F., Oberauer, K., Wilhelm, O., Suss, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *J Exp Psychol Gen, 136*(3), 414-429. doi:10.1037/0096-3445.136.3.414

Schwaninger, A., Hardmeier, D., & Hofer, F. (2005). Aviation Security Screeners Visual Abilities & Visual Knowledge Measurement. *IEEE Aerospace and electronic systems magazine, 20*.

Schwark, J., Sandry, J., & Dolgov, I. (2013). Evidence for a positive relationship between working-memory capacity and detection of low-prevalence targets in visual search. *Perception, 42*(1), 112-114. doi:10.1068/p7386

Schwark, J., Sandry, J., Macdonald, J., & Dolgov, I. (2012). False feedback increases detection of low-prevalence targets in visual search. *Attention Perception Psychophysics, 74*(8), 1583-1589. doi:10.3758/s13414-012-0354-4

See, J. E., Warm, J. S., Dember, W. N., & Howe, S. R. (1997). Vigilance and Signal Detection Theory: An Empirical Evaluation of Five Measures of Response Bias. *Human Factors, 39*(1).

Sen, A., & Goel, N. (1981). Functional relation between personality types and some impirically derived TSD parameters in a visual searching task. *Psychological Studies, 26*, 23-27.

Smallwood, J., Davies, J. B., Heim, D., Finnigan, F., Sudberry, M., O'Connor, R., & Obonsawin, M. (2004). Subjective experience and the attentional lapse: task engagement and disengagement during sustained attention. *Conscious Cogn, 13*(4), 657-690. doi:10.1016/j.concog.2004.06.003

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education., 2*, 53-55. doi:10.5116/ijme.4dfb.8dfd

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136.

Unsworth, N., Heitz, R., Schrock, J., & Engle, R. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3).

Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance Requires Hard Mental Work and Is Stressful. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*(3), 433-441. doi:10.1518/001872008x312152

Williams, L. (1966). The effect of target specification on objects fixated during visual search. *Perception & Psychophysics, 1*(9), 315-318.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychon Bull Rev, 1*(2), 202-238.

Wolfe, J. M. (2014). When do I quit? The search termination problem in visual search. *The Influence of Attention, Learning, and Motivation on Visual Search*.

Wolfe, J. M., Horowitz, T. S., & Kenner, N. M. (2005). Rare items often missed in visual searches. *Nature, 435*(26), 439-440.

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of experimental psychology: General, 136*(4), 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*(2), 121-124. doi:10.1016/j.cub.2009.11.066

Zhang, J., & Mueller, S. T. (2005). A note on ROC analysis and non-parametric estimate of sensitivity. *Psychometrika, 70*(1), 203-212. doi:10.1007/s11336-003-1119-8