2010

Empirical Evaluation Of The Effectiveness Of Eye Tracking-based Search Performance Diagnosis And Feedback Methods

Meredith Carroll
University of Central Florida

EMPIRICAL EVALUATION OF THE EFFECTIVENESS OF EYE TRACKING-BASED SEARCH PERFORMANCE DIAGNOSIS AND FEEDBACK METHODS

by

MEREDITH BELL CARROLL
B.S. University of Virginia, 2001
M.S. Florida Institute of Technology, 2003

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Psychology
in the College of Sciences
at the University of Central Florida
Orlando, Florida

Spring Term
2010

Major Professor: Mustapha Mouloua

ABSTRACT

In today's complex combat environments, troops are often faced with increasingly challenging tasks different from those experienced in the past. Warfighters must be trained in adaptive perceptual skill sets, such as search strategies that enable them to detect threats across any number of environmental, cultural, and situational conditions. The goal of the present study was to explore how advanced technology, specifically eye tracking, can be used to increase understanding of perceptual processes such as search and detection and provide tools that can be used to train search skills. Experiment 1 examined a method of diagnosing perceptual performance in order to be able to identify the perceptual root cause of target detection deficiencies and how these impact overall target detection performance. Findings indicate the method can be used to pinpoint where in the perceptual process a target miss originated, whether due to ineffective search strategy, inability to detect the subtle cues of the threat or inability to recognize these cues as indicative of a threat. Experiment 2 examined the training effectiveness of providing trainees with process level tailored feedback which incorporates elements of expert and trainee scan patterns. Findings indicate that providing trainees with elements of either expert or trainee scan patterns has the ability to significantly improve the search strategy being employed by the trainee. This work provides strong support for the use of eye tracking based perceptual performance diagnosis methods and training strategies in improving trainee search performance for complex target detection tasks.

To Matthew,
My Husband, Mentor and Best Friend

ACKNOWLEDGMENTS

Phil 4:13

This dissertation would not have been possible without the support and encouragement of many individuals. First, I would like to thank my committee members. Thank you to my advisor, Dr. Mustapha Mouloua who provided support, both academically and emotionally throughout the entire program. Thank you to Dr. Steve Fiore whose contribution to this dissertation went above and beyond and truly enabled the success of this study. Thank you to Dr. Valerie Sims who provided guidance and encouragement throughout the dissertation process. Thank you to Dr. Laura Milham, whose insight as a committee member, expertise as a colleague and encouragement as a friend has shaped me both professionally and personally.

This research was sponsored by the Office of Naval Research within the Observational Skills Enhancement and Retainment Virtual Environment (ObSERVE) effort under the Human Performance Training and Education (HPT&E) program. I would like to thank the HPT&E Program Manager, Dr. Roy Stripling, as well as the many individuals in the Marine Corps who made this research possible. Thank you to the Marine Corps Warfighting Lab (MCWL), the Marine Corps Training and Education Command (TECOM), the Combat Hunter Instructors at the Marine Corps School of Infantry and the Scout Sniper Instructors at the Marine Corps Scout Sniper School.

I would also like to thank the many friends and family members who have provided a helping hand, an open ear or a supportive shoulder. Most of all, thank you to my husband, Matthew. Matthew, you are my stronghold without which this dissertation as well as many other of my achievements would not have been possible.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS/ABBREVIATIONS

AOI	Areas of Interest
CAVS	Computer Assisted Visual Search
CR	Correct Rejection
DDDS	Difficult Detect Difficult Search
DDES	Difficult Detect Easy Search
DM	Decision Making
DV	Dependent Variable
EDDS	Easy Detect Difficult Search
EDES	Easy Detect Easy Search
FA	False Alarm
FIT	Feature Integration Theory
H	Hit
HCI	Human Computer Interaction
HF	Hidden Figures
HIP	Human Information Processing
IED	Improvised Explosive Device
I MEF	I Marine Expeditionary Force
ITS	Intelligent Tutoring System
IV	Independent Variable
IW	Irregular Warfare
KCR	Knowledge of Correct Response
KR	Knowledge of Results
KSA	Knowledge, Skills and Abilities
MCWL	Marine Corps Warfighting Lab
M	Miss
NTDS	Target Not Present, Difficult Search
NTES	Target Not Present, Easy Search
ObSERVE	Observational Skills Enhancement and Retainment Virtual Environment
ONR	Office of Naval Research
OP	Observation Post
SA	Situation Awareness
SME	Subject Matter Expert
SO	Spatial Orientation
TARGETs	Targeted Acceptable Responses to Generated Events or Tasks
TDS	Target Present, Difficult Search
TES	Target Present, Easy Search
VE	Virtual Environment
VZ	Visualization

CHAPTER ONE: INTRODUCTION

"The best sensor and weapon on the battlefield is a well-trained, situationally-aware Soldier, Sailor, Airman or Marine."

- Gordon England, Deputy Secretary of Defense
Memorandum for Secretaries of the Military Departments
April 24, 2006

In today's complex combat environments, troops are often faced with increasingly challenging tasks that are different from those experienced in the past. Enemy use of Irregular Warfare (IW) techniques has motivated the military to explore new tactics and concepts such as Distributed Operations, a type of maneuver warfare in which small, highly capable units are spread across a large area of operations to provide spatial advantage (Ucko, 2007). A key tenet of this concept is training Warfighters down to the lowest rank of soldier to obtain and exploit information in order to improve cognition and decision making at the small unit level. Warfighters must have the ability to identify a range of threats, including snipers, Improvised Explosive Devices (IED), and suspicious activities and behaviors which could indicate the presence of such threats. The skills necessary to detect such threats are commonly referred to in the military as observation skills.

The preliminary findings of a recent study supported by the I Marine Expeditionary Force (I MEF), the Marine Corps Warfighting Lab (MCWL), and the Office of Naval Research (ONR) suggests that observational skills are critical to Situational Awareness (SA) and tactical decision making (Carroll, Milham, Champney, Eitelman, & Lockerd, 2007). The ability to detect static and dynamic environmental threats such as slight changes in movement, color, and texture aids in detection of threats. A major challenge to the successful performance of threat detection, however, is that tactics are continually evolving. It is not sufficient to train troops to look for specific environmental and behavioral cues, given the adaptation of the enemy. Warfighters must be trained in adaptive perceptual skill sets, such as search strategies that enable them to detect threats across any number of environmental, cultural, and situational conditions. Thus, it is crucial to systematically examine the effect of training on perceptual skills as they relate to successful threat detection.

Military Threat Detection Skills

Threat detection has been studied across a range of domains from military aviation to airport security to radiography. Threat detection in a military aviation domain is defined by Smith, Johnston, and Paris (2004) as the process of evaluating relevant cues (in their case aircraft) in the vicinity of ones environment and determining how much of a threat they represent by gathering and reviewing relevant information and deciding on what actions to take. Fiore, Scielzo, and Jentsch (2004) define threat detection in an airport screening task as consisting of the ability to rapidly recognize cues in the environment and interpret the meaning and importance of these cues. Nodine, Mello-Thoms, Kundel, and Weinstein (2002) break the radiographic interpretation threat detection task down more granularly with respect to the perceptual components, describing the task as consisting of a search for an abnormality, the recognition of an abnormality and the decision made regarding the abnormality.

The focus of this research is on the perceptual aspects of threat detection in a military Observation Post (OP) task. An OP is "a position from which military observations (visual, audible, or other means) are made" and reported in order to provide situation awareness for the commanding force (Army, 1992, Chapter 5, Section II, Paragraph 5-6). OPs "are used during screening and reconnaissance operations" and from the OP, "the squad reports the enemy size, activity, location, and disposition to the commander." (Army, 1992, Chapter 5, Section II, Paragraph 5-6). In short, the OP task entails static observation of a designated area for what can be extended periods of time. A task analysis of this task revealed decomposition of this task into key competencies of visual search, detection of anomalous cues, recognition of these cues as indicative of a threat, decision to respond and action response (Carroll et al., 2007). This research will focus primarily on the first of these competencies, visual search.

Visual Search

In a visual search task, subjects look for a target item among a number of distracting items (Leonards, Rettenbach, Nase, & Sireteanu, 2002). Visual search is theorized to have two components, 1) an effortless component in which stimulus are processed preattentively (Sireteanu & Rettenbach, 2000), essentially "popping out" at the observer and 2) an effortful component in which attention must be serially allocated to objects in the environment (Treisman & Souther, 1985). These components are influenced respectively by 1) characteristics of the environment (e.g., saliency of targets and distracters) and 2) learned search strategies. These search strategies can range from a very structured systematic search in a regular pattern (e.g., alternate up and down scan moving from left to right, (Wang, Lin, & Drury, 1997), to a less structured strategy such as searching by area or searching by component/object (Chabukswar, Gramopadhye, Melloy, & Grimes, 2003).

In current military OPs, threats such as snipers and IEDs are searched for amid an urban backdrop bustling with distracters. In order to successfully detect threats in such a complex and dynamic environment, it is necessary to develop efficient search strategies which allow the greatest number of threat identifications over the least amount of space and time. Subject Matter Experts (SMEs) from the US Marine Corps were interviewed to obtain detailed descriptions of the search strategies used to perform the search task within an OP. Both a Scout Sniper and an Observation Training Instructor identified a similar process used during the static observation task, described as follows:

> Search begins with a hasty search of the environment, from near to far, aimed at identifying the high priority areas which demand attention most. This is following by a detailed search in which high priority areas are explored further, moving right to left, starting at ground level and working your way up and starting near and working your way out. Following this, low priority areas are searched, spending less time (80% high, 20% low). During this search, if a potential threat appears, this systematic search is temporarily halted to turn attention to the new element which is quickly evaluated and attention is turned back to detailed search, picking up where left off.

In the hasty search described above, experts rely on the preattentive search component, hoping to quickly identify obvious threats. In the detailed search, experts turned to the search strategies they have developed in their experiences to identify more subtle targets.

In describing typical novice performance, these SMEs identified two common mistakes made during search including 1) students not being good judges of which areas are high priority, and 2) students not scanning in a systematic sequence. This is inline with findings in the literature which indicate experts spend more time than novices attending to relevant aspects of the stimulus (Jarodzka, Scheiter, Gerjets, & van Gog, 2009) and novices have less defined scan patterns (Burgert et al., 2007; Jarodzka et al., 2009; Kasarskis, Stehwien, Hickox, Aretz, & Wickens, 2001). The less structured scan pattern is likely due to lack of a systematic search strategy with performance being influenced most by bottom up processes which draw attention to salient features in the environment (Jarodzka et al., 2009). Given this, the two components of search which will be the focus of this research are "where" (i.e., where to look) and "how" (i.e., in what sequence to look).

Training Visual Search

Key to the training of visual search skills is that trainees not only have the opportunity to practice these skills, but also that relevant training strategies are incorporated to ensure learning. Training strategies are training interventions that can be employed in practice environments that will optimize learning, transfer, and retention (Cannon-Bowers, Rhodenizer, Salas, & Bowers, 1998). Four principles for effective training strategies require that they: 1) present relevant information and concepts, 2) demonstrate Knowledge, Skills and Abilities (KSA) to be learned, 3) create opportunities for trainees to practice the skills, and 4) provide feedback to trainees regarding practice (Salas & Cannon-Bowers, 2001).

There has been a great deal of research into the training of search skills in tasks such as radiographic interpretation (Nodine, Krupinski, & Kundel, 1990; Nodine et al., 2002), airframe inspection (Sadasivan, Greenstein, Gramopadhye, & Duchowski, 2005), circuit board inspection (Nalanagula, Greenstein, & Gramopadhye, 2006), and airport baggage screening (Fiore et al., 2004). Training interventions such as presentation of expert scan (Nalanagula et al., 2006; Sadasivan et al., 2005), metacognitive strategies (Chapman, Underwood, & Roberts, 2002; Nodine et al., 2002) and attentional weighting strategies (Hagemann, Strauss, & Canal-Bruland, 2006; Williams, Ward, Knowles, & Smeeton, 2002) have proven successful in improving search performance.

Despite these extensive findings in the training science literature, few perceptual skills training strategies validated in the literature have reached training practitioners in the field. Common practice in current Marine Corps observation training is to provide feedback consisting of Knowledge of Correct Response (KCR) in the form of pointing out all threats in the scene/scenarios or Knowledge of Results (KR) in the form of pointing out all threats missed. This form of purely outcome feedback has been shown successful in some domains (e.g., teacher in service training; (Leach & Conto, 1999); however, in other domains outcome feedback may not be at a granular enough level to facilitate trainees identifying and improving process level skills which show performance decrements (Davis, Carson, Ammeter, & Treadway, 2005; Goodman, Wood, & Hendrickx, 2004). For example, in an inspection task, when compared to outcome level feedback, process level feedback has been shown to facilitate detection of more targets and development of a more systematic search strategy (Chabukswar et al., 2003).

One problem with outcome level feedback is that it typically gives little guidance on how performance needs to change in order to increase performance levels. Such is the case with visual search. KCR (where all threats were located) or KR (where missed threats were located) has an element of "where" the trainee should be looking inherent in the feedback as trainees are made aware not only of the threat missed, but the location of the threat as well. However, this type of feedback does not address high priority areas in which no threats were located or the sequence in which the trainee should be searching, both process level performance aspects which may be in need of remediation for performance to improve. Additionally, this feedback does not support the trainee in abstracting these locations to a higher level of search strategy, allowing them to apply this strategy in novel situations and environments.

Challenges to Training Visual Search

Many perceptual skills training strategies from the training science literature have not reached the field because of the multiple challenges associated with training perceptual skills such as search. The primary challenge is the ability to adequately measure and diagnose search performance to facilitate process level feedback. In order to be able to debrief at a subtask/process level, it is necessary to be able to distinctly measure a sub process (i.e., search) and effectively discriminate performance on separate sub processes (e.g., search and detection). For instance, in the threat detection task, a performer may not have indicated a threat because either a) the performer did not utilize effective search strategies and hence did not search the area in which the threat was located (search error), or b) the performer effectively searched, but did not detect the cues in the environment indicative of a threat (detection error), or c) the performer effectively searched, detected the indicator cues, however, did not recognize these cues as a threat (recognition error). Diagnosis at this level determines where in the perceptual process the process level error which led to an outcome error occurred and can facilitate feedback to target this specific process. In current military threat detection training practice, measurement and diagnosis is difficult as trainee to instructor ratios are typically high, placing high workload demands on instructors during performance monitoring and assessment stages, and perhaps more importantly, it is extremely difficult to monitor and diagnose the perceptual process of search as it is a subtle process inaccessible to instructor observation.

A second challenge is in obtaining the ability to demonstrate search skills. Perceptual skills such as search and detection are subtle or internal processes which are unobservable. As a result, many training strategies consisting of demonstration are infeasible. As a result, search training typically consists of verbal description of skill performance, not actual demonstration. One difficulty with this approach, however, is that research suggests that expert recall of certain motor or perceptual events is incomplete, and potentially erroneous (Cleeremans & McClelland, 1991). Expert performers often operate on “autopilot” and are not always aware of the cues to which they are attending (Klein, 1998). It may be more appropriate (i.e., more efficient and accurate) to incorporate effective training strategies which allow experts to demonstrate their expertise rather than try to verbalize them (Sidani & Gonzalez, 1994).

Addressing the Challenges

A promising solution to the above challenges is the use of eye tracking technology to measure visual performance. With the advancements in eye tracking technology, information about a person’s visual attention, once inaccessible, is becoming more attainable. Visual attention can provide important insights to the information used in task performance, such as the importance of various features or cues (Raab & Johnson, 2007). Several studies (Jarodzka et al., 2009; Mello-Thoms et al., 2008; Raab & Johnson, 2007; White, Hutson, & Hutchinson, 1997) have used eye tracking to extract information about scan strategies. These studies have demonstrated that eye tracking can aid in the assessment of perception through measurement of visual attention during observation via gaze, scan path, and fixation data. These metrics can identify which Areas of Interest (AOI) were gazed upon and the amount of time the AOIs were gazed upon.

With respect to the challenge of measuring and diagnosing search performance, researchers studying radiographic interpretation, have used eye tracking data to diagnose where in the perceptual-cognitive process (i.e., search, recognition, decision) errors occur. Nodine et al. (2002) make distinctions in the classification of errors in misdiagnosis in radiographic interpretation based on fixation duration, where lack of fixation is interpreted as a searching error, fixation for less than 1000 milliseconds and lack of indication of abnormality is interpreted as a recognition error, and fixation for greater than 1000 milliseconds and failure to indicate as abnormality is a decision making error. Such a method could be extended to the threat detection domain. If effective fixation duration thresholds could be established for discriminating the different perceptual processes (i.e., search, detection, recognition), this method could be used in the diagnosis of process level threat detection errors and discrimination of search verses detection errors. This would facilitate the process level feedback needed to effectively target the perceptual root cause of performance deficiencies.

With respect to the challenge of utilizing training strategies which consist of demonstration, eye tracking technology provides the ability to capture and present expert search demonstrations. What remains is the need to determine how to effectively present this granular data in order to facilitate effective training of all aspects of search skills. Few studies have attempted to target multiple aspects of search strategies such as search location and sequence. One consideration in targeting search location is that there are two unique aspects of "location" which warrant consideration. The first is the perceptual aspect of the location or what the location visually looks like, which allows identification of a specific area. For clear cut areas like windows and doors which have very distinct and familiar appearances, a verbal description may suffice to allow trainee identification of these areas. However, for more elusive and intangible areas like "shadows" (i.e., shaded areas which provide concealment) or negative space (i.e., areas in which there is nothing to draw attention, but may prove effective for concealing threats), it may be necessary to provide visual representation to allow trainee recognition. The second aspect of the location is the conceptual aspect of why the area is a target area (i.e., high priority area which should be searched), an understanding of which is necessary to abstract specific locations (e.g., shadows) to a higher level of categorization (e.g., a place which provides concealment), which would facilitate generalization to new environments and situations.

There is an opportunity to leverage training strategies from the training science literature which have been proven successful in improving search skills and extend them to address all critical aspects of search performance. For example, feed forward training of expert scan patterns have been tested in multiple domains including airframe inspection and electronic circuit board inspection and have shown great success (Nalanagula et al., 2006; Sadasivan et al., 2005). This strategy is promising as it could be used to target both location (perceptual) and sequence aspects of search. This strategy could be enhanced to also target the conceptual aspect of location (i.e., why it must be searched) by layering on auditory elaborative feedback describing why locations should be searched to support trainees in extracting higher level scan strategies. This feed-forward strategy could be extended into a feedback strategy and tailored to the specific errors a trainee has committed by also including elements of the trainee scan path to highlight differences between expert and trainee, a technique which has yet to be developed or empirically tested.

In summary, it is necessary to first be able to effectively measure search performance and discriminate search from other perceptual processes such as detection and second to develop process level feedback which 1) provides more specific feedback aimed at the threat detection sub process of visual search, 2) incorporates feedback strategies shown to be most effective in facilitating performance improvements on this skill, including both location and sequence aspects of search, and 3) provides support for the development of generalized search strategies. This feedback could be further enhanced by tailoring it to trainees' specific performance decrements, to better target areas in which the trainee is in need of improvement. The result would be tailored feedback, based on process level performance diagnosis, aimed to target all critical aspects of search, which is hypothesized to provide powerful training results.

Experimental Hypotheses

This line of research will explore the effectiveness of utilizing eye tracking to address the challenges of training visual search discussed above by empirically evaluating:

1. If search performance can be effectively measured and discriminated from other perceptual processes (i.e., detection) utilizing eye tracking
2. If newly developed search feedback strategies increase training value over traditional feedback strategies used in the military domain

Through a series of two experiments, two main hypotheses will be tested. First, it is hypothesized that search performance metrics which utilize eye tracking and behavioral data can effectively and reliably measure the search component of threat detection and discriminate between errors made in these two consecutive stages of perceptual performance (i.e., search and detect). Second, it is hypothesized that new search feedback strategies which utilize eye tracking to allow demonstration of search skills will result in greater performance gains over traditional feedback strategies used in the field. Specifically, it is hypothesized that elements of expert and trainee scan path data will significantly improve training performance over traditional KR feedback, both individually and when used jointly, with the additive effect of joint elements resulting in the greatest performance improvements.

The following section will first summarize the literature pertaining to visual search theory and practice, including characteristics of visual search performance, factors that influence visual search, and training techniques to support visual search.

CHAPTER TWO: LITERATURE REVIEW

Visual Search

Visual search, the task of finding a target among distracters (Verghese, 2001), is a ubiquitous skill used both in work and daily life. From searching for an exit sign in a shopping center to searching for an aircraft on a radar screen, visual search is required to function in almost any environment. Given the varied applications of this skill, visual search tasks have a broad range of characteristics on which they can vary. Visual search can be systematic, in which few fixation locations are repeated, or random, in which there is no memory for previous locations (Arani, Karwan, & Drury, 1984). Visual search tasks may require eye movement (such as in scanning a three dimensional scene), or may not include eye movement (such as in searching for a particular shape in a small array of letters). Visual search tasks can vary in difficulty from search among several distracters of similar appearance to search among few distracters of distinctly different appearance from the target. This broad range of characteristics and the performance differences associated are precisely what make visual search such a complex skill. For decades, researchers have been working towards decomposition of this skill to understand both the underlying process as well as the biological mechanisms driving human performance of visual search. The following sections detail theories of visual search and how these theories and findings from the training science literature can be leveraged to effectively train visual search.

Visual Search Theory

At a very basic level, there are generally considered two types of visual search: preattentive and attentive. A great deal of research has been conducted regarding the distinction between these two types of visual search and the various factors that affect performance during each (Shiffrin & Schneider, 1977; Treisman, Vieira, & Hayes, 1992). Pre-attentive search has been referred to under several names including efficient search, parallel search, easy or effortless search (Leonards et al., 2002) and automatic detection (Treisman & Souther, 1985). In this type of search, the targets are proposed to contain elementary features which are processed preattentively (Sireteanu & Rettenbach, 2000), essentially drawing attention to themselves. In this type of search, the target "pops out" at the observer, requiring little conscious effort. Such is the case when looking for the letter T on a page of letter Os or for a red baseball cap in a sea of blue baseball caps. Attentive search, also referred to as inefficient search (Leonards et al., 2002), serial search (Treisman & Souther, 1985), and controlled search (Shiffrin & Schneider, 1977), refers to effortful search in which attention must be serially allocated to objects in the environment to detect a target.

The distinction between these two types of search is consistent with the view that search is driven by both bottom up and top down processes (Itti & Koch, 2001), wherein bottom up processes drive attention due to salient features in an environment or stimulus (e.g., salient target features) and top down processes drive attention through the application of search strategies such as the direction of attention to locations of high priority.

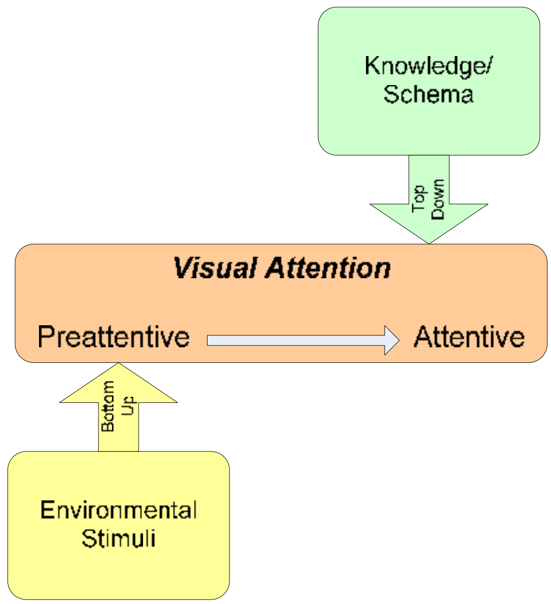


Figure 1. Drivers of Visual Attention

Many search tasks are comprised of both types of search. In fact, McCarley, Kramer, Wickens, Vidoni, & Boot (2004) discuss a multi-stage model of visual search wherein the visual search component can be broken down into an early "orientation" stage in which an efficient or pre-attentive search occurs resulting in detection of very salient targets. Following this is an attentive or inefficient search stage in which potential target locations are scanned via a succession of fixations in an attempt to detect less conspicuous targets. This stage requires the observer to scan the appropriate regions in the scene as driven by appropriate search strategies.

Two stage models of visual search such as this have been the cornerstone of many visual search theories. Treisman's Feature Integration Theory (FIT; Treisman & Gelade,

1980) is based on a two stage model wherein visual features (e.g., color, orientation) are processed and organized first into feature maps and then saliency maps that serve to guide attention to conspicuous objects in a scene. FIT proposes that features are extracted in a pre-attentive stage across an entire scene and these features are then bound into objects once attention is directed to this location in an attentive stage. From this theory, Treisman proposed a more practical description of the two types of search: feature search (similar to pre-attentive search) and conjunctive search (similar to attentive search) wherein feature search is based on one unique target feature and therefore does not require attention as the target "pops out", and conjunctive search is based on more than one feature requiring attention to bind these features together to detect a target (Treisman & Souther, 1985).

The Guided Search model, another two stage model, also proposes that in the efficient stage a quick snapshot of the scene is taken to build a saliency map, then in the second stage a detailed search based on specific features is performed (Wolfe, 1998). However, in the Guided Search Model, the preattentive stage reduces objects in a scene to their component features which serve to activate feature detectors associated with the target, effectively guiding search in the inefficient stage starting with objects most highly activated (Hoffman, 1999). Despite differences, the implications of both of these models is that deployment of attention in the inefficient stage can be influenced by explicit top down information provided with respect to search strategies (both what features to look for and how/where to look).

Interestingly, there has been research which also points to the ability to influence search through implicit top down knowledge. Chun and Jiang (1999) explored contextual cueing and found that perceptual schemas specify how objects covary in the world around us, and that such a schema can be created through target/context covariation, resulting in implicit influence on attention deployment in search. Chun and Jiang found that past target locations produced faster search as opposed to novel target locations, however, participants could not explicitly state these locations. This implicit learning, however, was robust only when the information is selectively attended and relevant and predictive (i.e., of target location). These findings are incredibly promising with respect to training visual search as implicit operations have been shown to be durable over time, robust across interference and exhibit high capacity (Chun & Jiang, 1999).

Much of the visual search theory, however, is based on experimentation performed on very simple laboratory tasks. It is necessary to consider visual search in an operational setting to ensure theoretical foundations in visual search theory align with the visual search in practice

Visual Search Theory in Practice

The two stage model presented above, is extremely relevant when referring to search within the military threat detection domain. The two stage model is consistent with the what/where distinction commonly referred to in visual processing (Levine, 2000). The preattentive stage identifies where to direct attention and the attentive stage subsequently identifies the object being attended. This maps very closely to the two types of search taught in US Marine Corps training: hasty or general search and detailed search *(MCI 03.35c Infantry Patrolling*, 1996). A hasty search requires a quick scan of the environment to identify high threat areas, which should “pop out”. Once these have been identified, a more detailed search of these areas ensues, utilizing prescribed search strategies (e.g., searching near to far in 180 degree arcs).

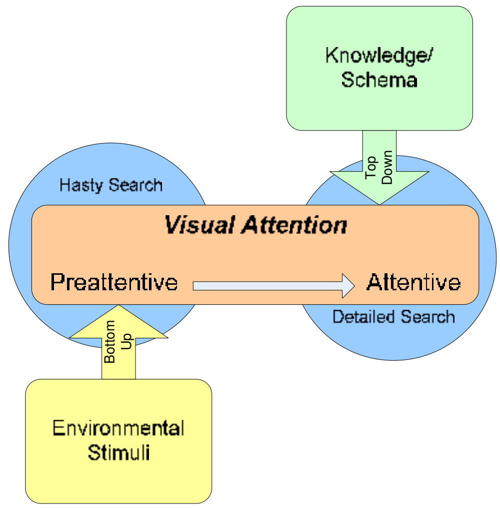


Figure 2. Visual Search Theory in Practice

What Influences Visual Search?

The above theory can serve as a foundation for understanding how best to influence visual search through training. There are many factors that affect preattentive search, the most influential of which is salience. Itti & Koch (2001) propose a model of bottom up pre-attentive search based on saliency and include such environmental features as color, intensity, orientation and motion. Pre-attentive search has been proposed by researchers to be difficult to alter, ignore or suppress (Shiffrin & Schneider, 1977), therefore may not prove fruitful to target with training intervention. However, there is research to suggest otherwise. For instance, after extensive practice, military observers have shown efficient detection of targets in cluttered scenes that novice observers search for inefficiently, suggesting that they learned, through practice, to preattentively process the target (Doll & Home, 2001). Leonards et al. (2002) also found that practice of inefficient search which included either unique features or differences in brightness distributions led to efficient search. Also in support is the large body of research surrounding perceptual learning to increase sensitivity to stimulus and stimulus features. Massed practice or exposure to stimulus at or below just noticeable difference thresholds has shown to increase sensitivity for processes such as contrast sensitivity (Sowden, Rose, & Davies, 2002), motion direction discrimination (Burns, Nettelbeck, McPherson, & Stankov, 2007), and figure ground segregation (Yi, Olson, & Chun, 2006) as well as more complex stimulus such as x-ray images (Sowden, Davies, & Roling, 2000). Such a strategy might render preattentive search a feasible option to target with training.

Attentive search is widely hypothesized to be influenced by several factors, the most notable of which is the top down visual search schema and strategy being employed. Additional factors include the number of relevant conjunctive features in a target or distracter objects (Wolfe, Cave, & Franzel, 1989), target type and search field size (Wang et al., 1997) and environmental conditions such as clutter (Doll & Home, 2001). Another factor which affects inefficient search performance is known as the edge effect, a phenomenon in which observers scan paths tend to avoid the edges of a display (Parasuraman, Boff, Kaufman, & Thomas, 1986). Researchers have found evidence of this effect by looking at target detection in search as a function of eccentricity (Parkhurst, Law, & Niebur, 2002).

It is believed that the attentive search stage provides the best opportunity for influencing visual search through training. Additionally, given the nature of the target task (i.e., military threat detection) including complexity of the search environment, wide field of view, large variety of distracters, and covert nature of the enemy, the attentive stage of search is likely the most critical for successful threat detection. As visual search strategies have been shown to be easily influenced through training (Chabukswar et al., 2003; Chapman et al., 2002; Gramopadhye, Drury, & Sharit, 1997; Underwood, 2007; Wang et al., 1997) the focus of the research herein will be the molding of visual search through the training of search strategy.

Visual Search Strategies

There are several types of search strategies which can be employed ranging from a very structured systematic search in a regular pattern (e.g., alternate up and down scan moving from left to right; (Wang et al., 1997), to less structured strategies such as searching by area or searching by component/object (Chabukswar et al., 2003) to attentional weighting strategies which focus search on cues/areas deemed most critical (Rezec & Dobkins, 2004). For military threat detection, a task analysis (Carroll et al., 2007) revealed that several of these search strategies are employed by subject matter experts. The first is similar to the attentional weighting strategy and prescribes where to search by identifying high priority areas where there is a high potential for threat (e.g., areas of concealment). The second strategy, a less structured and systematic search prescribes how to search, requiring high priority areas to be searched before low priority areas, always searching the environment near to far and right to left and allocating the majority of time to high priority areas (80/20, high priority/low priority). Lastly, a more structured systematic search prescribes searching from right to left in 50 meter arcs, temporarily halting this pattern for emerging environmental elements, then turning back to it. As presented above, the most common mistakes made by trainees during search for military threats include: 1) failure to judge which areas are high priority, and 2) failure to scan in a systematic sequence (e.g., no smooth patterns from right to left, not searching closest areas first). Based on these, the research herein will target a portion of the search strategies discussed above, specifically: 1) Where to search (high vs. low priority areas) and 2) Sequence of search (systematic sequence encompassing high then low priority, near to far, left to right/right to left). The below section explores options for how to train these elements of visual search.

<u>Training Techniques to Support Visual Search</u>

A common misconception in training design is that practice equals training; however, effective training of any task relies on the integration of effective training strategies (Cannon-Bowers et al., 1998). The training science literature was examined to identify visual search training strategies which have proven successful in past research and application. These strategies were then evaluated for relevancy to the military threat detection task and those most relevant served as the basis for development of innovative visual search training strategies to target military threat detection.

Training Strategies

McCarly et al. (2004) performed a study in which the effect of practice on visual search performance in a simulated airport baggage screening task was examined. Results indicated that practice alone did not improve the effectiveness of visual search, and that increased performance was due entirely to sensitivity increases with respect to detection of the target. These findings highlight the need to provide feedback that can help observers improve search strategies. There has been a great deal of research surrounding the training of search skills across a wide variety of domains and targeting an array of search strategies (Chabukswar et al., 2003; Gramopadhye et al., 1997; Salas & Cannon-Bowers, 2001; Wang et al., 1997). The visual search training strategies reviewed primarily fall into six categories: 1) Performance Feedback, 2) Process Feedback, 3) Attentional Weighting, 4) Difficulty Variation, 5) Metacognitive Strategies, and 6) Expert Performance Models. Of these, process level feedback, attentional weighting, metacognitive strategies and expert performance models show the most promise.

Performance Feedback

Using a circuit board inspection task, Wang et al. (1997) examined whether search strategy was trainable, and whether systematic, natural or random search strategies led to better defect detection. To train systematic search, subjects were instructed to move their eyes in a regular pattern across the board alternating up and down from left to right, with specific fixation positions defined based on the size of the circuit board. Trainees were then given 24 boards to scan as practice and were given feedback with respect to whether they had followed instructions or not (i.e., knowledge of results). Similar treatment for random search included instructions to follow no pattern and 24 practice trials with knowledge of results feedback. Findings indicated that practice with knowledge of results performance feedback could significantly change search strategy, both for the better with systematic search and for the worse with random search.

Knowledge of results performance feedback is the feedback method currently used in the field and is thus the current standard.

Process Feedback

Chabukswar et al. (2003) explored the effects of online process plus performance feedback compared to performance (outcome) feedback for a visual circuit board inspection task. Process feedback included both statistical (e.g., percent of area covered) and graphical (i.e., graphical representation of area covered) and performance feedback included items such as number of defaults detected and time to detect. The process plus performance group not only detected more defaults, but seemed to develop a more systematic search strategy than the performance feedback group. Gramopahye et al., (1997) examined the effectiveness of performance feedback and cognitive feedback, both statistical and graphical, in improving search performance for an airframe inspection task. Cognitive feedback consisted of information regarding areas which the trainee had already searched, represented statistically by percentages or graphically via scan pattern indicated with shaded markers on the airframe. Although the performance feedback group scored as high on the performance measures as the graphical cognitive feedback group, the graphical group showed the best combined response in performance and strategy. As the statistical cognitive feedback group actually performed worse after training, Gramopahye et al. suggest that the visual feedback in the graphic display served as the "bridge between cognitive data and action" (p. 342).

Nodine et al. (1990) developed a training strategy for radiographic interpretation in which scanned areas with dwell times greater than 1000 milliseconds and for which no indication of lesions occurred were interpreted as detection without recognition and were fed back for reanalysis resulting in increased recognition of initial misses. Similarly, Nodine, Kundel, Mello-Thoms, and Weinstein (2001) utilized a Computer Assisted Visual Search (CAVS) strategy which fed back regions of interest that received prolonged visual dwell (greater than or equal to 1000 ms) and highlighted them on the display so the regions could be re-evaluated and found improved detection performance. Such strategies are based on findings such as those reported in Nodine et al. (2002) that approximately 70% of lesions that are not reported in mammogram reading attract visual attention, as measured by the amount of visual dwell in the location of the lesion, implying that such misses are covert negative decisions.

Attentional Weighting

Attentional weighting, focuses on targeting aspects of important cues to attend. Exogenous orienting or highlighting is a technique which has been used to train the use of attentional weighting search strategies. Hagemann, Strauss, and Cañal-Bruland (2006) found that highlighting relevant cues such as areas of the trunk, arm and racket at the critical times during badminton training led to significant increase in test performance. Williams, Ward, Knowles, & Smeeton (2002) used a freeze frame and slow motion video playback to highlight critical cues to attend in anticipating the direction of tennis strokes. Critical cues were derived from expert performance data extracted via eye tracking. Williams et al. found significant performance improvements which also transferred to subsequent field exercises resulted from instruction which included 1) explicit instruction of critical cues and their associated outcome, and 2) guided discovery with the use of verbal probes encouraging trainees to look at a certain area of the body and draw conclusions about the relation of cues to outcomes. Crowley, Medvedeva, and Jukic

(2003) developed a perceptual intelligent tutoring system for pathology diagnosis which incorporated a training strategy which includes "visual hints" by moving a viewer position to an area of interest, highlighting critical features and providing textual information about the type of feature. The researchers found this method led to improved diagnosis performance.

Difficulty Variation

Research suggests that level of difficulty during training can affect the development of effective search strategies by facilitating development of a more generalizable search strategy. Doane, Alderton, Sohn, & Pellegrino (1996) explored the effect of discrimination difficulty in a simple polygon discrimination task and found that more difficult stimulus tasks to discriminate leads to development of more effective and more global search strategy and hence better transfer than with easier stimulus tasks to discriminate. Schmidt and Bjork (1992) found supporting evidence for this as well and suggested that training which maximized performance during training may not support transfer or generalizability to operational performance enhancements. Operational performance enhancements may be better facilitated by more challenging and diverse training conditions that result in degraded speed and accuracy during skill acquisition (Schmidt & Bjork, 1992). Schmidt & Bjork found that relative to standard practice conditions, three practice conditions, namely random practice, infrequent or faded feedback and variation in practice, slowed the rate of improvement during training resulting in lower training performance at the end of practice, but resulted in enhanced post training performance.

Metacognitive Strategies

Chapman et al. (2002) developed a training intervention aimed at improving visual search associated with the driving task, which incorporated both elements of metacognition and expert performance data. Through a series of five training modules which utilized videos, trainees would 1) practice visual search for driving hazards, indicate hazards, commentate on what they are looking at in their search, 2) explore in slow motion areas indicated as hazards, commentate on why they are hazards and listen to an expert commentary on why these areas are hazards, 3) practice visual search while being prompted during pauses to indicate what just happened or what happened next, 4) re-explore in full speed motion areas indicated as hazards and commentate on why they are hazards and 5), practice visual search for driving hazards, indicate hazards, commentate on what they are looking at in their search. Results showed not only significant immediate effects on visual search, but some of these effects remained for 3-6 months and many of them transferred to an actual driving task. Nodine et al. (2002) present a strategy which focuses on how long to dwell on suspect areas or cues and consideration of confidence in threat. Nodine et al. (2002) found that in a mammography diagnosis task that prolonged dwell-time on a potential lesion did not notably increase the number of lesions discovered and did increase the error rate for lesion detection. These researchers suggest mentor-guided feedback with instructions to trust only the more confident and early decisions and to quit searching when unsure to improve detection performance in search task.

Expert Performance Models

Expert search performance as illustrated through expert scan paths has been used to train visual search in multiple domains. Sadasivan, et al. (2005) examined the effect of a feedforward of expert scan pattern training strategy in an airframe inspection task. In this study, expert scan paths as well as indications of fixation duration as collected via an eye tracker were presented as a static overlay to an airframe image to trainees. This training strategy resulted in 30% greater performance improvements in defect detection accuracy than a practice condition. Similar research (Mehta, Sadasivan, Greenstein, Gramopadhye, & Duchowski, 2005) compared the training effectiveness of different types of expert feedforward training strategies and found that presentation of expert scan data with a decaying trace (fixations remaining on the screen for a brief period of time before disappearing) resulted in a mean gain in number of defects detected after training five times greater than a practice condition. Additionally, Nalanagula et al. (2006), examined the effect of a similar feedforward strategy for a circuit board inspection task with an additional comparison of static, dynamic and hybrid (static + dynamic) expert scan feedforward strategies. The results indicated that feedforward of expert scan paths improved circuit board defect detection by 26% higher levels than those without. Results also indicated that dynamic or hybrid display techniques, which include the development of the expert's search pattern in 'real' time (i.e., showing the scan unfold to illustrate, not only the pattern, but the chronological and temporal components of the scan) are better suited as feedforward training displays than static displays which only display snapshot representations of scan paths and regions of interest.

Eccles, Walsh, & Ingledew (2006) performed a study examining expert-novice differences in visual attention allocation in an orienteering task in which significant differences were found (Eccles et al., 2006). Based on these findings, the researchers propose presenting expert models of attention allocation to show trainees how to allocate attention properly to relevant cues, including verbal guidance on when and how to allocate visual attention to relevant cues in the environment.

The success of expert feedforward training strategies is not surprising given findings in research on expertise. Data from naturalistic studies suggests that experts have an ability to rapidly recognize critical cues, evaluate a situation and determine an appropriate plan of action, a phenomenon called "recognition-primed decision making" (Klein, 1993). Novices and intermediates consistently fail to "see" the same critical cues for the identical situation. If novices could be trained to utilize search strategies similar to experts, development of expertise may be accelerated. This becomes quite feasible with the availability of eye tracking technology. Multiple studies (e.g., Raab & Johnson, 2007; White et al., 1997) have used eye tracking to extract information about expert scan strategies. Eye tracking can aid in the assessment of perception through measurement of visual attention during observation via gaze, scan path, and fixation data. The metrics can identify what Areas of Interest (AOI) were gazed upon and the amount of time the AOIs were gazed upon to drive the development of feedback.

Several of these training strategies lend themselves to training search strategy. Presentation of expert scan paths provides strong support for influencing trainee scan strategy, including both location to search and sequence of search. Metacognitive strategies incorporating trainee scan paths could allow trainees to explore their own performance and reflect on areas in which their performance differed from intentions or planned strategies. Attentional weighting strategies which utilize highlighting can be leveraged to direct trainee attention to high priority areas or areas in which their performance differed from expected. Additionally, attentional weighting strategies which provide background information about features could be used to elaborate on why an area should be searched, targeting the conceptual aspect of search. Process feedback, both graphical and statistical, could be used to provide trainees information with regards to how to allocate their attention both spatially and temporally.

Tailored Feedback

The strategies identified above could be made more effective by tailoring them to specific performance decrements. The tailoring or individualizing of feedback to address specific trainee performance decrements has the ability to positively impact performance. Providing trainees with information relevant to what performance areas are in need of improvement allows them to focus on these areas during future performance. Bloom (1984) defined what is referred to as the “2 Sigma Problem” in which trainees who received one-on-one instruction or tutoring perform two standard deviations above those receiving traditional classroom (i.e., group) training. Bloom believed that through the tutoring process (i.e., one-on-one instruction) that all students have the potential to reach these levels. Bloom described this traditional tutoring process as one which provides constant feedback facilitating a corrective process between the tutor and the trainee. In order to mirror this method, training scientists have responded with the tailoring of feedback to target specific performance decrement.

Tailoring of feedback can happen at two levels: 1) tailor to the type of skill, and 2) tailor to a specific error. With respect to the first level, different feedback strategies can more effectively target certain types of knowledge or skills as different types of learning tasks require different instructional strategies and methods (Mory, 2004). Utilizing a feedback strategy which can most effectively impact the target skill allows feedback to be tailored to the specific skill decrement. For instance, it may be most effective to incorporate a strategy such as those discussed above which have been shown effective in improving visual search. With respect to the second level, as the main function of feedback lies in the correction of errors (Mory, 2004), tailoring the feedback to specific errors allows trainees a better understanding of deficiencies in their performance and how to improve upon these. Incorporating elements of trainee search in comparison to expert search would allow the feedback to be tailored to specific trainee decrements.

As a result, the above strategies could potentially be combined to provide a powerful training solution which supports trainees in extracting information from both expert scan paths and their own scan paths to guide improvements in search strategy. This level of tailored feedback, however, demands error analysis (i.e., what type of error was made); therefore a necessary component of tailored feedback is performance measurement and diagnosis.

Performance Measurement and Diagnosis

In order to provide effective feedback, it is necessary to capture relevant performance to facilitate the diagnosis of performance deficiencies. Brannick and Prince

(1997) describe performance measurement as an investment in which one purchases information to inform decisions or actions. In the case of training, these decisions/actions are how to provide training remediation. There are multiple levels of performance measures which can be collected including outcome and process measures. Outcome measures provide information regarding the overall outcome of performance (e.g., mission performance) indicating successful or not (Eddy, 1998). Process measures provide a more granular level of measurement which facilitates monitoring of processes which contributed to outcome performance (Eddy, 1998). These provide a richer set of data by which performance can be assessed and assist in identification of specific deficiencies or performance breakdowns.

Multiple aspects of performance can be measured to facilitate performance assessment. Behaviors can be measured to determine task steps which a trainee is performing, including time and accuracy of steps. Behavioral measures allow direct measure of procedural skills (e.g., Are the correct steps being performed, in the right order?). Communications are often measured to determine what information is being exchanged, between who and in what format (Smith-Jentsch, Zeisig et al., 1998). These also can be used to assess procedural skills as well as team coordination skills (e.g., Are the trainees sharing the correct information with the right teammates?).

Based on performance measures, performance diagnosis can then facilitate tailored feedback. Performance diagnosis is the analysis of performance measures to provide a consolidated view of performance and performance errors and facilitates identification of the underlying causes of performance outcomes and deficient processes to allow instructors to provide meaningful feedback to correct these deficiencies (Salas, Rosen, Burke, Nicholson, & Howse, 2007). As errors are considered to be valuable opportunities to clarify misunderstandings in learners (Mory, 2004) this diagnosis serves as the cornerstone of tailored feedback. The following sections describe methods for measuring and diagnosing visual search performance.

Measuring Visual Search Performance

While skills such as procedural and team coordination skills are relatively accessible for measurement due to their outward nature, this is not the case with measurement of perceptual skills such as visual search. Task performance components which are purely perceptual are not easily measured as perceptual processes are internal processes. Scanning behaviors, a manifestation of visual attention (Itti & Koch, 2001; Treisman & Souther, 1985) can be measured at a very gross level based on observable head movement, however, it is not possible to assess at a granular level what objects trainees are visually attending. With respect to detection, this is a purely psychophysical process which is inaccessible to observers. Overall outcome of the perceptual process (e.g., Was a target indicated as a threat?) is accessible and inference of process level performance results can be made. However, ability to measure performance at a granular level has not been possible in the past as instructors/researchers are not able to assess the perceptual state (i.e., they only have access to the action resulting from the combination of these processes). This level of assessment does not provide instructors the granularity necessary to build an accurate picture of the perceptual processes which took place during performance.

With increasing advancements in measurement technology, however, this is becoming more feasible. For example, eye tracking offers researchers an additional set of metrics to aid in measuring performance of perceptual skills. While easily observable actions and communications provide important information regarding performance (e.g., trainee did not engage threat), they often do not provide the level of data necessary to diagnose why certain performance decrements occur (e.g., he did not engage the target because 1) he did not search effectively, 2) he searched effectively, but did not detect the threat, etc.). Subtle physical behaviors such as scanning patterns or internal perceptual processes such as detection that would reveal these answers are not currently accessible via behavioral metrics. Researchers have started using eye tracking to measure perceptual processes such as visual attention in a driving task (Underwood, Chapman, Brocklehurst, Underwood, & Crundall, 2003) and visual search in a mammogram diagnosis task (Mello-Thoms, Nodine, & Kundel, 2002). Wang, Chignell, and Ishizuka (2006) used eye tracking in an Intelligent Tutoring System (ITS) to monitor users attention and interests to personalize agent behaviors. Jodlowski and Doane (2004) utilized eye tracking in development of a model of pilot action planning during simulated flight for intelligent tutoring, which uses eye tracking to model user knowledge based on which flight instruments the user fixates. Such advances provide invaluable data in understanding how performance unfolded both with respect to scan path and fixation durations. Although eye tracking facilitates the measurement of perceptual performance at a very granular level, it is necessary to transform this detailed data into meaningful and actionable performance diagnoses. The following section discusses diagnosis methods for visual search.

Diagnosing Visual Search Performance

The diagnosis of errors within procedural knowledge and skills performance have proven successful through event-based methods such as event based knowledge elicitation (Fowlkes, Salas, Baker, Cannon-Bowers, & Stout, 2000) and event-based approach to training (Fowlkes, Dwyer, Oser, & Salas, 1998) in which Targeted Acceptable Responses to Generated Events or Tasks (TARGETs) are used to elicit expected knowledge or procedural responses. For these competencies, such diagnosis methods equate to measurement of behavioral actions and communications to verify if all steps in the procedures were followed and which steps were omitted or performed incorrectly. Given adequate performance measurement design, this type of diagnosis can be made fairly accurately. Other knowledge and skill types have proven more of a challenge for performance diagnosis. For example, diagnosis of specific breakdowns within team performance has proven challenging as it is difficult to separate mutual performance monitoring, backup and feedback (Smith-Jentsch, Johnston, Payne, Cannon-Bowers, & Salas, 1998). Raters found it challenging to evaluate whether team members were monitoring one another's activities unless someone on the team provided feedback or exhibited backup behavior (Smith-Jentsch, Johnston et al., 1998).

With respect to the threat detection task, diagnosis has previously been limited by the limitations of performance measurement tools. But as discussed above, with the emergence of eye tracking technology, information about a person's perceptual state, once inaccessible, is becoming more available. Several researchers have utilized eye tracking to diagnose perceptual performance deficiencies. For example, researchers studying radiographic interpretation have used eye tracking data to diagnose where in the perceptual process (i.e., search, recognition, decision) errors occur. Nodine et al. (2002) make distinctions in the classification of errors in misdiagnosis in radiographic interpretation based on fixation duration, where lack of fixation is interpreted as a searching error, fixation for less than 1000 milliseconds and lack of indication of abnormality is interpreted as a recognition error, and fixation for greater than 1000 milliseconds and failure to indicate as abnormality is a decision making error. A visual dwell time of 1000 milliseconds is equated to detection as it is considered a significant allocation of visual attention (Nodine et al., 2002). Manning & Ethell (2002) used a similar method to classify whether a radiographic interpretation error was due to lack of detection or recognition. In this study missed lesions were dwelt on for an average of half the time of detected nodules but within a time thought to be acceptable for detection to occur (900 milliseconds); hence, the researchers interpretation was that observers were making recognition errors although they detected the nodules (Manning & Ethell, 2002). The researchers' reasoning was that detection of a specific feature may occur but the decision that it is an abnormality (recognition) depends on higher order cognitive processes (Manning, Leach, & Bunting, 2000). Errors in detection and decision making were determined based on fixation duration as it was seen as indicative of depth of information processing of image. Similarly, Nodine, Krupinski, and Kundel (1990) tested the hypotheses that long durations indicated detection but not necessarily recognition of perturbations in chest images. This hypothesis was supported by the found effectiveness of their detection algorithm which detected true nodules solely by localizing them on the basis of the observers' long gaze durations.

Interestingly, similar methods have been used to decompose the mental rotation task into separate processes of search, transformation and confirmation stages based on fixation duration (Just & Carpenter, 1976). The researchers propose that eye fixations can "reveal the sequence of mental operations" during a mental rotation task (Just & Carpenter, 1976, p. 459), which is precisely the goal of using such data in perceptual performance diagnosis. These diagnosis methods could be extended to the military threat detection domain, however, research is needed to identify effective fixation duration thresholds for discriminating between perceptual processes. Additionally, it is necessary to evaluate the validity and reliability of such diagnosis methods for a military threat detection task.

Unique Contributions of Present Study

This study has multiple unique contributions to the scientific community. This is the first study to systematically and empirically evaluate the effectiveness of fixation duration-based metrics in diagnosing process level aspects of target detection. Although several researchers to data have used these types of process level metrics (Manning & Ethell, 2002; Mello-Thoms, Dunn, Nodine, Kundel, & Weinstein, 2002; Nodine et al., 2002), none have performed a systematic evaluation of the validity and reliability of these metrics based on actual observer scan data. For instance, in Nodine et al. (2002) the determination of 1000 milliseconds as the threshold for a fixation duration-based metric to discriminate between missed chest nodules due to recognition or decision making errors was established based on previous research. Specifically, it was based on research conducted by Hillstrom and Logan (1998) which explored visual search skills on a simple laboratory conjunctive search task. The 1000 millisecond fixation duration was considered to be a significant allocation of visual attention as this was inline with typical response times to detect a target of interest in this conjunctive search task. Mello-Thoms et al. (2002) also used this threshold based on the same study as it showed that the mean response time for searching and identifying a target in a conjunction of features ranges from 800-1000 milliseconds. Not only did this threshold come from a research paradigm using a simple laboratory task, but the times were based on response times, not fixation times as eye tracking was not employed (Hillstrom & Logan, 1998). This present study will use previous research as a foundation, but derive the fixation duration-based metrics from scan data collected during the study and evaluate the validity and reliability of the metrics with data collected during the study.

This study is also the first to empirically examine the presentation of trainee scan patterns and comparison of trainee to exert scan patterns as a training intervention. Although multiple studies have explored the use of expert scan as a training intervention (Mehta et al., 2005; Nalanagula et al., 2006; Sadasivan et al., 2005), none to date have examined the effects of presentation of trainee scan data on search performance or target detection performance. This is likely due to the technology required to facilitate such feedback. In order to be able to present a training intervention which incorporates trainee scan data, researchers must have access to technology which not only allows the collection of trainee scan data, but also facilitates the near real time presentation of this scan data over the associated stimulus. Additionally, in order to ensure learning it is necessary to present the trainee scan data in a manner that allows trainees to extract meaningful information from the scan. Currently, there is no commercial-off-the-shelf technology that has this capability. This study was part of a larger effort funded by the Office of Naval Research in which training technology was being developed. As such, the capability to effectively present trainee scan data as a training intervention and allow trainees to compare their scan data to expert scan data was developed. As a result, the present study was able to explore this training intervention and the ability for it to impact a trainee's search strategy.

The results of this study will provide new theoretical contributions to the training science community and evidence to support or contradict current theories related to search skills, including performance measurement and diagnosis and the impact of training interventions. Additionally, the results will have generalizable implications for military threat detection training, specifically on the effectiveness of using eye tracking in search performance assessment and feedback. The following chapter will detail the experiments conducted in the present study.

CHAPTER THREE: PRESENT STUDY

Based on the above reviewed literature two key methods were developed which aimed to improve search skills. First, a diagnosis method for deducing where in the perceptual process breakdowns occur (i.e., search, detect, recognize) based on eye tracking data was developed. Second, an innovative feedback strategy for improving search strategies through the presentation of elements of expert and trainee scan data was developed. The present study aimed to examine the effectiveness of these methods in training military search skills by empirically evaluating:

1. If search performance can be effectively measured and discriminated from other perceptual processes (i.e., detection) utilizing eye tracking
2. If newly developed search feedback strategies add training value over traditional feedback strategies used in the military domain

Experiment 1

The goal of Experiment 1 was to validate that the search metrics in the diagnosis method developed can effectively measure search skills and reliably differentiate between search and detection errors. This method, developed based on the diagnosis method used by Nodine et al. (2002), aims to diagnose where in the perceptual process a breakdown occurs during military threat detection based on fixation durations. Specifically when a target is missed (i.e., not indicated as a threat), based on whether there was a fixation on the target and how long that fixation was, the perceptual root cause of the error is diagnosed as being a search error, a detection error or a recognition error. In order for this diagnosis method to be successful, it was necessary to identify an effective threshold for diagnosing between detection and recognition errors (the threshold between search and detection is simply 0 milliseconds, i.e., whether the target was fixated or not).

The threshold proposed for the above radiography examination task seemed inadequate given the different nature of the task (much more ill-defined targets). Research surrounding an Intelligence Imagery Analyst task seemed more applicable to the military threat detection task. In evaluating the ability to use eye tracking to improve Imagery Analyst's accuracy, Hale et al. (2007) found significantly greater average fixation durations associated with missed targets than non-targets. As fixations on non-targets are assumed to be scanned areas in which detection did not occur, given the significant difference in fixation duration, some fixations on missed targets are assumed to be areas in which detection occurred, however recognition did not. This data suggests that a reasonable threshold for detection within this domain and potentially similar domains is 300 milliseconds. Although a threshold for the military detection task was to be derived from data collected within this experiment, this initial threshold served as a point of reference which was used to guide experimentation.

The perceptual performance diagnosis method evaluated is illustrated in Table 1 below. This is hereafter referred to as target-based diagnosis as it is performed based on a target being missed. Specifically, if a target was missed and was not fixated on, a search error has occurred. If a target was missed and fixated for less than the determined fixation duration threshold, a detection error has occurred. If a target was missed and fixated for greater than the determined fixation duration threshold, a recognition error has occurred.

Table 1. Target-based Diagnosis Method (adapted from (Nodine et al., 2002))

		Lack of Behavioral (Key press) Target Indication	
		Target Absent	**Target Present**
Fixation Duration	**Zero ms**	Correct Answer	Search Error
	Short Duration*	Correct Answer	Detection Error
	Long Duration*	Correct Answer	Recognition Error

* Threshold between short/long duration will be set based on data collected within this study

There was an opportunity to enrich this diagnosis method. While this method was sufficient for discerning why a target was not indicated as a threat (i.e., whether it due to a search, detection or recognition error), it was not comprehensive in assessing search skills beyond target-specific performance. For instance, if there were certain high priority areas that should be included in a search, trainees may have searched the small portion of these areas in which targets were located, however, may have neglected many of these areas which do not contain targets in that specific instance. Although search strategies utilized were insufficient, the target-based diagnosis method would indicate good search performance as all targets were attended. The errors associated with non-target areas would have gone undiagnosed.

Therefore it was also important to include a higher level diagnosis of search errors in which “where” the trainee attended as indicated by scan path is compared against where they should have attended, providing a more global measure of search performance. Therefore the target-based diagnosis method was extended by adding a trial-based component to the evaluation of search performance, wherein search performance levels are diagnosed independent of targets. This trial-based method is illustrated in Table 2. Specifically, search performance is measured by percentage of high priority areas scanned.

Table 2. Trial-based Diagnosis Method

	Metric Description	
Metric	**Behavioral**	**Fixation Data**
Search Performance	NA	% of high priority areas scanned
Detection Error	**Target NOT Indicated as Threat**	Fixated on Target for Short Duration*
Recognition Error	**Target NOT Indicated as Threat**	Fixated on Target for Long Duration*

* Threshold between short/long duration will be set based on data collected within this study

Given the extension of the Nodine et al. (2002) diagnosis to the military threat detection domain and the extension of the method from target-based to trial-based, it was necessary to evaluate this method to validate that it can effectively diagnose where breakdowns in perceptual performance occur and that it is predictive of outcome performance. The focus of the first experiment in this effort was 1) to validate that the target-based performance diagnosis metrics could discriminate between search, detection and recognition errors in the military threat detection domain and 2) to validate that the newly developed trial-based performance diagnosis metrics are able to measure process level aspects of performance that are predictive of outcome performance (i.e., search and detection).

To accomplish this, it was necessary to ensure both search and detection errors occurred. A task analysis revealed several factors which contribute to the occurrence of search and detection errors. For search, these parameters include large number of high priority areas in the scene which demand attention, large amounts of clutter and the presence of distracters. As each of these increases, difficulty to search the area increases and search errors will result. For detection, parameters which contribute to errors include the levels of occlusion of a threat (e.g., 90% of a sniper rifle will be occluded in a good sniper hide) and orientation of the threat (e.g., rifle pointed straight at you is more difficult than one visible in a canonical view), both of which affect the visual size of the threat. By varying these parameters and creating opportunities for search and detection breakdowns, the ability of the diagnosis method to effectively discriminate between search and detection errors could be examined.

Diagnosis Metrics

The first metrics are the target-based metrics and were designed to determine the perceptual root cause of the error (i.e., whether a search error, a detection error or a recognition error) when a target is missed. The target-based metrics are calculated based on fixation durations on the target as follows:

Table 3. Components of Target-based Diagnosis Method

Error Type	Fixation Duration on Target
Search Error	0 ms
Detection Error	0ms < X < 300ms
Recognition Error	300ms</=X

*300ms detection threshold determined based on data collected in this study

The second set of metrics are trial-based metrics and were based on the target-based metrics but were designed to describe overall search and detection performance and trial outcome performance. The trial-based metrics are calculated as follows:

Table 4. Components of Trial-based Diagnosis Method

Metric Name	Trial (Scenario) Based Performance Metrics
Search Performance	Percentage of high priority areas that were fixated
Detection Errors	Number of targets fixated on but not indicated as threats
Outcome Performance	Percentage of targets indicated as threats via a key stroke & mouse click (Hit Rate).

Hypothesis 1

It is hypothesized (Hypothesis 1) that the above target-based metric can effectively (reliably and validly) diagnose where in the perceptual process (i.e., search, detect, recognize) breakdowns occur using eye tracking and behavioral data.

Prediction 1.1

It is predicted that there will be a significant difference between average fixation durations associated with non-threat fixations and threat fixations not indicated as threats (misses) as threat fixations will have instances in which detection occurs, resulting in a significant allocation of visual attention. This difference will serve as the basis for the threat detection threshold.

Prediction 1.2

It is also predicted that for misses (i.e., trials in which the participant did not indicate a threat or incorrectly indicated a threat), there will be either instances of 0 millisecond fixations on this threat (search errors), short fixations (detection errors) or long fixations (recognition errors). However, it is anticipated there will be fewer missed targets with long fixations (recognition errors) as there was only one type of threat (Dragunov sniper rifle) for which detection might prove challenging, but once detected, misinterpretation as a non-target is unlikely.

Hypothesis 2

It is also hypothesized (Hypothesis 2) that the above trial-based metrics can effectively (reliably and validly) discriminate between search and detection performance using eye tracking and behavioral data, by demonstrating:

- Reliable discrimination between search and detection performance
- Prediction of mission Outcome Performance as indicated by Hit Rate

Prediction 2.1

It is predicted that Search Performance (% of high threat areas attended) and Detection Errors (number of threats which were attended for short period of time but not indicated as threats) will both be significantly correlated with Outcome Performance (as indicated by Hit Rate), both accounting for a significant amount of independent variance.

Method

Participants

Prior to recruiting any participants and collecting data, a power analysis was performed to estimate the number of participants needed to obtain sufficient statistical power. A power analysis was performed using the software G Power 3.0 (Faul, Erdfelder, Lang, & Buchner, 2007). Using a small to moderate effect size of .3, for a repeated measures within group ANOVA with 1 group and 40 repetitions, the software recommended a total sample size of 9 participants. In addition, the same power analysis was performed using the effect size from a study which explored different fixation durations associated with threat detection in an intelligence analyst search task (Hale et al., 2007). Using this effect size of .223 resulted in a recommended sample size of 15.

Twenty participants took part in this experiment. They consisted of nine males and eleven female participants and ranged from 18 to 31 (mean=22.3, s.d.=4.2) years of age. Most participants were student volunteers from a Southeastern University and were recruited through a University system and given either extra credit or monetary compensation for their participation. Additionally, students and non students were recruited via flyers.

Task

Participants performed the experimental task using screenshots from the ObSERVE (Observational Skills Enhancement and Retainment Virtual Environment) desktop simulation testbed, a Delta 3D static Virtual Environment (VE). The screenshots were presented via a personal computer including a Hewlett Packard Compaq dc5800 Microtower desktop computer with Intel Core 2 Duo with 2.66GHz processors, 4GB RAM, GeForce 9800 GTX+ video card, a Dell 19 inch High Definition flatscreen monitor, and conventional keyboard and mouse. Participant eye movement data was collected via an easyGaze^{TM} eye tracker, a noninvasive desk mounted eye tracker located in front of the flat screen display. Performance data was collected from the eye tracker and the desktop computer via a performance assessment system which calculated behavioral and eye tracking metrics.

Figure 3. Experimental Apparatus

Participants performed a static military threat detection task. The task involved trainees being stationed at a virtual Observation Post (OP) in a Middle Eastern urban environment. In the virtual OP environment, participants were stationed either in a building looking out of a window, on a rooftop or in an alleyway. Participants were given 45 seconds to search the environment from this static location, by scanning the display. Participants could respond before the 45 second limit, however, after 45 seconds, their view of the environment was obstructed and they were prompted to indicate if there was a threat or not in the environment by clicking the Y button to indicate there was a threat and the N button to indicate no threat. Upon participant indication of threat present, the scenario view was again presented and they were then asked to point out the location of the threat via a mouse click. They were given approximately 10 seconds to do so. Upon mouse click to indicate location of threat or N button to indicate no threat, the trainee was advanced to next scenario. Prior to performance, participants received pre-training, in which they were taught scanning strategies consisting of what areas are high priority and low priority as well as general rules for sequence of scan (e.g., high priority first, near to far). Within the pre-training, participants were also given training on threats they were searching for, specifically the appearance of the Dragunov sniper rifle at various levels of occlusion and various orientations. Participants then performed 40 scenarios in which there was either 0 or 1 threat present.

Experimental Design

Experiment 1 was a within subjects repeated measures design. All participants performed 40 trials, 20 of which had targets present. To ensure ample opportunity for breakdowns in search and detection performance, search and detection difficulty were manipulated. Half of the 40 trials had easy levels of search difficulty and half had difficult levels of search difficulty. Of the 20 trials with targets, half of the trials had easy detection levels and half had difficult detection levels. Difficulty breakdowns by trial are illustrated in Table 5 and operationalization of search difficulty and detection difficulty are presented in Tables 6 and 7, respectively.

Table 5. Experiment 1 Difficulty Levels by Trial

		Trials (40)		
		Target Present (20)		Target Not Present (20)
		Detection Difficulty		
		Easy	Diff	
Search Difficulty	Easy	5	5	10
	Diff	5	5	10

Table 6. Operationalization of Search Difficulty

		# High Priority Areas	
		0 – 30 Areas	Over 30Areas
Amount of Clutter	Low (0-30 pieces)	E	D
	High (Over 30 pieces)	D	D

Table 7. Operationalization of Detection Difficulty

		Orientation	
		0 (Canonical) -45 degrees	46 degrees - 90 degrees
Occlusion	Low (0-50%%)	E	D
	High (51-99%)	D	D

Dependent Variables

The following Dependent Variables (DVs) were measured:

Table 8. Dependent Variables

- Behavioral indication of threat: Y/N
- Location (and object associated) of mouse click to indicate threat
- Fixation Locations
- Fixation Durations
    - Threat Fixations
    - Non Threat Fixations
- Search Performance = % High priority areas fixated
- Search Errors = Number of threats not fixated and not indicated as threat
- Detection Errors = Number of threats fixated < 300ms but not indicated as threat
- Recognition Errors = Number of threats fixated >/=300ms but not indicated as threat
- Outcome Performance = % of threats correctly indicated (hit rate)
- Time to detect threat (Reaction time)

Procedure

Upon arrival, participants completed an informed consent form and a series of questionnaires and tests listed below in Table 9.

Table 9. Questionnaires

1. Demographics
2. Visual Acuity
3. Color Blindness
4. Spatial ability
    a. Spatial Orientation
    b. Visualization
    c. Hidden Figures
5. Cognitive Load
6. Visual/Verbal Learning Style

Participants then received pre-training targeting both search and detection knowledge and skills necessary to perform the task. Pre-training was presented via a power point presentation in which screen shots from the simulation were utilized to present target and scene examples. Participants then performed the eye tracker calibration process and began experimental trial performance. Participants performed the threat detection task over a series of 40 trials with no feedback given during training. Participants were then debriefed on the study and their participation.

Results

The analyses reported below were performed in SPSS 11.5 for windows, and all alpha levels were set to .05, unless otherwise specified. Due to anticipated eye tracking data loss caused by excessive participant movement as well as poor candidacy of participants due to causes unknown, all data was screened to identify cases in which significant eye tracking data was missing. To identify these cases, eye tracking based metrics were graphed using histograms to identify outliers. Outlier data sets were further analyzed to determine if there was significant eye tracking data loss. Those cases for which there was none or little eye tracking data for more than two trials were dropped. This data screening process resulted in the exclusion of one of the 20 participant cases which had significant amounts of eye tracking data loss, resulting in 19 participants involved in analysis reported below.

Individual Differences

To assess individual difference effects on performance, the aptitude and demographic variables presented in Table 10 were correlated with Outcome Performance, Detection Errors, Search Performance and Reaction Time.

Table 10. Aptitude and Demographic Variables

Aptitude
- Visual Acuity
- Color blindness
- Spatial Orientation
- Visualization
- Hidden Figures Ability
- Visual Verbal Learning Style
- Cognitive Load

Demographic
- Gender
- Age
- Highest Level of Education
- SAT Score
- Gaming Experience
- Military Training Experience
- Hunting Experience
- Danger level of neighborhood growing up

Table 11 presents the correlations of aptitude variables with performance and reaction time variables and Table 12 presents the correlations of demographic variables with performance and reaction time variables.

Table 11. Aptitude Variable Correlations with Performance and Reaction Time

Aptitude	Outcome Performance	Detection Errors	Search Performance	Reaction Time
Visual Acuity	R = .21 Sig = .35 N = 22	R = -.10 Sig = .70 N = 17	R = .17 Sig = .46 N = 22	R = .17 Sig = .45 N = 22
Color blindness	R = .21 Sig = .36 N = 22	R = .00 Sig = 1.0 N = 17	R = -.09 Sig = .71 N = 22	R = -.23 Sig = .30 N = 22
Spatial Orientation	R = .36 Sig = .10 N = 22	R = -.61** Sig = .01 N = 17	R = -.16 Sig = .48 N = 22	R = -.07 Sig = .77 N = 22
Visualization	R = .47* Sig = .03 N = 22	R = -.54* Sig = .03 N = 17	R = -.20 Sig = .38 N = 22	R = -.08 Sig = .73 N = 22
Hidden Figures Ability	R = .36 Sig = .10 N = 22	R = -.26 Sig = .31 N = 17	R = .23 Sig = .31 N = 22	R = .15 Sig = .49 N = 22
Visual Verbal Learning Style	R = -.26 Sig = .32 N = 22	R = -.32 Sig = .23 N = 17	R = -.06 Sig = .80 N = 22	R = -.07 Sig = .77 N = 22
Cognitive Load	R = -.10 Sig = .66 N = 22	R = -.26 Sig = .32 N = 17	R = .30 Sig = .18 N = 22	R = .62** Sig = .00 N = 22

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

Table 12. Demographic Variable Correlations with Performance and Reaction Time

Demographics	Outcome Performance	Detection Errors	Search Performance	Reaction Time
Gender	R = -.28 Sig = .20 N = 22	R =.21 Sig = .43 N = 17	R = -.10 Sig = .67 N = 22	R = -.11 Sig = .64 N = 22
Age	R = -.38 Sig = .08 N = 22	R = .41 Sig = .11 N = 17	R = .31 Sig = .16 N = 22	R = .22 Sig = .32 N = 22
Highest Level of Education	R = -.44* Sig = .04 N = 22	R = .27 Sig = .29 N = 17	R = .02 Sig = .93 N = 22	R = -.09 Sig = .68 N = 22
SAT Score	R = -.03 Sig = .93 N = 14	R = .31 Sig = .39 N = 10	R = -.21 Sig = .47 N = 14	R = -.24 Sig = .40 N = 14
Gaming Experience	R = .02 Sig = .92 N = 22	R = -.06 Sig = .83 N = 17	R = -.31 Sig = .16 N = 22	R =- .34 Sig = .12 N = 22
Military Training Experience	R = -.05 Sig = .84 N = 22	R = .33 Sig = .20 N = 17	R = .63** Sig = .00 N = 22	R = .37 Sig = .09 N = 22
Hunting Experience	R = .38 Sig = .08 N = 22	R = -.45 Sig = .07 N = 17	R = .15 Sig = .51 N = 22	R = .14 Sig = .52 N = 22
Danger level of neighborhood growing up	R = -.14 Sig = .54 N = 22	R = -.26 Sig = .32 N = 17	R = -.28 Sig = .20 N = 22	R = -.26 Sig = .24 N = 22

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

No consistent patterns of individual difference effects on performance emerged, although the spatial ability aptitude of Spatial Orientation (SO) was highly and significantly correlated with number of Detection Errors and the spatial ability aptitude of Visualization (VZ) was moderately and significantly correlated with Outcome Performance and number of Detection Errors. Based on these findings, SO and VZ were tested as covariates for the following analyses; however, as neither were statistically significant covariates, they were excluded in the analyses.

Analysis 1.1: Average Fixation Durations

Analysis 1.1 was performed to identify a fixation duration threshold which represents the difference between 1) a missed target fixated, however, not for a significant amount of time indicating the participant did not detect the target (detection error) and 2) a missed target fixated for a significant amount of time indicating a level of detection (recognition error). To identify this threshold, average fixation durations were calculated for the following types of Area of Interest (AOI) categories included in Table 13.

Table 13. AOI Categories

- Non-target fixations in trials without targets (AOI Type 1)
- Non-target fixations in trials with missed targets that were not fixated (AOI Type 2)
- Non-target fixations in trials with missed targets that were fixated (AOI Type 3)
- Target fixations in trials with missed targets that were fixated (AOI Type 4)
- Non-target fixations in trials with targets that were found (Hits; AOI Type 5)
- Target fixations in trials with targets that were found (Hits; AOI Type 6)

The average fixation durations and standard deviations in milliseconds for each of these AOI categories are summarized in Table 14 below.

Table 14. Average Fixation Durations

	Trials with:			
	No Target	Missed Target Not Fixated	Missed Target Fixated	Hit Target
Target Avg Fix Duration (ms)	NA	NA	212 (42)[4]	294 (56)[6]
Non-Target Avg Fix Duration (ms)	204 (11)[1]	206 (18)[2]	207 (19)[3]	241 (22)[5]

* AOI Type Number

A repeated measures ANOVA was performed between average fixation durations associated with these AOI categories. The analysis revealed that type of AOI had a significant effect ($F(5, 90) = 27.85$, $p<.01$, partial $\eta^2 = .61$.) on average fixation duration. Post hoc tests were performed using Tukey's LSD to determine where significant differences lie. Post hoc analyses revealed significant differences between average fixation durations of AOI types 1 and 5 above ($p<.01$), 1 and 6 ($p<.01$), 2 and 5 ($p<.01$), 2 and 6 ($p<.01$), 3 and 5 ($p<.01$), 3 and 6 ($p<.01$), 4 and 6 ($p<.01$), and 5 and 6 ($p<.01$).

Given the lack of significant difference between AOI Type 3 and 4 average fixation durations as predicted in prediction 1.1, the results of the above analysis prevent determination of a fixation duration threshold between detection and recognition errors. As such, multiple fixation durations thresholds were evaluated in analysis 1.2 and 2.1 to determine an effective threshold.

Analysis 1.2: Missed Target Eye Tracking Response

Analysis 1.2 was performed to further analyze missed targets based on eye tracking data. Using behavioral metrics of whether a target was indicated or not (Yes/No response) and once indicated if the target was correctly located (via a mouse click), responses were categorized into Hit (H), Miss (M), False Alarm (FA), and Correct Rejection (CR) as illustrated in Table 15.

Table 15. Target Behavioral Response Categorization

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="3">Said Response</th></tr>
<tr><th colspan="2">Yes</th><th>No</th></tr>
<tr><th rowspan="2">Target Present</th><th>Yes</th><td>H
(Correct Location)</td><td>M
(Incorrect Location)</td><td>M</td></tr>
<tr><th>No</th><td colspan="2">FA</td><td>CR</td></tr>
</table>

Based on the four detection thresholds identified above (200ms, 250ms, 300ms, 350ms), misses were further analyzed to determine if the miss was due to a search error, a detection error, or a recognition error based on fixation duration as follows:

Table 16. Missed Target Eye Tracking Response Categorization

	Fixation Duration
Search Error	0 ms
Detection Error	0ms < X < detection threshold (ms)
Recognition Error	Detection threshold (ms) </= X

A total of 162 misses occurred across all participants over the 40 scenarios. Table 17 summarizes the breakdown (number and percentage) of the errors which fall into each of the three error categories across the 4 detection thresholds under evaluation.

Table 17. Missed Target Eye Tracking Response Statistics (Number (Percentage))

	200Threshold	**250Threshold**	**300Threshold**	**350Threshold**
Search Error	121 (74%)	121 (74%)	121 (74%)	121 (74%)
Detection Error	14 (9%)	19 (12%)	27 (17%)	32 (20%)
Recognition Error	27 (17%)	22 (14%)	14 (9%)	9 (6%)

Analysis 2.1: Sequential Multiple Regression

Analysis 2.1 was performed with two goals. The first goal was to evaluate if the Search Performance and Detection Error metrics were meaningful and unique metrics which were predictive of Outcome Performance. The second goal was to determine the optimal fixation duration detection threshold (fixation duration on a target over which a level of detection is assumed) by comparing regression models created based on an array of different thresholds. Four Sequential Multiple Regression analyses were performed with Search Performance and Detection Errors as predictors and Outcome Performance as criterion. Detection Errors varied between these four analyses based on the four different fixation duration thresholds selected based on findings from analysis 1.1: 200ms, 250ms, 300ms and 350 ms.

200ms Threshold

A sequential multiple regression was performed with predictors of Search Performance and Detection Errors calculated using a 200 millisecond fixation duration threshold and Outcome Performance as a criterion. R was not significantly different from zero at the end of each step. After the second step with all Independent Variables (IVs) in the equation, $R = .52$, $F(2, 16)=2.97$, $p>.05$. After step 1, with Detection Errors in the equation, $R^2 =.20$, $F_{inc}(1, 17)=4.25$, $p>.05$. Detection performance when detection threshold is set to 200 did not contribute to a significant amount of Outcome Performance variance. After step 2, with Search Performance added to Detection Errors to predict Outcome Performance, $R^2=.27$, (adjusted $R^2=.18$), $F_{inc}(1, 16)=1.55$, $p>.05$.

The addition of search performance did not significantly improve the amount of variance accounted for. Results are presented in Table 18.

Table 18. Sequential Multiple Regression for 200ms Detection Threshold

Source	**SS**	**df**	**MS**	**F**	**Sig**
Regression	.10	2	.05	2.97	.08
Residual	.26	16	.02		
Total	.35	18			
Predictor	**Coeff**	**SE**	**t**	**Sig**	
Constant	.45	.15	2.97	.01	
Detection Errors	-0.08	.04	-2.32	.03	
Search Performance	.004	.003	1.25	.23	
R^2 = .27	**R^2_{adj} = .18**				

250ms Threshold

A sequential multiple regression was performed with predictors of Search Performance and Detection Errors calculated using a 250 millisecond fixation duration threshold and Outcome Performance as a criterion. R was not significantly different from zero at the end of each step. After the second step with all IVs in the equation, R = .63, F (2, 16)=5.30, p<.05. After step 1, with Detection Errors in the equation, R^2 =.35, F_{inc} (1, 17)=9.08, p<.01. After step 2, with Search Performance added to Detection Errors to predict Outcome Performance, R^2=.40, (adjusted R^2=.32), F_{inc} (1, 16)=1.34, p>.05. The addition of search performance did not significantly improve the amount of variance accounted for by detection errors alone. Results are presented in Table 19.

Table 19. Sequential Multiple Regression for 250ms Detection Threshold

Source	SS	df	MS	F	Sig
Regression	.14	2	.07	5.29	.02
Residual	.21	16	.013		
Total	.35	18			
Predictor	**Coeff**	**SE**	**t**	**Sig**	
Constant	.51	.14	3.72	.00	
Detection Errors	-0.09	.03	-3.15	.01	
Search Performance	.003	.003	1.16	.27	
R^2 = .40	**R^2_{adj} = .32**				

300ms Threshold

A sequential multiple regression was performed with predictors of Search Performance and Detection Errors calculated using a 300 millisecond fixation duration threshold and Outcome Performance as a criterion. R was significantly different from zero at the end of each step. After the second step with all IVs in the equation, $R = .81$, $F(2, 16)=15.63$, $p<.01$. After step 1, with Detection Errors in the equation, $R^2 =.51$, $F_{inc}(1, 17)=17.81$, $p<.01$. After step 2, with Search Performance added to Detection Errors to predict Outcome Performance, $R^2=.66$, (adjusted $R^2=.62$), $F_{inc}(1, 16)=7.08$, $p<.05$, significantly improving the amount of variance accounted for. Results are presented in Table 20.

Table 20. Sequential Multiple Regression for 300ms Detection Threshold

Source	**SS**	**df**	**MS**	**F**	**Sig**
Regression	.23	2	.12	15.63	.00
Residual	.12	16	.01		
Total	.35	18			
Predictor	**Coeff**	**SE**	**t**	**Sig**	
Constant	.43	.10	4.21	.00	
Detection Errors	-.09	.02	-5.48	.00	
Search Performance	.01	.002	2.66	.02	
R^2 = .66	**R^2_{adj} = .62**				

350ms Threshold

A sequential multiple regression was performed with predictors of Search Performance and Detection Errors calculated using a 350 millisecond fixation duration threshold and Outcome Performance as a criterion. R was significantly different from zero at the end of each step. After the second step with all IVs in the equation, R = .81, F (2, 16)=14.97, p<.01. After step 1, with Detection Errors in the equation, R^2 =.54, F_{inc} (1, 17)=19.59, p<.01. After step 2, with Search Performance added to Detection Errors to predict Outcome Performance, R^2=.65, (adjusted R^2=.61), F_{inc} (1, 16)=5.34, p<.05, significantly improving the amount of variance accounted for. Results are presented in Table 21.

Table 21. Sequential Multiple Regression for 325ms Detection Threshold

Source	**SS**	**df**	**MS**	**F**	**Sig**
Regression	.23	2	.12	14.96	.00
Residual	.12	16	.01		
Total	.35	18			
Predictor	**Coeff**	**SE**	**t**	**Sig**	
Constant	.47	.10	4.50	.00	
Detection Errors	-.08	.02	-5.36	.00	
Search Performance	.01	.002	2.31	.03	
$R^2 = .65$	**$R^2_{adj} = .61$**				

Results revealed that the 300 millisecond threshold resulted in the model which accounts for the greatest amount of Outcome Performance variance. Intercorrelations of performance variables based on this 300 millisecond threshold and reaction time are presented in Table 22.

Table 22. Intercorrelation of Performance and Reaction Time Variables (N=22)

	1	**2**	**3**	**4**
1. Outcome Performance	1			
2. Detection Errors	-.67**	1		
3. Search Performance	.09	.37	1	
4. Reaction Time	.06	.33	.84**	1
Mean	.57	1.55	49.07	25.70
Standard Deviation	.13	1.34	10.33	7.46

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

Additional Analyses

Additional analyses were performed to explore the effects of search difficulty and detection difficulty on Outcome Performance, Detection Errors, Search Performance and Reaction Time.

Detection Difficulty and Search Difficulty Effects

Participant scores on these four metrics for the twenty trials which contained a target were aggregated (averaged) for each individual participant across trials for each of the following 4 categories: Easy Detect Easy Search (EDES), Easy Detect Difficult Search (EDDS), Difficult Detect Easy Search (DDES), Difficult Detect Difficult Search (DDDS). Four repeated measures ANOVAs were performed with IVs of search difficulty and detection difficulty and DVs of Outcome Performance, Detection Errors, Search Performance and Reaction Time. A MANOVA was not used due to lack of high correlations between all dependent variables.

Outcome Performance

A repeated measures ANOVA was performed on Outcome Performance (hit rate) for the four difficulty categories mentioned above. The analysis revealed that detection difficulty had a significant effect ($F(1, 18) = 238.77$, $p<.01$, partial $\eta^2=.93$) on Outcome Performance with performance decreasing with increasing detection difficulty. The analysis also revealed that search difficulty had a significant effect ($F(1, 18) = 9.15$, $p<.01$, partial $\eta^2=.34$) on Outcome Performance with performance levels increasing with increasing levels of search difficulty. There was also a significant interaction ($F(1, 18) =$

5.73, $p<.05$, partial $\eta^2=.24$). The means and standard deviations are presented in Table 23 and the data are illustrated in Figure 4.

Table 23. Detection Difficulty and Search Difficulty Effects on Outcome Performance

	Search Difficulty	
Detection Difficulty	**Easy**	**Difficult**
Easy	mean = .78 s.d.= .15	mean =.81 s.d.= .16
Difficult	mean = .24 s.d.= .17	mean = .41 s.d.= .24

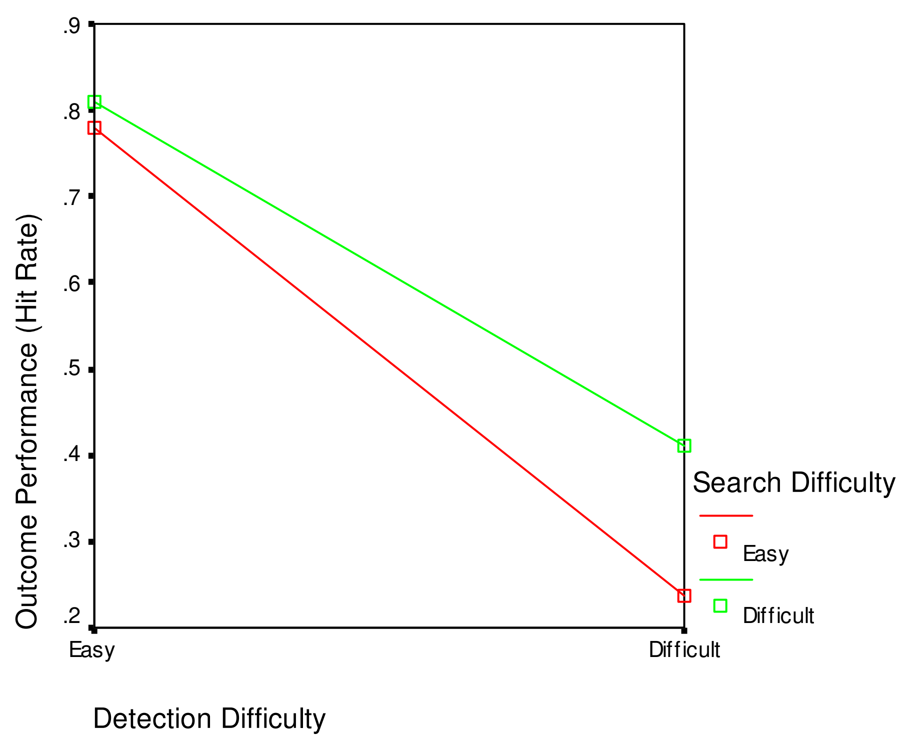


Figure 4. Detection Difficulty and Search Difficulty Effects on Outcome Performance

Detection Errors

A repeated measures ANOVA was performed on Detection Errors for the four difficulty categories mentioned above. The analysis revealed that detection difficulty had a significant effect ($F(1, 18) = 16.88$, $p<.01$, partial $\eta^2=.48$) on Detection Errors with errors increasing with increasing detection difficulty. The analysis revealed that search difficulty had no significant effect on Detection Errors ($p>.05$) and there was no significant interaction between detection difficulty and search difficulty ($p>.05$). The means and standard deviations are presented in Table 24 and the data are illustrated in Figure 5.

Table 24. Detection Difficulty and Search Difficulty Effects on Detection Errors

	Search Difficulty	
Detection Difficulty	**Easy**	**Difficult**
Easy	mean = .02 s.d.= .06	mean = .02 s.d.= .06
Difficult	mean = .11 s.d. = .12	mean = .15 s.d. = .19

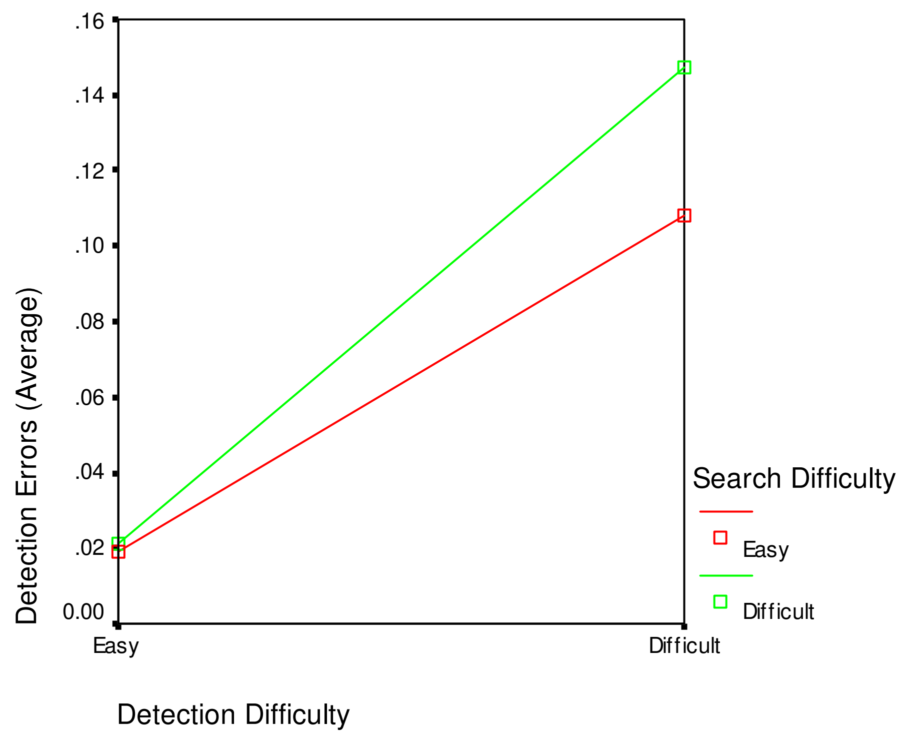


Figure 5. Detection Difficulty and Search Difficulty Effects on Detection Errors

Search Performance

A repeated measures ANOVA was performed on Search Performance for the four difficulty categories mentioned above. The analysis revealed that detection difficulty had a significant effect ($F(1, 18) = 32.49$, $p<.01$, partial $\eta^2=.64$) on Search Performance with performance increasing with increasing detection difficulty. The analysis also revealed a significant effect of search difficulty ($F(1, 18) = 5.25$, $p<.05$, partial $\eta^2=.23$) on Search Performance, with performance increasing as search difficulty increases. There was also a significant interaction ($F(1, 18) = 5.40$, $p<.05$, partial $\eta^2=.23$). The means and standard deviations are presented in Table 25 and the data are illustrated in Figure 6.

Table 25. Detection Difficulty and Search Difficulty Effects on Search Performance

	Search Difficulty	
Detection Difficulty	**Easy**	**Difficult**
Easy	mean = 34.96 s.d.= 10.92	mean = 35.05 s.d.= 10.51
Difficult	mean = 41.40 s.d.= 8.74	mean = 48.96 s.d.= 10.81

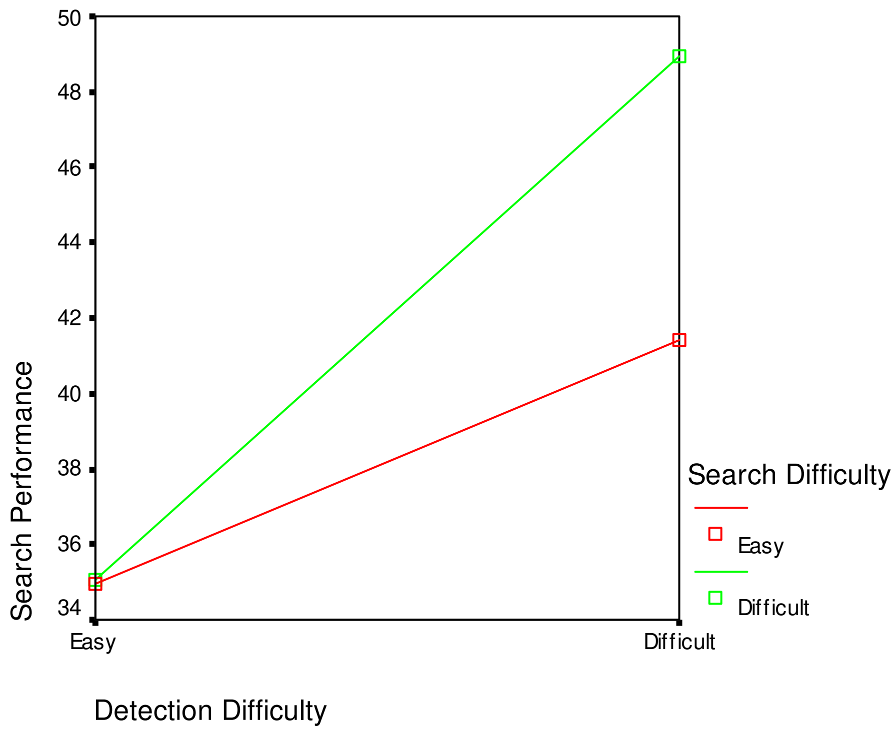


Figure 6. Detection Difficulty and Search Difficulty Effects on Search Performance

Reaction Time

A repeated measures ANOVA was performed on Reaction Time for the four difficulty categories mentioned above. The analysis revealed that detection difficulty had a significant effect ($F(1, 18) = 68.49$, $p<.01$, partial $\eta^2=.79$) on Reaction Time with

Reaction Time increasing with increasing detection difficulty. The analysis revealed no significant effects of search difficulty on Reaction Time (p>.05) and there was a not a significant interaction (p>.05). The means and standard deviations are presented in Table 26 and the data are illustrated in Figure 7.

Table 26. Detection Difficulty and Search Difficulty Effects on Reaction Time

	Search Difficulty	
Detection Difficulty	**Easy**	**Difficult**
Easy	mean = 13.46 sec s.d.= 5.83 sec	mean =14.00 sec s.d.= 7.48 sec
Difficult	mean = 24.56 sec s.d. = 7.50sec	mean = 23.71 sec s.d. = 6.39 sec

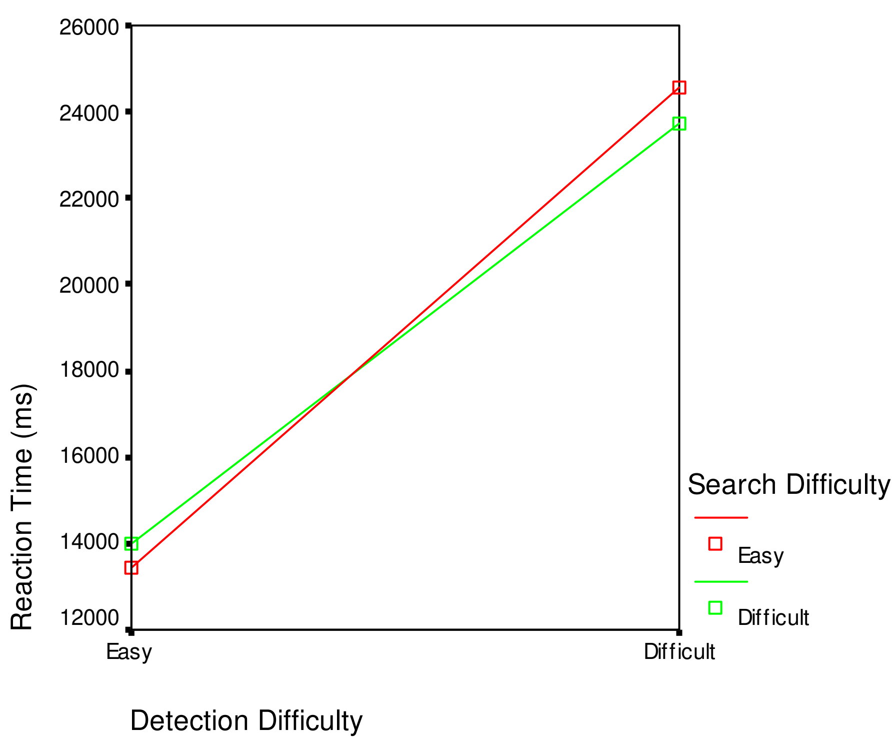


Figure 7. Detection Difficulty and Search Difficulty Effects on Reaction Time

Search Difficulty and Target Presence Effects

In order to assess search difficulty effects on performance across all trials, including interactions with target present or not, participant scores on Search Performance and Reaction Time for all trials were aggregated across trials for each of the following 4 categories: Target Not Present, Easy Search (NTES), Target Not Present, Difficult Search (NTDS), Target Present, Easy Search (TES), Target Present, Difficult Search (TDS). Outcome Performance and Detection Errors were not included as they required a target present to calculate. Two repeated measures ANOVAs were performed with IVs of search difficulty and target present and DVs of Search Performance and Reaction Time.

Search Performance

A repeated measures ANOVA was performed on Search Performance for the four categories mentioned above. The analysis revealed that target presence had a significant effect ($F(1, 18) = 92.30$, $p<.01$, partial $\eta^2=.84$) on Search Performance with performance decreasing with a target present. The analysis also revealed significant effects of search difficulty ($F(1, 18) = 8.59$, $p<.01$, partial $\eta^2=.32$) on Search Performance with performance increasing with increasing search difficulty. There was not a significant interaction ($p>.05$). The means and standard deviations are presented in Table 27 and the data are illustrated in Figure 8.

Table 27. Search Difficulty and Target Presence Effects on Search Performance

	Target Presence	
Search Difficulty	**Target Not Present**	**Target present**
Easy	mean = 52.65 s.d.= 13.83	mean =37.97 s.d.= 8.82
Difficult	mean = 55.95 s.d. = 10.44	mean = 42.00 s.d. = 8.97

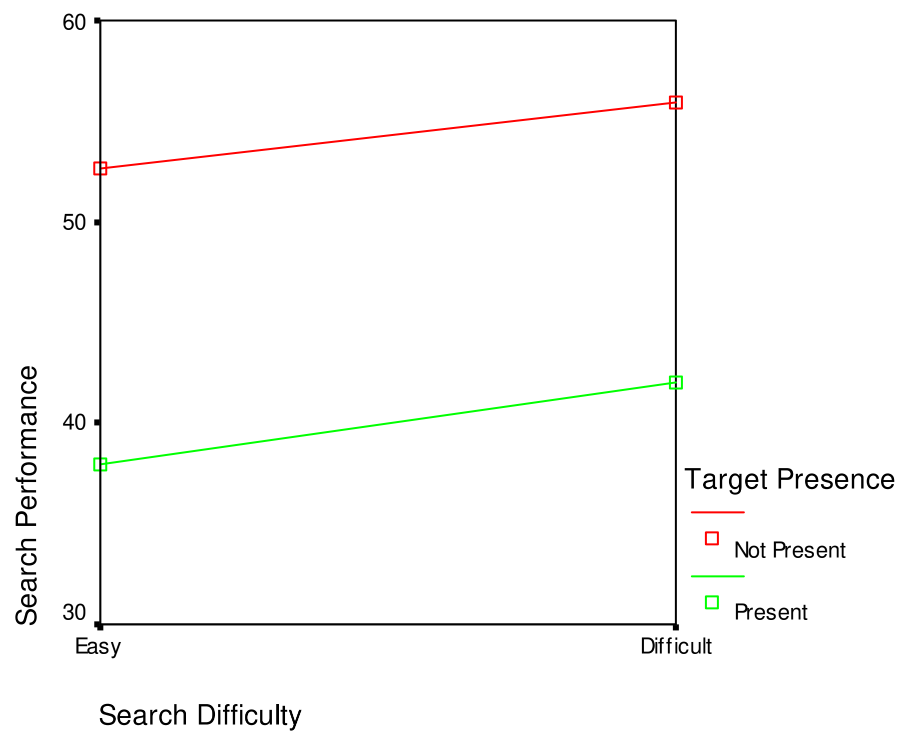


Figure 8. Search Difficulty and Target Presence Effects on Search Performance

Reaction Time

A repeated measures ANOVA was performed on Reaction Time for the four difficulty categories mentioned above. The analysis revealed that target presence had a significant effect ($F(1, 18) = 63.10$, $p<.01$, partial $\eta^2=.78$) on Reaction Time with Reaction Time decreasing with a target present. The analysis also revealed that search difficulty had a significant effect ($F(1, 18) = 9.92$, $p<.01$, partial $\eta^2=.36$) on Reaction Time with Reaction Time increasing with increasing search difficulty. There was also a significant interaction ($F(1, 18) = 6.83$, $p<.05$, partial $\eta^2=.28$). The means and standard deviations are presented in Table 28 and the data are illustrated in Figure 9.

Table 28. Search Difficulty and Target Presence Effects on Reaction Time

	Target Presence	
Search Difficulty	**Target Not Present**	**Target present**
Easy	mean = 28.40 sec s.d.= 10.46 sec	mean =18.42 sec s.d.= 5.95 sec
Difficult	mean = 33.01 sec s.d. = 10.06 sec	mean = 18.85 sec s.d. = 5.97 sec

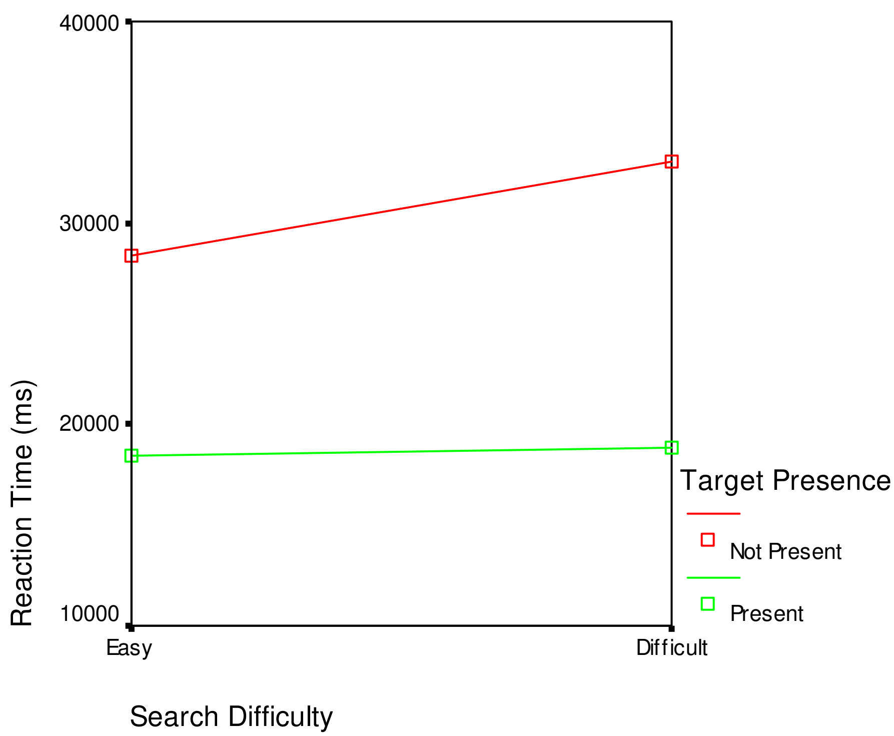


Figure 9. Search Difficulty and Target Presence Effects on Reaction Time

Discussion

Experiment 1 aimed to examine the effectiveness of utilizing eye tracking data to measure and diagnose search performance and performance deficiencies, by investigating 1) how effectively the method could identify where in the perceptual process breakdowns occur (i.e., search, detection, recognition) and 2) how reliably it could predict Outcome Performance and differentiate Search Performance from other perceptual processes (i.e., detection). Based on existing research, it was hypothesized that the diagnosis method would provide training practitioners and training science researchers with an effective method that facilitates identification of the perceptual root cause of target misses, allowing them to better understand process level performance deficiencies and tailor training remediation to address these deficiencies.

The results provide significant support for the effectiveness of the eye tracking-based diagnosis method. Utilizing fixation duration measurements, an effective threshold was determined which facilitated the target-based diagnosis of the perceptual root cause of target misses, specifically whether the miss was due to a search error, detection error or recognition error. This level of perceptual performance diagnosis gives practitioners and researchers access to perceptual processes previously inaccessible. Additionally, when these metrics were extended to diagnose trial based performance, they proved to be valid and reliable metrics predictive of Outcome Performance. These findings show promise for the effectiveness of utilizing eye tracking to better understand perceptual performance, including performance on sub processes such as search and detection and how these contribute to Outcome Performance. Such findings have far reaching implications for practitioners and researchers alike. Findings and implications are discussed in greater detail in the sections below.

Target-based Diagnosis Method

The goal of this analysis was to determine meaningful parameters with which to discriminate and classify perceptual errors. The results indicate that, in fact, the method can discriminate three different types of errors which occur at 0 millisecond fixations, between 0 and 300 millisecond fixations and beyond 300 millisecond fixations, respectively. Essentially, this indicates that there are three error categories for perceptual skills that are commensurate with the literature (i.e. search, detect, and recognize). Being able to successfully measure these with advance technologies allows researchers to not only better understand how people process information perceptually, but also, the effects different influences have on the perceptual process (whether training remediation, neurological disorder or medical treatment).

Fixation Duration Threshold

Hypothesis 1 predicted that the average fixation duration target-based metric can effectively (reliably and validly) diagnose where in the perceptual process (i.e., search, detect, recognize) breakdowns occur using eye tracking and behavioral data. This hypothesis was fully supported. A fixation duration based threshold was identified which allowed the classification of errors into three categories of search, detection and recognition. Although this threshold was not determined as originally anticipated (based on the average fixation durations of non-target verses. missed target fixations, which was not significantly different), an alternate method resulted in success. Based on the 300 millisecond threshold from the literature (Hale et al., 2007) and the average fixation durations resulting from Analysis 1.1, multiple potential thresholds were identified. These four thresholds (200ms, 250ms, 300ms, 350ms) were then used to categorize target misses into error categories and to calculate trial-based metrics to correlate with Outcome Performance. Based on the findings from these two analyses it was determined that 300 milliseconds was the most meaningful and effective threshold for categorizing perceptual errors into detection and recognition errors. The findings are discussed further in the following section.

Missed Target Responses

Hypothesis 1 (Prediction 1.2) which predicted that for missed targets there would be either instances of 0 millisecond fixations on this threat (search errors), short fixations (detection errors) or long fixations (recognition errors) was fully supported. Using detection thresholds of 200, 250, 300, and 350 milliseconds, in all cases, misses fell into all three categories above. The majority of misses (75%) fell into the search error category, regardless of detection threshold. This was not surprising as all performers were novices. As search was the first step in the target detection process, in most cases, these novices did not successfully accomplish this initial step, preventing progression to subsequent steps. This is inline with extant research which has found that novices lack defined search strategies and spend less time than experts scanning relevant aspects of the environment (Jarodzka et al., 2009).

At the lowest threshold (200ms) 9% of misses fell into the detection error category and 17% fell into the recognition error category. As the detection threshold was raised, as expected, an increasing number of misses fell into the detection error category. Greater numbers of misses fell into the recognition error category than the detection error category for thresholds of 200 and 250 milliseconds. Greater numbers of misses fell into the detection error category than the recognition error category for thresholds of 300 and 350. At the highest threshold (350ms) 20% of misses fell into the detection category and 6% fell into the recognition category.

Error breakdowns associated with 200 and 250 millisecond thresholds were not inline with expected outcomes. Given the homogeneous nature of the target, it is unlikely that more misses were due to recognition errors than detection errors. As all targets were Dragunov sniper rifles which had very distinct features (long, straight black barrel, brown, angular butt). Once these features were detected, it is unlikely participants would fail to recognize these cues as a portion of the rifle, except in the most difficult target conditions in which targets were heavily occluded and in non-canonical orientation. Therefore, these findings suggest a threshold of either 300 or 350 millisecond as a more meaningful detection threshold.

Trial-based Diagnosis Method

The goal of this analysis was to assess the validity and reliability of the metrics involved in extending the target-based diagnosis method used to identify the perceptual root cause of errors to the trial based level to assess overall threat detection performance. The results indicate the metrics are both reliable and valid as demonstrated by their effective prediction of Outcome Performance as well as their unique contributions to Outcome Performance. Essentially, this indicates that not only can eye tracking be used to effectively diagnose perceptual root cause of individual errors, but also to characterize performance on a larger scale over multiple scenarios and when no target misses occur.

Hypothesis 2 predicted that the trial-based diagnosis method could discriminate between Search Performance and Detection Performance, as well as predict Outcome Performance based on these two metrics, and this hypothesis was fully supported. The meaningfulness and uniqueness of the metrics which comprised this diagnosis method (Detection Errors, Search Performance) was demonstrated by showing that both metrics were significantly and independently correlated with Outcome Performance.

As Outcome Performance was measured via Hit Rate, and misses (failures in Outcome Performance) were categorized into one of three error types (search error, detection error, recognition error) in the target based diagnosis method, an almost perfect prediction of Outcome Performance could be made with a combination of these three variables. Given this, detection is a necessary element for successful Outcome Performance and thus Detection Errors were hypothesized to contribute a significant amount of variance to Outcome Performance. As the goal with the trial-based diagnosis method was to also measure a unique element of search performance not being captured by the target-based method (independent of target location, is the trainee searching high priority areas?) it was hypothesized that in addition to Detection Errors, Search Performance, as measured by percentage of high priority areas scanned, would contribute a significant amount of variance to Outcome Performance as well. These hypotheses were supported by the results of the sequential multiple regression in which approximately 66% of the variance was accounted for by the two metrics (approximately 51% by Detection Errors and another 15% by Search Performance). These results suggest that the trial-based diagnosis method can discriminate between search performance and detection performance, as well as predict Outcome Performance based on these two metrics.

The above regression findings are based on a detection threshold of 300 milliseconds. This threshold was determined to be the most optimal threshold by comparing the amount of variance accounted for by models based on 200, 250, 300 and 350 millisecond thresholds. The Outcome Performance regression model based on a 200 millisecond threshold did not account for a significant amount of the variance, only 27%, with 20% due to Detection Errors and 7% due to Search Performance. The Outcome Performance regression model based on a 250 millisecond threshold did account for a significant amount of the variance, approximately 40%, with a significant amount due to Detection Errors (35%) and an insignificant amount (5%) due to Search Performance. The Outcome Performance regression model based on a 300 millisecond threshold also accounted for a significant amount of the variance, 66%, with both Detection Errors (51%) and Search Performance (15%) contributing a significant amount of variance. The Outcome Performance regression model based on a 350 millisecond threshold also accounted for a significant amount of the variance, 65%, with both Detection Errors (54%) and Search Performance (11%) contributing a significant amount of variance.

As increasing the detection threshold from 300 to 350 milliseconds increased the opportunity for (and in this case the total number of) Detection Errors, it as a result increased the amount of variance incorporated into the model (i.e., variance once excluded as recognition error variance was now included as detection error variance). If the 350 millisecond threshold was indeed a better or even equally good threshold for discriminating detection and recognition errors, an increase in amount of overall variance accounted for should have resulted. As increasing the detection threshold to 350 milliseconds did not increase the variance accounted for, 300 milliseconds was chosen as the optimal fixation duration threshold to distinguish between detection and recognition errors. This threshold is also in line with fixation duration data from previous research (Hale et al., 2007), which found significantly different average fixation durations associated with hit targets (723 milliseconds), non-target false alarms (483 milliseconds) missed targets (366 milliseconds), and correctly rejected non-targets (275 milliseconds).

Average Fixation Durations

The one prediction which was not supported was prediction 1.1 which stated that there would be a significant difference between average fixation durations associated with non-target fixations and missed target fixations. This was originally intended to serve as the basis for the fixation duration threshold. One possible explanation for this was that it was due to the nature of the scenarios. Specifically, lack of significant differences may be due to the fact that misses were comprised more of detection errors than recognition errors. As detection errors would have a fixation duration similar to that of a non-target fixation, if these were highly disproportional to the number of recognition errors (which would have higher fixation duration thresholds than non-targets), the average fixation duration associated with a miss would be decreased.

This is a likely explanation as the uniformity of the target would make the detection of the target features much more challenging than recognition that the target features were associated with the target of interest. As discussed above, all targets were

Dragunov sniper rifles and although they were presented in a range of different orientations and levels of occlusion, given the unique features of the rifle (i.e., long straight black barrel, brown angular rifle butt) it is likely that once a feature was detected, recognition would occur fairly easily. If the number of detection and recognition errors were greatly disproportional (many more detection errors than recognition errors), the average fixation duration associated with missed threats would be much lower and much closer to non-threat average fixation durations than in cases in which equivalent number of detection and recognition errors occurred. Hale et al. (2007) found difference as predicted, however, the task involved locating a more heterogeneous group of targets (white work trucks) which could vary in appearance. Future experiments could investigate this prediction using scenarios in which the target of interest varied in perceptual appearance such adding IED targets ranging from gas tanks to garbage heaps. Such a situation would likely increase the number of recognition errors, hence increasing the average fixation duration associated with missed threats.

Difficulty Effects

Difficulty effects were examined via an exploratory analysis in an attempt to understand what effect detection difficulty (as operationalized via target orientation and levels of occlusion) and search difficulty (as operationalized via amount of clutter and number of high priority areas) had on performance accuracy, including Outcome Performance, Detection Errors, Search Performance and Reaction Time.

Detection Difficulty and Search Difficulty

The key finding with respect to difficulty was that detection difficulty seemed to be the driving force behind performance. For very easy to detect targets, the targets were found accurately and quickly, likely due the preattentive nature of the search required to detect these salient targets. The more difficult the target was to detect, the lower the hit rate and the greater the reaction time. This suggests that key drivers in both detection accuracy and reaction time are target aspects such as target difficulty. Additionally, the number of Detection Errors was significantly influenced by detection difficulty. Increasing detection difficulty levels led to more Detection Errors. Search Performance was also significantly affected by detection difficulty. Interestingly, as detection difficulty increased, Search Performance also increased. A plausible explanation is that difficult target detection scenarios may force performers to shift from a pre-attentive or parallel search to an attentive serial search, resulting in a more analytic/systematic search process, hence increasing the number of high priority areas scanned.

Search difficulty did impact performance; however, it appeared to be less influential than detection difficulty and actually moderated by it. Outcome Performance was significantly affected by search difficulty, however, not as expected. As search difficulty increased, Outcome Performance actually increased. As hypothesized above, this is likely due to performers being forced to systematically search for targets, resulting in performers scanning a larger percent of the high priority areas leading to improved detection of targets. There was also an interesting interaction. Search difficulty had very little effect on Outcome Performance when targets were easy to detect. In these cases, performers quickly found the targets due to their salient nature, not requiring a systematic search and therefore distracters had little effect. However, when targets were difficult to detect, a serial self terminating search resulted and search difficulty (which is based partially on number of distracters) had a greater influence on whether these targets were found. Search Performance was also significantly affected by search difficulty with Search Performance increasing with increasing search difficulty. However, there was a significant interaction and in looking at the data, search difficulty only significantly affected Search Performance when detection difficulty was high and participants were forced into a serial self terminating search. When difficulty was low, search difficulty had no impact as detection was likely driven by bottom up processes due to the saliency of the target. Detection Errors were not significantly affected by search difficulty. It seemed that the effect of target difficulty had much more effect on Detection Errors than the addition of clutter.

These findings help shed light on the impact that search and detection difficulty manipulation can have on performance. Additionally, the results indicate that the detection and search difficulty variations were strong enough to ensure breakdowns in both search and detection performance, however, it appears the search difficulty manipulation was not nearly as strong as the detection difficulty manipulation. To manipulate search difficulty, the same background scene was used and clutter (including high priority area clutter) was added to the environment. Given that the background contained a great deal of clutter, in future attempts to manipulate search difficulty, it may be more effective to use different backgrounds which vary in clutter and number of high priority areas.

Target Presence and Search Difficulty

Both the presence of a target and search difficulty affected performance. With the addition of a target, Search Performance increased over trials without a target. This is not surprising; within target present trials there were trials in which targets were very easy to detect, leading to almost immediate indication of a target. In such cases, few high priority areas were likely searched before detection. However, for trials with no target, performers were likely to perform a serial self terminating search, continuing to search until the trial time was up and covering a greater percent of high priority areas. Reaction Time was also significantly influenced by presence of a target. Trials in which targets were present had shorter Reaction Times compared to trials in which no targets were present. This is also likely due to the almost immediate detection of easy to detect targets vs. the exhaustive search for targets in trials in which they were not present.

Search difficulty significantly impacted Search Performance, with increasing difficulty levels leading to increased performance. As hypothesized above, this was likely due to the performers being forced to use a more systematic serial self terminating search in more difficult search situations. Reaction Time was also significantly affected by search difficulty with Reaction Time increasing with increasing levels of difficulty. As trials became more difficult to search, performers took longer to search the environment; however, a significant interaction indicates that this is really only in trials without targets. For trials in which targets were present, search difficulty had much less influence on Reaction Time than trials in which there were no targets present. This is likely due to the extreme influence that detection difficulty had on Reaction Time, leaving little variability due to search difficulty.

Experiment 2

The goal of Experiment 2 was to explore the training effectiveness of a newly developed search training strategy which incorporated elements of expert scan and trainee scan. This training strategy provided process level search feedback which incorporated proven feedback strategies and provided support for the development of generalized search strategies. The experiment aimed to evaluate the effectiveness of the feedback strategy components on the search and detection metrics validated in Experiment 1 to determine training value added over traditional Knowledge of Results (KR) feedback. The following sections describe the newly developed eye tracking-based search training strategy.

Visual Search Training Strategy

An innovative training strategy referred to as the "Expert + Trainee" search training strategy was developed by integrating aspects of expert performance models, metacognitive strategies, attentional weighting strategies and process level feedback. These strategies were combined into a feedback method which allowed the trainee to compare their search strategy to an expert's, including where they scanned and how their scan unfolded. This feedback strategy had four key elements.

1. *Expert Scan*

First, the presentation of expert scan data was used as the foundation of the training strategy. This strategy provided trainees with a model of "good" performance, addressed both location and sequence aspects of search and has empirically proven successful in improving visual search in a range of domains.

2. *Trainee Scan*

Second, in order to support development of trainee metacogntion, trainee scan data was presented to allow trainees to explore their actual performance and understand how it differed from their intended or perceived performance.

3. *Highlighting of Differences between Expert and Trainee Scan*

Third, differences between expert and trainee performance were highlighted to illustrate areas in need of improvement. .

4. *Elaborative Feedback*

Fourth, given the complexity of scan data it may prove challenging for a novice to extract key elements intended to improve trainee performance and support the establishment of generalizable search strategies. Therefore auditory elaboration on elements of these scan paths was included to facilitate learning.

These elements were combined into a two component module in which trainees were first presented with a "Where" component which targets the location aspect of search followed by a "How" component which targets the sequence aspect of search. These components are discussed and illustrated in the following section.

"Where" Component of Search Training Strategy

The "Where" component was designed to aid trainees in developing an understanding of what areas are high vs. low priority and should or should not be searched, as well as developing a strategy which balances the time spent searching these two types of areas. This component presents expert scan data (to illustrate desired/target performance) and trainee scan data (to facilitate metacognition) side by side with fixations color coded into high and low priority area fixations (in order to highlight differences in search strategy, i.e., allocation of attention to high vs. low priority areas). Areas that an expert scanned that a trainee did not scan are highlighted with semi-transparent, color-coded (high vs. low priority) squares to illustrate differences between target and actual performance and hence areas in need of improvement.

The module then steps through the objects that an expert fixated on, while a) highlighting associated fixations (i.e., with a green outline) and areas the expert scanned that the trainee did not scan (i.e., by filling in the boxes) and b) presenting auditory elaborative feedback of an expert commenting on the area and why it should/should not be scanned. This auditory elaborative feedback was designed to aid in abstraction of a specific scan pattern to higher level strategies of where they should look in novel situations (e.g., “You should always search areas in the shadows such as alleys”). This feedback addresses both perceptual and conceptual aspects of “where to look” and is illustrated in Figure 10 below.

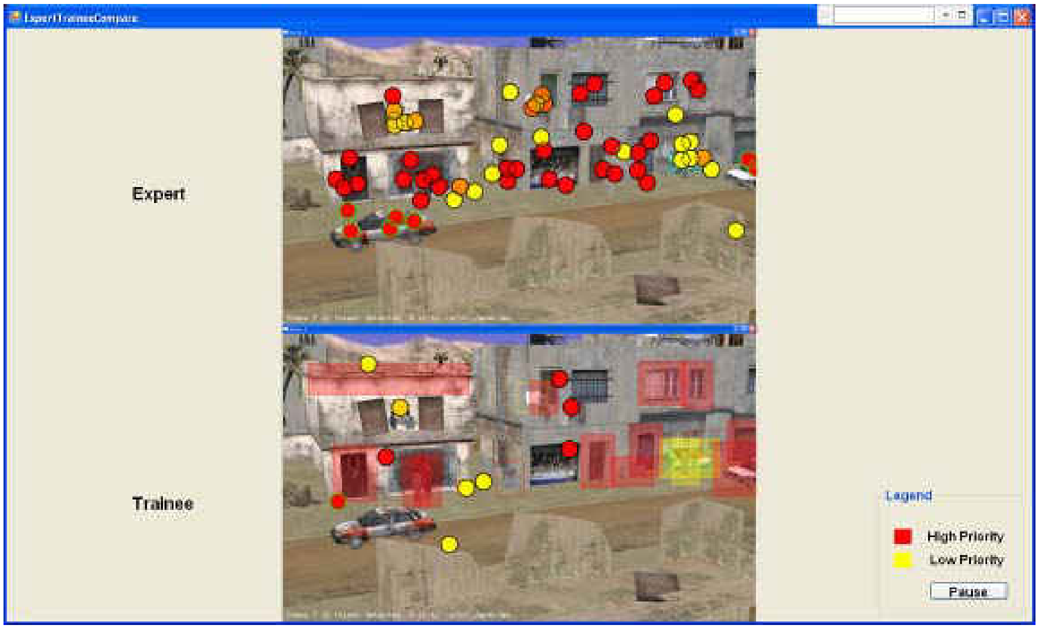


Figure 10. Search Location Feedback

"How" Component of Search Training Strategy

The "How" component was designed to aid trainees in developing systematic search strategies. This component presents the dynamic unfolding of an expert scan next to a the dynamic unfolding of a trainee scan, alternating scan segments between the two to allow an understanding of where the expert started verses the trainee and differences in how their scan paths unfolded. Specifically, the firsts ten fixations of an expert scan are presented, including color-coding of those associated with high vs. low priority areas. Upon completion of this unfolding, the firsts ten fixations of the trainee scan are presented, including color-coding of those associated with high vs. low priority areas. Upon completion of this unfolding, both sets of ten fixations turn to gray. This pattern continues, alternating between the two scan paths and turning old portions of the scan paths to gray until both scans are complete. The trainee is cued auditorily to ensure attention to the correct scan through prompts such as "The expert started scanning here", "You started scanning here", "Then the expert scanned here", "Then you scanned here", "This is the last place the expert scanned", and "This is the last place you scanned". This is illustrated in Figure 11 below.

Figure 11. Search Sequence Feedback

Although the presentation of expert scan paths has been empirically proven to enhance search strategy across multiple domains (Nalanagula et al., 2006; Sadasivan et al., 2005), it is necessary to explore the effectiveness of the above proposed extensions to this method. Not only is it necessary to explore the effectiveness of this training strategy as a whole compared to traditional Knowledge of Results (KR) feedback, it is also important to determine which elements of this feedback (expert scan, trainee scan, or both) contribute to the effect. Such is the goal of Experiment 2 of this effort.

Hypothesis 3

The newly developed search feedback strategy (expert scan + trainee scan) will improve search skills over traditional Knowledge of Results (KR) and both elements of expert scan and trainee scan will contribute to this success.

Prediction 3.1

Participants who receive training with elements of an expert scan will perform significantly better than those who do not receive elements of an expert scan in feedback in Search Performance, Detection Performance and mission Outcome Performance in the simulation post test.

Prediction 3.2

Participants who receive training with elements of a trainee scan will perform significantly better than those who do not receive elements of a trainee scan in feedback in Search Performance, Detection Performance and mission Outcome Performance in the simulation post test.

Prediction 3.3

Participants who receive training with elements of both expert and trainee scans will perform significantly better than those who do not receive both elements in feedback in Search Performance, Detection Performance and mission Outcome Performance in the simulation post test.

Method

Participants

Prior to recruiting any participants and collecting data, a power analysis was performed to estimate the number of participants needed to obtain sufficient statistical power. A power analysis was performed using the software G Power 3.0 (Faul et al., 2007). Using standard deviations from a study which explored the effect of feedforward expert scan patterns on aircraft defect detection (Sadasivan, et al., 2005), an effect size of .385 was calculated. Using this effect size for a repeated measures within and between group ANOVA with 4 groups and 2 repetitions (pretest, posttest), the software recommended a total sample size of 48 participants.

Sixty six participants took part in this experiment. They consisted of thirty three males and thirty three female participants and ranged from 18 to 25 (mean=18.9, s.d.=1.5) years of age. All participants were student volunteers from a Southeastern University and were recruited through a University system or via flyers and given either extra credit or monetary compensation for their participation.

Task

Participants performed the same experimental task on the apparatus described in Experiment 1. The only difference was the addition of feedback presented by the performance assessment system via the flat screen display. Participants performed six pretest scenarios after which no feedback was given, eight training scenarios after which feedback was given according to experimental condition and six post test scenarios after which no feedback was given. Sadasivan et al. (2005), found a significant performance gain on airframe inspection after five trials of feedforward expert scan data and Nalanagula et al. (2006), found significant performance gains on circuit board inspection after three trials of feedforward expert scan data. Therefore it was determined that eight trials with feedback would provide ample opportunity to impact performance. Prior to performance, participants received pre-training, in which they were taught scanning strategies consisting of what areas are high priority and low priority, general rules for sequence of scan (e.g., high priority first, near to far) and details of the feedback they were to receive after each training trial, including how to interpret and use it to improve future performance.

Experimental Design

Experiment 2 was a 2x2x2 mixed factorial repeated measures design, with between subjects independent variables of Expert Scan (present vs. not present) and Trainee Scan (present vs. not present) and within subjects independent variable of trial (pretest vs. posttest). Feedback conditions are listed in Table 29 and described in detail in Table 30.

Table 29. Experiment 2 Feedback Conditions

IV 2: Trainee Scan	**IV1: Expert Scan**	
	Present	***Not Present***
Present	Expert + Trainee + KR+ Elaboration	Trainee + KR+ Elaboration
Not Present	Expert + KR+ Elaboration	KR + Elaboration

Table 30. Feedback Descriptions

Type of Feedback	Description
1. KR + Elaborative Feedback	a. First, audio and textual presentation of % of high threat areas attended. b. Second, presentation of scenario view with audio elaborative feedback regarding where they should be looking (i.e. high vs. low priority areas) and why. c. Third, presentation of scenario view for trainees to practice scanning scene.
2. Expert Scan + KR + Elaborative Feedback	a. First, audio and textual presentation of % of high threat areas attended. b. Second, "Where" component described above with expert scan data and elaborative feedback only (i.e., expert fixations highlighted & associated audio). c. Third, "How" component described above with expert scan data only.
3. Trainee Scan + KR + Elaborative Feedback	a. First, audio and textual presentation of % of high threat areas attended. b. Second, "Where" component described above with trainee scan data and elaborative feedback only. (i.e., trainee fixations highlighted & associated audio) c. Third, "How" component described above with trainee scan data only.
4. Expert Scan + Trainee Scan + KR + Elaborative Feedback	a. First, audio and textual presentation of % of high threat areas attended. b. Second, "Where" component as described above. c. Third, "How" component as described above.

*Feedback was same duration in time across conditions

Dependent Variables

The following Dependent Variables (DVs) were measured:

Table 31. Dependent Variables

- Behavioral indication of threat: Y/N
- Location (and object associated) of mouse click to indicate threat
- Fixation Locations
- Fixation Durations
  - Threat Fixations
  - Non Threat Fixations
- Search Performance = % High priority areas fixated
- Search Errors = Number of threats not fixated and not indicated as threat
- Detection Errors = Number of threats fixated < 300ms but not indicated as threat
- Recognition Errors = Number of threats fixated >/=300ms but not indicated as threat
- Outcome Performance = % (hit rate) of threats correctly indicated
- Pre/Post Knowledge Test
- Participant Feedback Form

Procedure

Upon arrival, participants completed an informed consent form and a series of questionnaires and tests listed below in Table 32.

Table 32. Questionnaires

1. Demographics
2. Visual Acuity
3. Color Blindness
4. Spatial ability
    a. Spatial Orientation
    b. Visualization
    c. Hidden Figures.
5. Cognitive Load
6. Visual/Verbal Learning Style

Participants then received pre-training targeting both search and detection knowledge and skills necessary to perform the task. Pre-training was presented via a power point presentation in which screen shots from the simulation were utilized to present target and scene examples. Screenshots from the feedback modules were utilized to present feedback examples in order to facilitate trainee interpretation of feedback modules. Participants then completed a paper and pencil knowledge test addressing search strategies including "Where to look". Participants then performed the eye tracker calibration process and began experimental trial performance. Participants performed the threat detection task over a series of 20 trials, receiving no feedback after the first six trials, feedback according to condition after trials 7-14 and no feedback after trials 15-20. Participants then completed a different paper and pencil knowledge test addressing

"Where to look". Participants then filled out a Feedback form and were debriefed on the study and their participation.

Results

The analyses reported below were performed in SPSS 11.5 for windows, and all alpha levels were set to .05, unless otherwise specified. Due to anticipated eye tracking data loss caused by excessive participant movement as well as poor candidacy of participants due to causes unknown, all data was screened to identify cases in which significant eye tracking data was missing. Outlier data sets were further analyzed to determine if there was significant eye tracking data loss; those cases for which there was none or little eye tracking data for more than two trials were dropped. This data screening process resulted in the exclusion of seven of the 66 participant cases, resulting in 59 participants involved in analysis reported below.

Individual Differences

To assess individual difference effects on performance, the aptitude and demographic variables presented in Table 10 above (See Experiment 1 Results section) were correlated with Outcome Performance, Detection Errors, and Search Performance both for pretest and post test. Tables 33, 34, 35 and 36 present the correlations of aptitude and demographic variables with pretest and posttest performance variables, respectively.

Table 33. Aptitude Variable Correlations with Pretest Performance

Aptitude	Outcome Performance	Detection Errors	Search Performance
Visual Acuity	R = .22 Sig = .09 N = 59	R = -.23 Sig = .08 N = 59	R = -.10 Sig = .46 N = 59
Color blindness	R = .15 Sig = .26 N = 22	R =-.13 Sig = .35 N = 59	R = -.06 Sig = .68 N = 59
Spatial Orientation	R = .31* Sig = .02 N = 59	R = -.26* Sig = .04 N = 59	R = .21 Sig = .11 N = 59
Visualization	R = .17 Sig = .20 N = 59	R = -.05 Sig = .72 N = 59	R = .13 Sig = .33 N = 59
Hidden Figures Ability	R = .36** Sig = .01 N = 59	R = -.08 Sig = .57 N = 59	R = -.02 Sig = .91 N = 59
Visual Verbal Learning Style	R = -.26 Sig = .32 N = 59	R = -.32 Sig = .23 N = 59	R = -.06 Sig = .80 N = 59
Cognitive Load	R = -.14 Sig = .31 N = 59	R = .03 Sig = .80 N = 59	R = .27 Sig = .04 N = 59

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

Table 34. Demographic Variable Correlations with Pretest Performance

Demographics	Outcome Performance	Detection Errors	Search Performance
Gender	R = -.06 Sig = .68 N = 59	R =.01 Sig = .95 N = 59	R = -.30 Sig = .02 N = 59
Age	R = .01 Sig = .96 N = 59	R = .06 Sig = .63 N = 59	R = -.10 Sig = .44 N = 59
Highest Level of Education	R =.06 Sig = .65 N = 59	R = .03 Sig = .84 N = 59	R = -.23 Sig = .08 N = 59
SAT Score	R = -.02 Sig = .91 N = 34	R = -.01 Sig = .97 N = 38	R = .03 Sig = .87 N = 34
Gaming Experience	R = .17 Sig = .20 N = 59	R = -.06 Sig = .56 N = 59	R = .28* Sig = .03 N = 59
Military Training Experience	No Participants had military experience so no variability.		
Hunting Experience	R = .10 Sig = .44 N = 59	R = .14 Sig = .29 N = 59	R = .12 Sig = .37 N = 59
Danger level of neighborhood growing up	R = .12 Sig = .38 N = 59	R = -.10 Sig = .44 N = 59	R = .13 Sig = .33 N = 59

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

Table 35. Aptitude Variable Correlations with Posttest Performance

Aptitude	Outcome Performance	Detection Errors	Search Performance
Visual Acuity	R = .04 Sig = .77 N = 59	R = -.14 Sig = .29 N = 59	R = -.02 Sig = .488 N = 59
Color blindness	R = .10 Sig = .46 N = 22	R = .10 Sig = .45 N = 59	R = -.10 Sig = .46 N = 59
Spatial Orientation	R = .25 Sig = .05 N = 59	R = -.66* Sig = .00 N = 59	R = .15 Sig = .25 N = 59
Visualization	R = .14 Sig = .29 N = 59	R = -.19 Sig = .15 N = 59	R = .12 Sig = .37 N = 59
Hidden Figures Ability	R = .28* Sig = .03 N = 59	R = -.38** Sig = .00 N = 59	R = -.05 Sig = .71 N = 59
Visual Verbal Learning Style	R = .19 Sig = .15 N = 59	R = -.05 Sig = .68 N = 59	R = .02 Sig = .90 N = 59
Cognitive Load	R = -.02 Sig = .89 N = 59	R = -.10 Sig = .44 N = 59	R = .10 Sig = .44 N = 59

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

Table 36. Demographic Variable Correlations with Posttest Performance

Demographics	Outcome Performance	Detection Errors	Search Performance
Gender	R = -.30* Sig = .02 N = 59	R =.16 Sig = .21 N = 59	R = -.20 Sig = .13 N = 59
Age	R = -.11 Sig = .40 N = 59	R = -.02 Sig = .91 N = 59	R = .07 Sig = .60 N = 59
Highest Level of Education	R =-.18 Sig = ..18 N = 59	R = .05 Sig = .71 N = 59	R = -.04 Sig = .74 N = 59
SAT Score	R = -.08 Sig = .66 N = 34	R = .15 Sig = .39 N = 38	R = -.04 Sig = .81 N = 34
Gaming Experience	R = .20 Sig = .12 N = 59	R = -.13 Sig = .34 N = 59	R = .13 Sig = .32 N = 59
Military Training Experience	No Participants had military experience so no variability.		
Hunting Experience	R = -.05 Sig = .69 N = 59	R = -.05 Sig = .71 N = 59	R = .02 Sig = .86 N = 59
Danger level of neighborhood growing up	R = .34** Sig = .01 N = 59	R = -.15 Sig = .27 N = 59	R = .24 Sig = .07 N = 59

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

No consistent patterns of individual difference effects on performance emerged, although the spatial ability aptitude of Spatial Orientation (SO) had a small but significant correlation with Outcome Performance and number of Detection Errors during pretest and number of Detection Errors during posttest. Additionally, the spatial ability aptitude of Hidden Figures (HF) had a small but significant correlation with Outcome Performance during pretest and number of Detection Errors and Search Performance during posttest. Gender also had a small but significant correlation with Search Performance during pretest and Outcome Performance during posttest. Based on these findings, SO, HF and gender were tested as covariates for the following analyses;

however, as none were statistically significant covariates, they were excluded in the analyses.

Analysis 3.1: Feedback Strategy Effects

Participant scores on Outcome Performance, Detection Errors and Search Performance were averaged across pre-training trials to created average pre-training scores and across posttest trials to create average posttest scores. Three pretest/posttest repeated measures ANOVAs were performed with between subjects IVs of expert scan presence and trainee scan presence for Outcome Performance, Detection Errors, and Search Performance. A MANOVA was not used due to lack of high correlations between all dependent variables.

Outcome Performance

A pretest/posttest repeated measures ANOVA was performed on Outcome Performance (hit rate) with between subjects variables of expert scan presence and trainee scan presence. The analysis revealed that trial had a significant effect ($F(1, 55) = 65.3$, $p<.01$, partial $\eta^2=.54$) on Outcome Performance with performance increasing from pretest to posttest. The analysis revealed no significant interaction between trial and expert scan ($p>.05$), no significant interaction between trial and trainee scan ($p>.05$) and no significant interaction between trial and expert scan and trainee scan ($p>.05$). The means and standard deviations are presented in Table 37 and Figure 12.

Table 37. Expert Scan and Trainee Scan Effects on Outcome Performance

	Expert Scan			
	Not Present		Present	
Trainee Scan	Pre-test	Post -test	Pre-test	Post-test
Not Present	mean = .48 s.d.= .06	mean = .73 s.d.= .06	mean = .55 s.d.= .06	mean = .75 s.d.= .07
Present	mean = .40 s.d.= .06	mean = .70 s.d.= .07	mean = .45 s.d.= .06	mean = .72 s.d.= .07

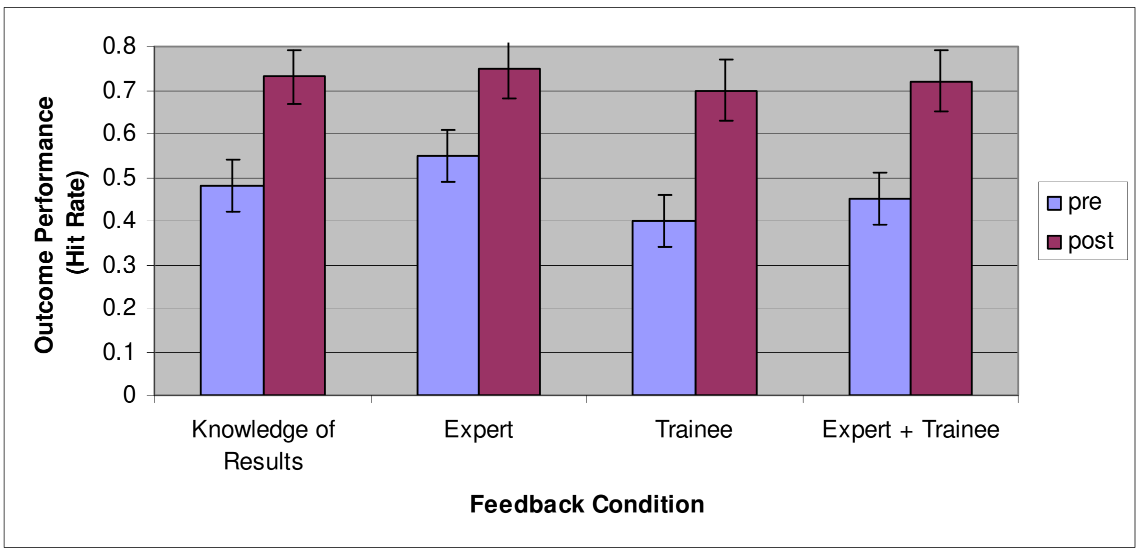


Figure 12. Feedback Condition Effects on Outcome Performance

Detection Errors

A pretest/posttest repeated measures ANOVA was performed on Detection Errors with between subjects variables of expert scan presence and trainee scan presence. The analysis revealed that trial had no significant effect (p>.05) on Detection Errors and that there was no significant interaction between trial and expert scan (p>.05), no significant interaction between trial and trainee scan (p>.05) and no significant interaction between trial and expert scan and trainee scan (p>.05). The means and standard deviations are presented in Table 38 and Figure 13.

Table 38. Expert Scan and and Trainee Scan Effects on Detection Errors

	Expert Scan			
	Not Present		Present	
Trainee Scan	**Pre-test**	**Post -test**	**Pre-test**	**Post-test**
Not Present	mean = .13 s.d.= .14	mean = .13 s.d.= .14	mean = .47 s.d.= .14	mean = .27 s.d.= .14
Present	mean = .29 s.d.= .15	mean = .14 s.d.= .15	mean = .21 s.d.= .15	mean = .21 s.d.= .15

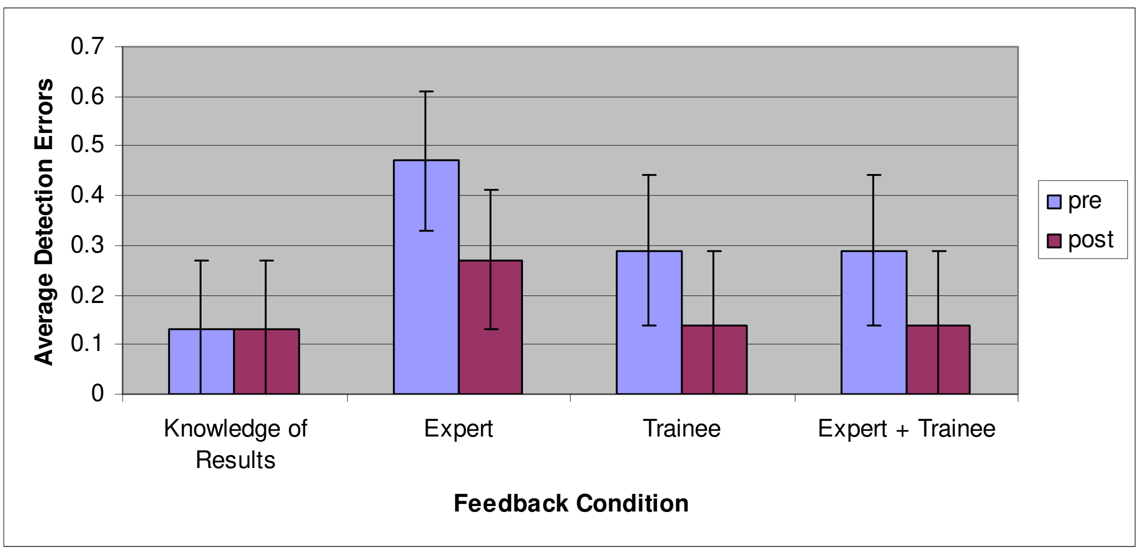


Figure 13. Feedback Condition Effects on Detection Errors

Search Performance

A pretest/posttest repeated measures ANOVA was performed on Search Performance with between subjects variables of expert scan presence and trainee scan presence. The analysis revealed that trial had a significant effect ($F(1, 55) = 77.10$, p<.01, partial η^2=.58) on Search Performance, with Search Performance increasing from pretest to posttest. The analysis also revealed that there was a significant interaction between trial and expert scan ($F(1, 55) = 4.21$, $p<.05$, partial η^2=.07), with greater pretest/posttest Search Performance improvements for participants who received elements of the expert scan than those that did not. The analysis also revealed a significant interaction between trial and trainee scan ($F(1, 55) = 4.34$, $p<.05$, partial η^2=.07) with greater pretest/posttest Search Performance improvements for participants who received elements of the trainee scan than those who did not. There was not a significant interaction between trial and expert scan and trainee scan ($p>.05$). The means and standard deviations are presented in Table 39 and Figure 14.

Table 39. Expert Scan and Trainee Scan Effects on Search Performance

	Expert Scan			
	Not Present		**Present**	
Trainee Scan	**Pretest**	**Post test**	**Pretest**	**Posttest**
Not Present	mean = 48.9 s.d.= 2.7	mean = 55.0 s.d.= 3.6	mean = 53.1 s.d.= 2.8	mean = 66.9 s.d.= 3.7
Present	mean = 55.4 s.d.= 2.9	mean = 69.3 s.d.= 3.8	mean = 52.6 s.d.= 2.9	mean = 71.1 s.d.= 3.8

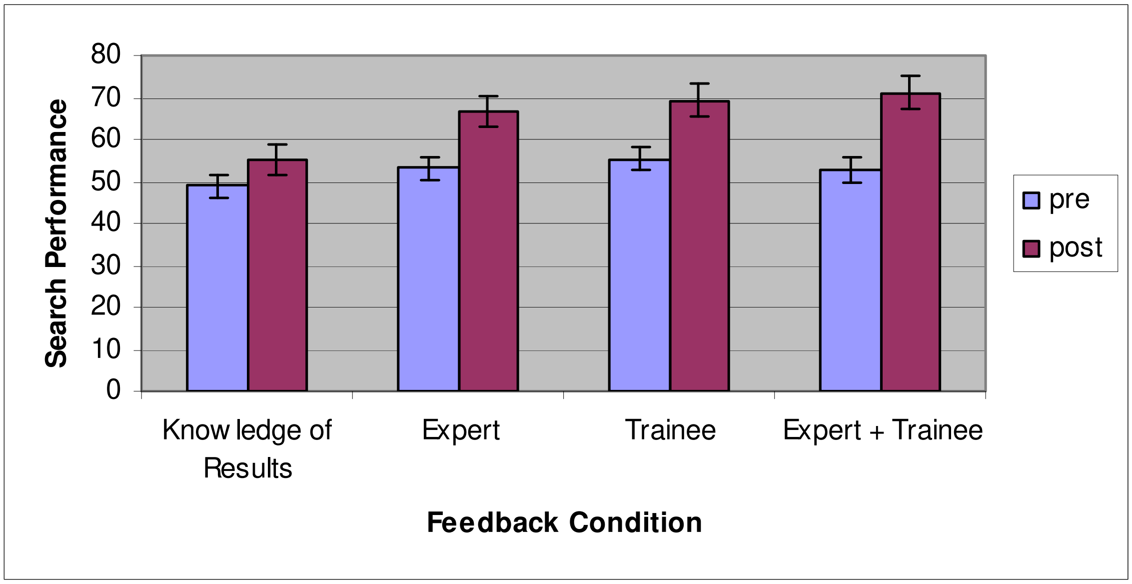


Figure 14. Feedback Condition Effects on Search Performance

Additional Analyses: Evaluation of Sensitivity and Response Criterion

Additional exploratory analyses were run to evaluate the signal detection theory metrics of Sensitivity and Response Criterion to determine if the training strategies affected these two variables. Trial had a significant effect on both Sensitivity (which increased from pretest to posttest) and Response Criterion (which decreased from pretest to posttest). Although the treatment conditions did not have a statistically significant effect on either of these metrics, with the addition of trainee scan data, Sensitivity trended toward a significant increase from pretest to posttest.

Sensitivity

A pretest/posttest repeated measures ANOVA was performed on Sensitivity with between subjects variables of expert scan presence and trainee scan presence. The analysis revealed that trial had a significant effect ($F(1, 55) = 10.67$, $p<.01$, partial $\eta^2=.16$) on Sensitivity, with Sensitivity increasing from pretest to posttest. This analysis also revealed that there was no significant interaction between trial and expert scan ($p>.05$) and no significant interaction between trial and trainee scan ($p>.05$), although this trended toward significance ($p=.054$). Additionally, there was no significant interaction between trial and expert scan and trainee scan ($p>.05$). The means and standard deviations are presented in Table 40 and Figure 15.

Table 40. Expert Scan and and Trainee Scan Effects on Sensitivity

	Expert Scan			
	Not Present		Present	
Trainee Scan	**Pre-test**	**Post -test**	**Pre-test**	**Post-test**
Not Present	mean = .85 s.d.= .22	mean = .93 s.d.= .21	mean = .85 s.d.= .23	mean = 1.02 s.d.= .21
Present	mean = .40 s.d.= .24	mean = 1.15 s.d.= .22	mean = .45 s.d.= .23	mean = .95 s.d.= .22

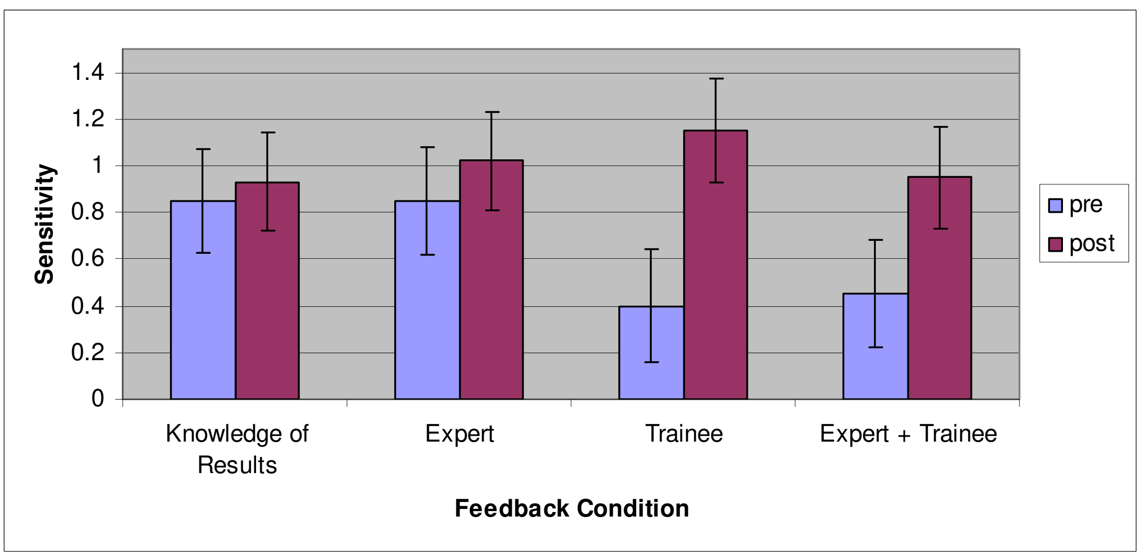


Figure 15. Feedback Condition Effects on Sensitivity

Response Criterion

A pretest/posttest repeated measures ANOVA was performed on Response Criterion with between subjects variables of expert scan presence and trainee scan presence. The analysis revealed that trial had a significant effect ($F(1, 55) = 70.90$, $p<.01$, partial $\eta^2=.56$) on Response Criterion, with Response Criterion decreasing from pretest to posttest. The analysis revealed no significant interaction between trial and expert scan ($p>.05$), no significant interaction between trial and trainee scan ($p>.05$) and no significant interaction between trial and expert scan and trainee scan ($p>.05$). The means and standard deviations are presented in Table 41 and Figure 16.

Table 41. Expert Scan and and Trainee Scan Effects on Response Criterion

	Expert Scan			
	Not Present		Present	
Trainee Scan	**Pre-test**	**Post -test**	**Pre-test**	**Post-test**
Not Present	mean = .44 s.d.= .08	mean = -.15 s.d.= .11	mean = .31 s.d.= .08	mean = -.11 s.d.= .11
Present	mean = .46 s.d.= .08	mean = .03 s.d.= .11	mean = .38 s.d.= .08	mean = -.07 s.d.= .11

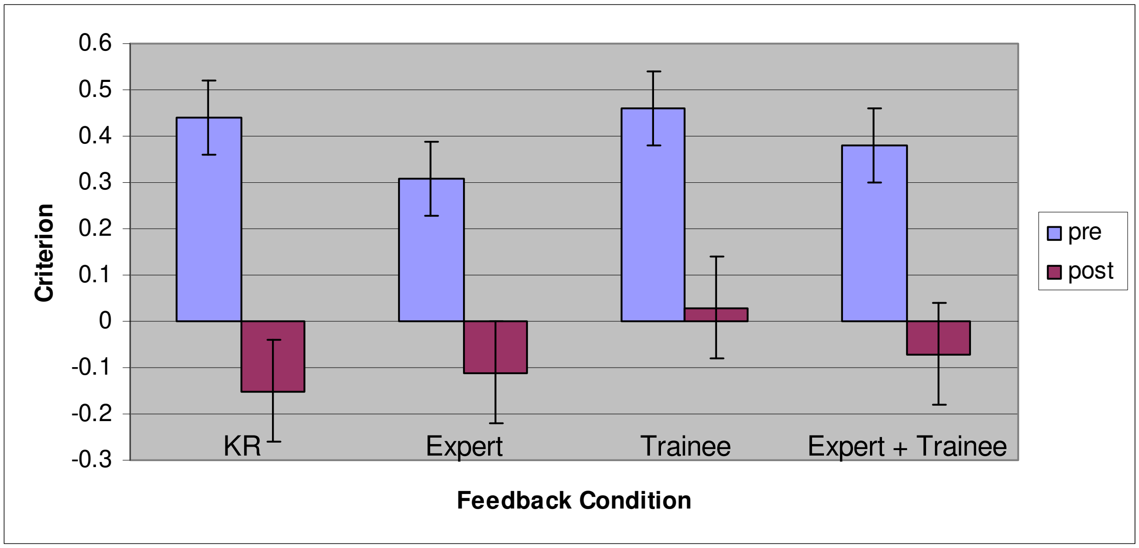


Figure 16. Feedback Condition Effects on Response Criterion

Discussion

Experiment 2 aimed to examine the training effectiveness of the eye tracking-based feedback strategy which incorporated elements of expert and trainee scan data. By allowing trainees to compare their scan to an expert scan and determine how their search strategy needed to change to achieve target performance, it was hypothesized that trainees would more effectively search the environment resulting in greater number of targets found and fewer targets that were searched but not detected. It was also hypothesized that both elements of the training strategy would contribute to the success of the strategy as the addition of expert scan would provide trainees with a model of "good" performance which they should strive to achieve and the trainee scan would provide trainee insight into how they were actually performing, increasing their metacognition and enabling them to understand issues with their performance. Together these elements were hypothesized to have an additive effect, allowing trainees to compare actual to desired performance to identify specific instance of how performance needed to change.

The study investigated this by exploring the effects that presence of each feedback element (expert scan, trainee scan) had on threat detection performance, specifically, Outcome Performance as indicated by hit rate, Search Performance, and number of Detection Errors. The results provide significant support for the effectiveness of this eye tracking-based training strategy. Both the addition of expert scan and trainee scan significantly improved how the trainee searched the environment. The combination of these two elements resulted in the greatest improvements to Search Performance, however, this additive effect was not significantly higher, statistically. Although trainee search strategies were significantly impacted by the training intervention, these effects did not extend to increase the number of targets found or decrease Detection Errors as predicted. Findings are discussed in greater detail in the sections below.

Presence of Expert Scan in Training Strategy

Hypothesis 3 (prediction 3.1) predicted that participants who received training with elements of an expert scan would perform significantly better than those who did not receive elements of an expert scan in feedback in Search Performance, Detection Performance (i.e., detection errors) and mission Outcome Performance in the simulation posttest. This prediction was partially supported. The presence of an expert scan in feedback led to a significant increase in Search Performance improvement from pretest to posttest. These participants increased the percentage of high priority areas scanned from pretest to posttest by a significantly greater percentage (16% improvement) than those who received feedback which did not contain an expert scan (10% improvement). This suggests that the presentation of expert scan alone provided trainees with enough information about desired performance that, in hand with trainee's innate awareness of how they were performing, they were able to make significant changes to how they searched the environment. Without the illustration of target performance (expert scan), trainees may not have been able to extend the elaborative feedback provided in the KR condition (e.g., you should always search alleys, because they provide cover and concealment) to actual locations in the scenarios, resulting in less impact on search strategies used.

However, presence of an expert scan in feedback did not lead to significant pretest/posttest improvements for Outcome Performance or decrease in number of Detection Errors. This is contradictory to findings in the literature in which presentation of expert scan data led not only to changes in search strategy, but also to increases in target detection (Nalanagula et al., 2006; Sadasivan et al., 2005). These somewhat contradictory results could be due to multiple reasons. The first is that there were limited pretest and posttest trials. There were only six pretest trials and six posttest trials, with each having four trials with targets and two trials without targets. This resulted in only four scores on which to compare pretest/posttest performance changes. Furthermore, the four scenarios used in these four trials varied in difficulty with the four difficulty categories discussed above (Easy Detect Easy Search, Difficult Detect Easy Search, Easy Detect Difficult Search, Difficult Detect Difficult Search). As discovered in Experiment 1, detection difficulty significantly impacted Outcome Performance and Detection Errors and this effect may have been stronger than effects resulting from the addition of an expert scan to feedback. Another possible, but less likely explanation is that although the addition of expert scan led to improvements in the first step in target detection (i.e., search strategies employed), it was not successful in impacting subsequent steps in the perceptual process of target detection and recognition. This explanation is less plausible due to 1) the findings of Experiment 1 which demonstrated that Search Performance accounted for a significant amount of variance associated with Outcome Performance and 2) findings from similar research in which presentation of expert scan has led to significant improvements in target detection (Nalanagula et al., 2006; Sadasivan et al., 2005). In all three of these studies, there were significantly more opportunities for target detection. Experiment 1 included 40 trials, 20 of which had targets; Sadasivan et al. (2005) pretest/posttest trials contained twenty two targets (i.e., defects) and Nalanagula et al. (2006) test trials contained 68 targets (defects). This provides support for the first explanation that lack of significant improvement in Outcome Performance and decrease in Detection Errors was due to the limited number of pretest/posttest trials.

Presence of Trainee Scan in Training Strategy

Hypothesis 3 (prediction 3.2) predicted that participants who received training with elements of a trainee scan would perform significantly better than those who did not receive elements of a trainee scan in feedback in Search Performance, Detection Performance (i.e., Detection Errors) and mission Outcome Performance in the simulation posttest. This prediction was also partially supported. The presence of a trainee scan in feedback led to a significant increase in Search Performance improvement from pretest to posttest. These participants increased the percentage of high priority areas scanned from pretest to posttest by a significantly greater percentage (also a 16% improvement) than those who received feedback which did not contain a trainee scan (10% improvement).

This suggests that the presentation of trainee scan data alone was able to impact performance as much as expert scan data, which is surprising. It would seem that information about actual performance has limited utility without information about desired performance and hence how actual performance needs to change. One plausible explanation is that the elaborative feedback presented with the trainee scan may have provided enough information about desired performance that with increased levels of metacognition resulting from the trainee scan data allowed trainees to make changes to how they scanned the environment. An alternate explanation is that although the expert scan feedback provided more meaningful information, trainees paid closer attention to the feedback when it contained elements of their own scan. Given a natural inclination to compare actual performance as remembered to actual performance captured by the system, participants may have attended more closely to the trainee scan feedback and gleaned more information from the feedback as compared to expert scan which may have seemed less interesting to participants.

Similar to findings associated with expert scan presence, presence of a trainee scan in feedback did not lead to significant pretest/posttest improvements for Outcome Performance or decrease in number of Detection Errors. This is hypothesized to be due to the same reason that the addition of expert scan did not significantly improve Outcome Performance or decrease Detection Errors, the limited number of pretest/posttest trials.

Presence of Expert and Trainee Scan in Training Strategy

Hypothesis 3 (prediction 3.3) predicted that participants who received training with both elements of expert and trainee scan would perform significantly better than those who did not receive both elements of feedback in Search Performance, Detection Performance (i.e., Detection Errors) and mission Outcome Performance in the simulation posttest. This prediction was not supported. Although participants who received the feedback which contained both elements had the greatest increase in the percentage of high priority areas scanned from pretest to posttest (19% improvement), this was not significantly greater than improvements demonstrated by expert scan only (16% improvement) and trainee scan only (16% improvement) conditions as illustrated by the lack of a significant interaction between expert scan presence, trainee scan presence and trial.

There are multiple plausible explanations for this finding. First, it may be due to the fact that presence of either set of scan data provides trainees with enough perceptual feedback so that, with the addition of the auditory elaboration they are able to adjust their scan strategies accordingly. One reason for the inclusion of the scan data is that while verbal description of how an environment should be scanned can effectively target conceptual aspects of the tasks (e.g., what areas should be scanned and why), it may not have the ability to effectively target perceptual aspects of the task (e.g., what do these areas look like). Expert scan feedback was intended to provide target performance (i.e., performance trainee should strive for), trainee scan was intended to provide metacognition of trainee performance (i.e., how the trainee is actually performing) and the combination of both was intended to highlight performance improvements needed to reach target performance levels. Perhaps trainees possessed high enough levels of metacognition that by merely being exposed to the expert scan allowed them to identify on their own how their performance needed to change. Additionally, the auditory elaboration may have provided enough information about target performance that with the trainee scan providing needed perceptual aspects associated with this elaboration that trainees were able to effectively adjust their search strategy. The combination of the two may not have provided significantly different information than the expert or trainee scan alone.

Alternately, the expert and trainee scan combination may have provided significantly different information, but the displays may have created excessive levels of workload preventing trainees from extracting the relevant information. Presentation of both scans with differences highlighted resulted in a great deal of information trainees would have to process in the same amount of time as the expert and trainee only conditions. In fact, the combination of the two scans and highlighted differences more than doubled the amount of visual information presented. Although the displays were designed to try and highlight differences, eliminating the need for trainees to absorb all information about each scan that was presented, this may not have been the case. When users are presented with large amounts of data that exceed workload limitations, information overload can occur resulting in failure to discover overall trends in the data (Chung, Chen, Chaboya, O'Toole, & Atabakhsh, 2005).

In fact, anecdotal evidence from interactions with the participants indicated that the feedback which contained both expert and trainee scan data was harder to follow than the other feedback modules. Specifically, although the “where” portion of the feedback seemed equivalent across conditions, the "how" portion which presents the dynamic unfolding of the scan data resulted in trainees "getting lost" or "bored". Participants in the expert and trainee scan only conditions watched the unfolding of the scan and were given the remaining time to explore the scan pattern. The condition which contained both scans had to watch twice as much scan data unfold which seemed to result in some trainees not paying attention to the entire feedback module, potentially due to being overloaded or overwhelmed. Hence, the feedback which contained both expert and trainee scan data may have created high levels of workload resulting in trainees not being able to extract significantly more information than from the expert or trainee scan alone displays.

Sensitivity and Response Criterion

Sensitivity and Response Criterion, two standard Signal Detection Theory metrics, were examined via an exploratory analysis in an attempt to gain a more thorough understanding of feedback effects on target detection performance. Given the limited number of trials, it was thought that analysis of Sensitivity and Response Criterion might provide a more granular look at target detection performance, resulting in findings not evident by looking at hit rate alone.

Sensitivity

Sensitivity, refers to the keenness or resolution a trainee has in their ability to detect the target (Wickens & Hollands, 2000). A trainee with high Sensitivity would have good ability to discriminate a target from a non-target and a trainee with low Sensitivity would have a poor ability (Macmillan & Creelman, 2005). This measure provides a supplemental look at Outcome Performance as it takes into consideration both hit rate and false alarm rate. The hope was that this measure would uncover findings not revealed by merely looking at Outcome Performance as it increased the pretest/posttest trials used in the analysis from four (those with targets) to six (all trials) and provided a more comprehensive breakdown of trainee performance.

Analysis revealed no significant effects of expert scan presence on Sensitivity, no significant effects of trainee scan on Sensitivity and no significant interaction. However, there was a trend identified with trainee scan effects on Sensitivity approaching significance ($p=.054$). The addition of trainee scan to feedback resulted in pretest/posttest Sensitivity increase of approximately .6 over those who did not receive trainee scan. Interestingly, this was greater than the Sensitivity increase found due to trial (.4). There was no Sensitivity increase due to the addition of expert scan. It is hypothesized that with increased pretest/posttest trials that this value would have reached a significant level. This suggests that the presence of trainee scan data in feedback can potentially increase Sensitivity to the target. This may be due to increased metacognition, specifically an increased understanding of areas which they failed to search. With trainees having a better understanding of areas they need to include in their search strategy in the future, they may effectively expanded the high priority areas searched and decreased the number of low priority areas search, leading not only to improved Search Performance but also to decreased false alarms resulting in increased target Sensitivity. Perhaps with a greater number of trials this Sensitivity would have manifested itself with improved Outcome Performance (hit rate).

Response Criterion

Response Criterion, refers to a trainees response bias; whether they are more liberal (i.e., prone to saying yes something is a target) and hence increasing both target hits and false alarms, or more conservative (i.e., prone to saying no) and hence having few false alarms but greater target misses (Wickens & Hollands, 2000). This measure also provides a supplemental look at Outcome Performance (hit rate) by taking into account false alarm rate as well.

Analysis revealed no significant effects of expert scan presence on Response Criterion, no significant effects of trainee scan on Response Criterion and no significant interaction. Although response bias became significantly more liberal between pretest and posttest, the different training strategies appeared to have no effect on Response Criterion.

CHAPTER FOUR: GENERAL DISCUSSION AND CONCLUSIONS

General Discussion

The goal of the present study was to explore how advanced technology, specifically eye tracking, can be used to increase understanding of perceptual processes such as search and detection as well as provide tools that can be used to impact perceptual performance. With the increasing demands of the war in the Middle East, increasing diagnosis of diseases and disorders and increasing complexity of systems being used by practitioners, there is an ever increasing need to understand how the human body works and how technology can be leveraged to enhance it. Central to this is a solid understanding of how humans collect and process information and what influences this process.

Process Level Perceptual Performance Measures

Experiment 1 aimed to examine a method of diagnosing perceptual performance in order to be able to identify the perceptual root cause of target detection deficiencies and how these impact overall target detection performance. Findings indicate the method can be used to pinpoint where in the perceptual process a target miss originated, whether due to ineffective search strategy, inability to detect the subtle cues of the threat or inability to recognize these cues as indicative of a threat. These findings are inline with the Human Information Processing (HIP) model (Wickens & Flach, 1988) which describes bottom up and top down processing conducted by individuals to build situation awareness (SA), conduct decision making (DM), and act upon the environment.

Individuals progress through the first three components of the HIP model (i.e. attention, sensation, perception) to perform a perceptual task and as illustrated by the results of this study, at any of these stages an information processing breakdown could occur. Breakdowns can occur at the attention stage during the process of spotlighting particular areas of the visual field (i.e., search; Levine, 2000), the sensation stage which refers to initial detection of a stimulus (Pike & Edgar, 2005), or the perception stage which involves the analysis of sensory information to construct a description of the surrounding world (i.e., recognition; Pike & Edgar, 2005). Regardless of where the breakdown occurs, SA and DM resulting from this process will be affected.

This ability to identify where the performance deficiency originated, or the "Root Cause" of the performance deficiency facilitates a foundation for understanding how to prevent reoccurrence of the outcome failure. This process has been used in accident investigation programs for years to identify how and why undesirable events occurred, in order to prevent reoccurrence (Rooney & Heuvel, 2004). The root cause analysis process is designed for use in categorizing, linking, and refining probable causes of events, in order to be able to specify workable corrective measures that prevent future events of the type observed (Rooney & Heuvel, 2004). This study illustrates that extending this methodology to performance measurement can provide an effective method for categorizing performance failures into more granular process level causes. Integration of eye tracking technology gives researchers and practitioners access to process level data that facilitates this level of analysis.

Such granular process level measures have allowed researchers to uncover several important findings. For instance, Mello-Thoms et al. (2002) discovered that faulty visual search is not the main reason why most breast cancer lesions are not detected in mammograms, but that perception and decision making errors are primarily responsible. Additionally, Manning, Ethell, Donovan and Crawford (2006) utilized process level eye tracking metrics to identify distinct differences between scan strategies used by experienced and inexperienced observers, resulting in recommendations for how to increase expertise. In Experiment 2 of the present study, the process level measures facilitated identification of performance differences between treatment groups that were not apparent through behavioral measures alone. There were clear differences in search strategies that would not have otherwise been evident. All three of these examples provide proof that eye tracking-based process level measures such as those developed and validated in this effort provide critical tools for use in the study of perceptual skills.

These measures also provide invaluable tools for training practitioners as they allow practitioners to focus training remediation on the most critical performance deficiencies. The feedback given to address an erroneous search strategy (e.g., providing information on how the environment should be searched) is quite different than that aimed to address issues with target recognition (e.g., what are the critical cues that together indicate a threat and why). By allowing instructors to tailor training based on process level performance deficiencies, there is an opportunity to accelerate learning.

Influences of Visual Search

Findings from both Experiment 1 and Experiment 2 are consistent with contemporary models of visual search. For instance, a model proposed by Itti & Koch (2001) consists of bottom up preattentive search based on the environmental influence of salience as manifested through factors such as orientation. Findings from Experiment 1 indicate that detection difficulty which was a product of target orientation and level of occlusion had a significant impact on performance and reaction time. This supports the theory that search is partially influenced by bottom up environmental factors. The same model (Itti & Koch, 2001) proposes that top down processes also drive attention through the application of search strategies. Experiment 2 findings indicate that participant search performance was altered through feedback aimed at impacting the search strategies being employed; hence supporting the theory that search is also partially influenced by top down processes.

Experiment 1 findings are also inline with models such as the Guided Search Model (Wolfe, 1998) and the Feature Integration Theory model (Treisman & Gelade, 1980). These theories predict that when search is preattentive and the target "pops out" at the observer that the number of distracters will have little effect on performance; however, when the search is attentive, requiring serial allocation of attention the performance will be affected by the number of distracters (Wolfe et al., 1989). This is evident through the impact of search and detection difficulty on Outcome Performance and Search Performance found in Experiment 1. When detection difficulty was low and the targets were processed preattentively, search difficultly, partially based on number of distracters, had little effect on Outcome Performance and Search Performance. However, when detection difficulty was high, requiring a serial attentive search, search difficulty had a significant impact on performance.

Experiment 1 findings are also consistent with predictions regarding the serial self terminating nature of attentive search (Treisman & Souther, 1985) which predicts that in this type of search, observers will search through the stimulus elements until the target is found; hence as number of distracter increase, reaction time will also increase. Experiment 1 results indeed show an increase in reaction time with increasing search difficulty which was partially based on number of distracters.

Training Effects on Visual Search

Experiment 2 aimed to examine the effect of providing trainees with process level tailored feedback which incorporates elements of expert and trainee scan patterns. Findings indicate that providing trainees with elements of either expert or trainee scan patterns has the ability to significantly improve the search strategy being employed by the trainee. These findings provide a foundation for understanding ways in which visual search strategies can be influenced. For instance, it appears that the presentation of exemplar perceptual performance can significantly improve search strategies. By presenting trainees with expert scan patterns, including all areas in which the expert searched as well as how the expert scan unfolded, significant improvements in number of critical areas covered in a trainee scan strategy may result. This very specific process level perceptual feedback which illustrates how the trainee should scan the environment enables the trainee to make significant changes to their search strategy.

Additionally, the presentation of actual perceptual performance, including all areas in which the trainee searched and how the trainee scan unfolded can also lead to significant improvements in number of critical areas covered in a trainee scan. This process level perceptual feedback, which illustrates how the trainee scanned the environment, enlightens trainees to how performance actually unfolded, allowing them to make significant changes to their search strategy. These results provide strong support for the use of either expert or trainee scan data in perceptual feedback. Interestingly, although the combination of these two elements (expert and trainee scan) resulted in the greatest Search Performance improvement, it was not significantly greater than improvements achieved by trainees who received the individual elements of expert or trainee scan alone. This could be due to limitations with the feedback interface, such as excessive workload, resulting in trainees not being able to extract all critical information about target performance and how actual performance needs to change.

Although the findings provide a foundation for understanding ways in which visual search strategies can be influenced, it is still necessary to understand how these training interventions influenced visual search. For instance, it may be the case that the feedback modules actually led to alterations in trainees' natural search strategy. Boot, Becic and Kramer (2009) provide evidence for a natural or "default" search strategy that observers bring to the table regardless of the task. Boot et al. (2009) found that observers persisted with their default search strategy (for instance, a covert search without eye movement) even when it proved maladaptive to the task at hand, however, when given simple instruction to change their search strategy, they were easily able to do so. Perhaps the instruction provided by the feedback presented in this study led to changes in participant natural search tendencies.

Research has shown that novice observers tend toward a less systematic search strategy than experts (Jarodzka et al., 2009) and in some cases what appears to be no strategy at all. Research has also shown that a systematic search strategy can be trained and leads to improved performance over random search strategies (Wang et al., 1997).

Perhaps the scan feedback facilitated a modification of novice natural search strategy (or lack thereof) towards a more systematic search strategy. Random search is defined as a search with no memory in which previously searched areas are revisited (Arani et al., 1984). Observers with a random search strategy sample the environment for the target without a clear search schema, resulting in sampling previously scanned areas and neglecting areas which have not been searched. As an observer develops a more thorough understanding of the task, search environment and techniques that can be employed to improve performance, the observer would seemingly develop a search schema which facilitates employment of more effective search strategies. The results would be a more systematic search in which an observer reduces the number of areas revisited and increases the total number of areas searched.

This provides a feasible explanation for the increase in percent of high priority areas searched after participants received the scan feedback. If participants were randomly searching the environment prior to receiving feedback the percentage of high priority areas searched would be limited due to the participants wasting time revisiting previously searched areas. Additionally, without an understanding of what constitutes a high priority area, high priority areas are no more likely to be searched than low priority areas. After receiving the feedback, participants were conceivably able to develop a search schema which included explicit knowledge of which areas were high priority and should therefore be searched first (e.g., windows are probably target locations). Additionally, participants were also conceivably able to develop a search strategy which allowed them to search the environment in a more structured way (e.g., start by searching all window first), reducing the number of areas revisited and increasing the number of relevant areas scanned.

The feedback may have not only influenced visual search through explicit knowledge, but also through implicit knowledge. Research has found that a search schema can be created through target/context covariation, resulting in implicit influence on attention deployment in search that cannot be explicitly stated (Jiang & Chun, 2001). Based on this it is possible that through scan data/scenario object covariation within the feedback modules participants were able to develop implicit knowledge of critical areas to search which influenced their visual search strategy.

As visual search depends on many low level processes and abilities, there are many aspects of visual search that the scan feedback likely did not influence. For instance, one aspect of search that the scan training strategies did not target was search speed, specifically the speed at which the eyes move while fixating. A search strategy of keeping fixations long results in slow search speed and Togami (1984) found that slower search speed (and hence longer fixation time) resulted in higher inspection accuracy. Sadasivan et al. (2005) presented fixation time as an important element of search and attempted to affect fixation time through training. Feedforward presentation of expert scan paths included a representation of time spent fixating on each area of interest in graphical form. Although the training intervention, which also included AOIs scanned and scan patterns, led to increased inspection accuracy, subjective feedback from trainees indicated they did not find the fixation time representation useful in performing the task. However, given the findings from Togami (1984), perhaps this is useful information to present and a more intuitive presentation method would yield more positive reactions.

Future development of the scan training strategies will aim to represent this critical element of search.

Visual search also depends on many aspects of the task/environment ranging from spatial and temporal uncertainty to distance and lighting effects to the effects of dynamic aspects of the environment. These are all factors that were beyond the scope of this study that will impact visual search and may influence the effectiveness of training strategies such as those examined in this study. The most critical aspect not addressed by this study is visual search in a dynamic environment. Warfighters are not typically faced with monitoring a static environment; instead, these observers are faced with identifying threats in situations unfolding dynamically. Research has shown dynamic aspects of an environment can capture attention and alter search even when these elements are irrelevant to the task (Franconeri & Simons, 2003). Such findings imply that the search strategies employed to successfully perform a search in a dynamic environment might be quite different than those employed in searching a static environment.

An important question thus emerges: Will these training strategies that successfully improved search strategies in a static environment transfer to a dynamic environment? Although traditional views of effective training transfer stress the importance of identical elements between training and transfer environments, more contemporary views of transfer suggest that what is important is the similarity in information processing of the two tasks (Ford, Smith, Weissbein, Gully, & Salas, 1998). In fact, recently, emphasis has been placed on teaching higher level processes such as metacognition to increase transfer as "similarity in stimulus conditions between training and transfer situations has rarely promoted far-reaching transfer" (Cox, 1997, p. 41). As a result, it is hypothesized that the scan feedback strategies herein will facilitate development of generalizable search strategies which will transfer to dynamic environments. Additionally, given the conceptual aspect of the feedback which focuses on why certain high priority areas should be searched, the feedback should facilitate development of a higher level search schema, also necessary for search in a dynamic environment. The dynamic environments in which Warfighters operate are comprised of static elements similar to those in the static scenarios of this present study, for instance when searching a building for potential sniper hides. Therefore this schema would prove extremely relevant to dynamic search. However, despite the theoretical and practical reasons which suggest such strategies may transfer to a dynamic task, this remains a question which needs to be evaluated empirically.

Theoretical Implications

This research provides empirical support for previous work related to visual search theory and the training of visual search. The results provide support for the Human Information Processing (HIP) model (Wickens & Flach, 1988) and the presence of distinct sub processes in the perceptual components of this process (attention, sensation, perception). The research findings indicate perceptual performance breakdowns fall into three categories of search errors, detection errors and recognition errors.

The findings are consistent with contemporary visual search models such as the Guided Search Model (Wolfe et al., 1989) and Feature Integration Theory (Treisman & Gelade, 1980) which propose both bottom up and top down influences of visual search. Present study research findings indicate visual search was indeed influenced by bottom up environmental characteristics such as target orientation as well as top down visual search strategies resulting from training intervention.

Findings are partially consistent with previous research which examined the training effects of expert scan on search performance (Nalanagula et al., 2006; Sadasivan et al., 2005). Current research findings indicate that similar to previous findings, expert scan data led to improvements in search strategy, however, contrary to previous research findings, these changes did not result in increased target detection accuracy.

These findings support the benefit of using eye tracking to assess perceptual performance and influence visual search through feedback. Hopefully these findings will help facilitate a deeper understanding of perceptual skills such as visual search and factors that influence these skills, as well as facilitate successful execution of future research to further investigate.

Practical Implications

These findings also have practical implications for both training practitioners and researchers. With respect to training practitioners, the diagnosis metrics validated in Experiment 1 provide practitioners with a set of tools to tailor training and potentially accelerate learning. With the ability to pinpoint process level perceptual performance deficiencies, feedback can be tailored to focus on those performance deficiencies which need improvement, decreasing training time wasted and increasing training efficiency. Such an increase could result in decreased training time to reach necessary performance standards. With respect to the military, this could result in a quicker deployment time or expansion of target skills trained, both critical needs of the military.

The feedback strategies found successful in Experiment 2 also provide practitioners with training methods that can effectively improve trainee search strategies. Currently, military trainees do not receive feedback on their search strategies at all as instructors do not typically have the ability to monitor search skills. These training strategies provide a means by which to effectively remediate search skills and have far reaching implications for current military observation training. If these methods could be used to improve the training effectiveness of current training practices, Warfighter search for enemy threats such as snipers and IEDs could be enhanced. Such an enhancement could lead to both a decrease in Warfighter deaths and an increase in enemy destruction or apprehension.

These finding also have implications for a range of other operational domains, such as baggage screening, air traffic control, industrial inspection and intelligence analysis. As the extraction of expert and trainee scan using eye tracking technology becomes less expensive and less complicated, the integration of these capabilities into elements of performance diagnosis and feedback can provide training value across a range of domains, resulting in increased search performance leading to reduced success of terrorists and a reduction in accidents and incidents.

These findings are relevant to the field of Human Computer Interaction (HCI) as well. The methods evaluated in the present study provide HCI practitioners with tools to evaluate effectiveness and usability of interfaces ranging from websites to Virtual Environments (VE). The ability to monitor a user’s scan patterns, including what aspects of a display or environment capture trainees attention or lead to deeper levels of visual interrogation allows a designer to understand the impact their interface has on users and how the interface may need to be change to achieve desired effects. These findings also have implications for the medical field. Methods evaluated herein provide medical professionals with tools to diagnose disabilities such as neurological disorders and determine effects, both intended and unintended, of treatments prescribed.

These findings also have practical implications for researchers. First, they provide a set of validated eye tracking-based metrics for use in investigation of perceptual performance. These metrics facilitate identification of the perceptual root cause of errors, specifically discriminating between search, detection and recognition errors, and can effectively characterize threat detection performance on a larger scale over multiple scenarios. The ability to effectively measure process level perceptual performance, such as search and detection could facilitate a range of research in the visual search domain. Not only could these metrics be used to study search and detection across varying characteristics of the search task (e.g., difficulty, size, complexity), such metrics could also be correlated with neurophysiologic metrics to gain a more thorough understanding of the underlying biological processes of search.

Limitations of Present Study

There were several limitations of the present study that should be discussed. First, the limited number of pretest and posttest trials may have inhibited the ability to find feedback effects. Results suggest that the benefits to the presence of expert and trainee scan data may not have been fully realized due to this limitation. Given the limited number of pretest and posttest trials and the strong effects of target difficulty, improvements in search strategy may not have been given ample opportunity to manifest themselves as increases in threat detection levels. Further research is needed to more fully understand potential training value of these strategies.

Additionally, limitations of the eye tracking technology may have attenuated effects of the feedback. The accuracy of eye tracking technology has limiting factors including inaccuracies caused by the movement of the participant and individual difference factors associated with the participant. As a result, some trainee fixations may not be accurately captured and presented in the feedback. Additionally, feedback displays only included fixations in which concentration of visual attention remained within a 50 pixel radius for 100 milliseconds. Given individual differences in how quickly individuals gather visual information during fixations, there may have been instances of visual attention allocation that the trainee was aware of but that were not classified as a fixation and hence not displayed in the feedback. Consequently, trainees may have detected inconsistencies between system reported trainee scan data and self perceived scan data, resulting in frustration or distrust in the system and hence diminished effects of the feedback modules. Future use of such technology should be accompanied by instructions regarding how the eye tracker calculates and displays fixations and limitations in the accuracy of the technology.

Lastly, the static nature of the experimental task limits the extent to which the findings generalize to skill performance in the field. Given the dynamic nature of the real world task and the known effects of movement on bottom up attentional mechanisms, further research is necessary to draw any concrete conclusions regarding the effectiveness of scan feedback strategies in impacting real world search performance.

Future Research

There are several interesting research questions that remain and warrant further research. First, the current study focused on identifying a universal fixation duration threshold between detection and recognition across all participants. Given individual differences, it may be more effective to identify individualized thresholds to discriminate between target detection and recognition. Togami (1984) examined the relationship between correct count rate in an inspection task and eye movements to determine individual differences. Results revealed individual differences in fixation time for a simple task comprised of counting dots of the same sample (Togami, 1984). This implies individuals likely have different time courses for target detection/recognition. Future studies could attempt to identify individualized fixation duration thresholds and the range of these thresholds.

An additional research question is whether the search performance improvements resulting from these training strategies will transfer to a dynamic task or more importantly to the performance of a military threat detection task in the field. It is hypothesized that the feedback strategies will lead to performance improvements that transfer, but this is an empirical question which needs to be examined experimentally. Currently, planning is underway to examine the effects these training strategies have on Marines both in training performance on the experimental task as well as transfer to performance in practical application exercises which require search for a range of threats in a dynamic environment.

Future work is needed to fully understand the potential benefits of utilizing eye tracking to improve training effectiveness both through performance assessment and training strategies such as feedback. With respect to this current line of research, future efforts aim to explore usability and workload issues associated with the scan feedback displays. Strides will be made to optimize these interfaces to allow trainees to fully extract the wealth of information provided without overloading the trainee.

Conclusion

With the emergence of more usable and cost effective eye tracking technology, the ability to both assess and affect trainee perceptual performance is increasing. Subtle or internal perceptual processes, such as search and detection, once inaccessible can now be observed both by trainees and instructors, providing unprecedented access to trainee performance and state.

The goal of the present study was to explore how eye tracking technology can be used to increase understanding of perceptual processes such as search and detection and provide tools that can be used to train search skills. Experiment 1 examined a method of diagnosing perceptual performance in order to be able to identify the perceptual root cause of target detection deficiencies and how these impact overall target detection performance. Findings indicate the method can be used to pinpoint where in the perceptual process a target miss originated, whether due to ineffective search strategy, inability to detect the subtle cues of the threat or inability to recognize these cues as indicative of a threat. Experiment 2 examined the training effectiveness of providing trainees with process level tailored feedback which incorporates elements of expert and trainee scan patterns. Findings indicate that providing trainees with elements of either expert or trainee scan patterns has the ability to significantly improve the search strategy being employed by the trainee. This work provides strong support for the use of eye tracking based perceptual performance diagnosis methods and training strategies in improving trainee search performance for complex target detection tasks.

APPENDIX A: INTERNAL REVIEW BOARD APPROVAL LETTER

University of Central Florida

University of Central Florida Institutional Review Board
Office of Research & Commercialization
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: 407-823-2901, 407-882-2012 or 407-882-2276
www.research.ucf.edu/compliance/irb.html

Notice of Expedited Initial Review and Approval

From : **UCF Institutional Review Board**
FWA00000351, Exp. 10/8/11, IRB00001138

To : **Meredith B. Carroll**

Date : **December 10, 2008**

IRB Number: **SBE-08-05964**

Study Title: **Empirical Evaluation of the Effectiveness of Eye-tracking-based Search Performance Diagnosis and Feedback Methods**

Dear Researcher:

Your research protocol noted above was approved by **expedited** review by the UCF IRB Vice-chair on 12/9/2008. **The expiration date is 12/8/2009.** Your study was determined to be minimal risk for human subjects and expeditable per federal regulations, 45 CFR 46.110. The categories for which this study qualifies as expeditable research are as follows:

6. Collection of data from voice, video, digital, or image recordings made for research purposes.

7. Research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies.

The IRB has approved a **consent procedure which requires participants to sign consent forms.** Use of the approved, stamped consent document(s) is required. Only approved investigators (or other approved key study personnel) may solicit consent for research participation. Subjects or their representatives must receive a copy of the consent form(s).

All data, which may include signed consent form documents, must be retained in a locked file cabinet for a minimum of three years (six if HIPAA applies) past the completion of this research. Any links to the identification of participants should be maintained on a password-protected computer if electronic information is used. Additional requirements may be imposed by your funding agency, your department, or other entities. Access to data is limited to authorized individuals listed as key study personnel.

To continue this research beyond the expiration date, a Continuing Review Form must be submitted 2 – 4 weeks prior to the expiration date. Advise the IRB if you receive a subpoena for the release of this information, or if a breach of confidentiality occurs. Also report any unanticipated problems or serious adverse events (within 5 working days). Do not make changes to the protocol methodology or consent form before obtaining IRB approval. Changes can be submitted for IRB review using the Addendum/Modification Request Form. An Addendum/Modification Request Form **cannot** be used to extend the approval period of a study. All forms may be completed and submitted online at http://iris.research.ucf.edu .

Failure to provide a continuing review report could lead to study suspension, a loss of funding and/or publication possibilities, or reporting of noncompliance to sponsors or funding agencies. The IRB maintains the authority under 45 CFR 46.110(e) to observe or have a third party observe the consent process and the research.

On behalf of Tracy Dietz, Ph.D., UCF IRB Chair, this letter is signed by:

Signature applied by Joanne Muratori on 12/10/2008 09:06:06 AM EST

Joanne Muratori

IRB Coordinator

APPENDIX B: INFORMED CONSENT FORM

Participant Informed Consent Form

I agree to participate in the study "Empirical Evaluation of the Effectiveness of Eye-tracking-based Search Performance Diagnosis and Feedback Methods."

I verify that I am 18 years or older to participate in this study.

In this research study, I will participate in a study targeted at assessing the effectiveness of search performance diagnosis and feedback methods through the measurement of search knowledge and skills as demonstrated through performance of a military threat detection task and a knowledge test. I am aware that the experiment will consist of one pretraining session of approximately 10 minutes on the ObSERVE training system: the session will focus on search strategies and threats included in the training scenarios. I am aware that following this I will perform training session lasting approximately 30 minutes during which, I will search the simulated environment for military threats. While participating in the study, I will use a laptop or desktop computer, a mouse and a keyboard to control my actions within the simulation environment. I am aware that an eye-tracker will be used to determine where I am looking within the simulation. I am aware that my performance with regard to training may be recorded by video camcorder and audio recordings for later data analyses (only) and immediately destroyed after data use. The performance on these tasks will remain completely confidential (see below). I am aware that I may be asked to discuss the search strategies utilized in performance of the task and rate the difficulty of the scenarios. I understand that I will either receive 1 extra credit unit per hour of participation or $10/hour for my participation in this study regardless of if I complete the study or not.

All data I will contribute to this study will be held in strict confidentiality by the researchers. That is, my individual data will not be revealed to anyone other than the researchers and their immediate assistants.

To insure confidentiality, the following steps will be taken: (a) only researchers will have access to the data in paper or electronic form. Data and consent forms will be stored in separate locked cabinets under the control of the principal investigator; (b) the actual forms will not contain names or other personal information. Instead, a number assigned by and only known to the experimenters will match the forms to each participant; (c) only group means scores and standard deviations, but not individual scores, will be published or reported.

MY PARTICIPATION IN THIS RESEARCH IS COMPLETELY VOLUNTARY. I CAN WITHDRAW MY PARTICIPATION AT ANY TIME WITHOUT PENALTY OR PERJURY. THIS INCLUDES REMOVAL/DELETION OF ANY DATA I MAY HAVE CONTRIBUTED. I DO NOT HAVE TO ANSWER ANY QUESTION THAT I DON'T WANT TO.

This research study has been reviewed and approved by the UCF Institutional Review Board. Questions or concerns regarding your rights as a research volunteer may be directed to the UCF IRB office.

IRB Coordinator
Institutional Review Board (IRB)
University of Central Florida (UCF)
12201 Research Parkway, Suite 501
Orlando, Florida 32826-3246
Telephone: (407) 823-2901

I have been given the opportunity to ask the research assistants any questions I may have. For any other questions regarding this research, I can contact Meredith Carroll.

Meredith Carroll	Email: meredith@designinteractive.net	Phone: (407) 706-0977 ext. 31 Fax: (407) 977-0980
Dr. Mustapha Mouloua	Email: mouloua@pegasus.cc.ucf.edu	Phone: (407) 823-2910 Fax: (407) 823-5862

I have read the procedure described above. I voluntarily agree to participate in the procedure and I have received a copy of this description.

Print Name: ______________________________

Signature: ______________________________ Date: ____________

UCF University of Central Florida IRB
IRB NUMBER: SBE-08-05964
IRB APPROVAL DATE: 12/9/2008
IRB EXPIRATION DATE: 12/8/2009

APPENDIX C: EXPERIMENT 1 DEMOGRAPHICS QUESTIONNAIRE

Background Information

Please complete the following questions. Any information you provide is voluntary and will be kept strictly confidential. A participant number will be assigned to your responses and in no way will your name be associated with this data. The information you provide will be used only for the purposes of this study.

1. Gender: _______ Male _______ Female

2. Age: _______

3. Handedness (check one)? _____Left-handed _____Right-handed

4. Year in school: ___Freshman ___Sophomore ___Junior ___Senior

5. Major: _________________________

6. In general how do you feel about working with computers?

________ I don't like working with computers.
________ I have no strong like or dislike for working with computers.
________ I like working with computers.
________ Other (please explain)
____

7. How would you describe your general level of computer experience?

________ ***None*** (I have never used any computer applications).
________ ***Low*** (I have used only 1 or 2 computer applications).
________ ***Moderately Low*** (I have learned and used between 3 and 10 different computer applications).
________ ***Moderately High*** (I have learned and used more than 10 different computer applications but have no programming skills).
________ ***High*** (I have used many different computer applications and have some programming skills).
________ ***Other*** (please explain)
____

8. Have you ever been in a virtual environment (VE)? YES_____ NO____

   If **YES**, how many times have you been in a VE? _________________

9. How would you describe your general level of gaming experience (i.e., playing video games)?

________ ***None*** (I have never played a video game).

________ ***Low*** (I have played a video game a few times in the past).
________ ***Moderately Low*** (I have played a video game a regularly in the past).
________ ***Moderately High*** (I currently play video games weekly).
________ ***High*** (I currently play video games daily).
________ ***Other*** (please explain)
____

10. What were your SAT scores: Math __________ Verbal: ______

11. Thinking only about the last 90 days, how often, on average, do you play console based video games in a given week?
    a. Under 3 hours
    b. 3 - 9 hours
    c. 10 - 16 hours
    d. 17+ hours

12. About how many of your total gaming hours are spent playing first-person shooter style video games such as Halo, Medal of Honor, and Splinter Cell?
    ____

13. Have you ever received any formal military training?
    a. Yes
    b. No

    If Yes, please describe: ____

14. How would you describe your general level of hunting experience?

________ ***None*** (I have never hunted).
________ ***Low*** (I have hunted a few times in the past).
________ ***Moderately Low*** (I have hunted regularly in the past).
________ ***Moderately High*** (I currently hunt weekly).
________ ***High*** (I currently hunt daily).
________ ***Other*** (please explain)
____

15. How would you describe the neighborhood in which you grew up?

________ ***Not Dangerous*** (No concern for safety at any time).
________ ***Minimally Dangerous*** (Occasional concern for safety).
________ ***Moderately Dangerous*** (Concern for safety some of the time).
________ ***Very Dangerous*** (Concern for safety at all times).
________ ***Other*** (please explain)
____

APPENDIX D: EXPERIMENT 2 DEMOGRAPHICS QUESTIONNAIRE

Demographics Questionnaire

Please complete the following questions. Any information you provide is voluntary and will be kept strictly confidential. A participant number will be assigned to your responses and in no way will your name be associated with this data. The information you provide will be used only for the purposes of this study.

1. Gender: _______ Male _______ Female

2. Age: _______

3. Handedness - Dominant (check one)? _____Left-handed _____Right-handed

4. Please indicate the highest level of education completed:

   - ☐ Grammar school
   - ☐ High school or equivalent
   - ☐ Some college
   - ☐ College Graduate (4 yrs)
   - ☐ Master's Degree (MS)
   - ☐ Doctoral Degree (PhD)
   - ☐ Professional Degree (MD,JD, etc.)

5. Major: _________________________

6. Do you have normal or corrected-to-normal hearing:

   - ☐ Normal/Corrected Hearing
   - ☐ Hearing Problems (please describe) _______________

7. Do you have normal or corrected-to-normal vision:

   - ☐ Normal
   - ☐ Corrected (**Circle One**: glasses / contacts)
   - ☐ Problems
     Please describe___________________

8. Do you have astigmatism in:

   - ☐ Left eye
   - ☐ Right eye
   - ☐ Both eyes
   - ☐ Neither eye

9. If you have astigmatism in one or both eyes, do you wear a toric contact lens or have specially shaped glasses lenses to correct for this in the affected eyes?

   - ☐ I wear a toric contact lens in the affected eye

☐ I have specially shaped glasses for the affected eye

10. Eye Color
   ☐ Blue
   ☐ Green
   ☐ Hazel
   ☐ Brown

11. Have you had any previous eye injuries?
   ☐ Yes (Describe: ________________________________ Approximate date:__________)
   ☐ No

12. Color-blindness
   ☐ Yes (Describe: ___________________)
   ☐ No

13. What were your SAT scores:

   Math ______ Verbal: ______ Not sure / Don't remember: ______

14. In general how do you feel about working with computers?

________ I don't like working with computers.
________ I have no strong like or dislike for working with computers.
________ I like working with computers.
________ Other (please explain) ____

15. How would you describe your general level of computer experience?

________ ***None*** (I have never used any computer applications).
________ ***Low*** (I have used only 1 or 2 computer applications).
________ ***Moderately Low*** (I have learned and used between 3 and 10 different computer applications).
________ ***Moderately High*** (I have learned and used more than 10 different computer applications but have no programming skills).
________ ***High*** (I have used many different computer applications and have some programming skills).
________ ***Other*** (please explain) ____

16. Have you ever been in a virtual environment (VE)? YES_____ NO____

   If **YES**, how many times have you been in a VE? _________________

17. How would you describe your general level of gaming experience (i.e., playing video games)?

________ ***None*** (I have never played a video game).
________ ***Low*** (I have played a video game a few times in the past).
________ ***Moderately Low*** (I have played a video game a regularly in the past).
________ ***Moderately High*** (I currently play video games weekly).

________ ***High*** (I currently play video games daily).
________ ***Other*** (please explain) ____

18. Thinking only about the last 90 days, how often, on average, do you play console based video games in a given week?

e. Under 3 hours
f. 3 - 9 hours
g. 10 - 16 hours
h. 17+ hours

19. About how many of your total gaming hours are spent playing first-person shooter style video games such as Halo, Medal of Honor, and Splinter Cell?

____

20. Have you ever received any formal military training?

c. Yes
d. No

If Yes, please describe: ____

____

21. How would you describe your general level of hunting experience?

________ ***None*** (I have never hunted).
________ ***Low*** (I have hunted a few times in the past).
________ ***Moderately Low*** (I have hunted regularly in the past).
________ ***Moderately High*** (I currently hunt weekly).
________ ***High*** (I currently hunt daily).
________ ***Other*** (please explain) ____

22. How would you describe the neighborhood in which you grew up?

________ ***Not Dangerous*** (No concern for safety at any time).
________ ***Minimally Dangerous*** (Occasional concern for safety).
________ ***Moderately Dangerous*** (Concern for safety some of the time).
________ ***Very Dangerous*** (Concern for safety at all times).
________ ***Other*** (please explain) ____

APPENDIX E: VISUAL ACUITY TEST

70 ft - 21 m	G
60 ft - 18 m	WV
50 ft - 15 m	G S B E
40 ft - 12 m	N O I H W
30 ft - 9 m	J H E R L C
20 ft - 6 m	N O S Z L E P H
15 ft - 4.5 m	U L Y T H B X P G O
10 ft - 3 m	S W M B W G C P T T
7 ft - 2.1 m	O H D C W N Y Z W A V
4 ft - 1.2 m	H N U O C I C R T W W D Q M V B F

APPENDIX F: COLOR BLINDNESS TEST

Color Vision Test Instructions

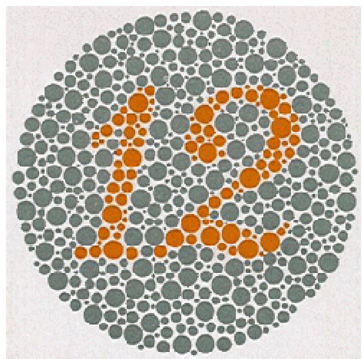
- In the following task, you will be presented with a number of questions that assess your ability to perceive numbers embedded within patterns.
- For each question, you will be asked to indicate what *number* you see revealed in the patterns of dots inside the picture.
  - If you do ***not*** see a number inside the pattern of dots, then write "NONE" on the answer sheet next to that question.
- There are a total of 12 questions. As you complete each question, record your response on the answer sheet provided.
- Should you have questions about this task, please feel free to ask for assistance at any time.
- Please do not write on the test booklet.

Please turn the page to continue. . .

Color Vision Test
Sample Item

Here is a sample item of the task you will perform. Please look at the picture below. What number do you see revealed in the pattern of dots below?

Sample Item



You should see the number "12" inside the pattern of dots. So, you would write "12" on the answer sheet on the space next to that question. Please make sure to complete all items. And please remember not to make any marks on the test booklet.

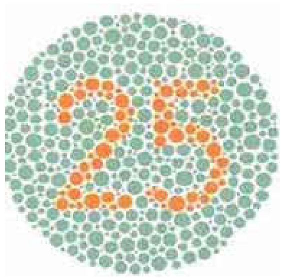
If you have any questions, please ask now. Otherwise, let the experimenter know that you are ready to begin.

Please wait until the instruction is given before turning the page to begin. . .

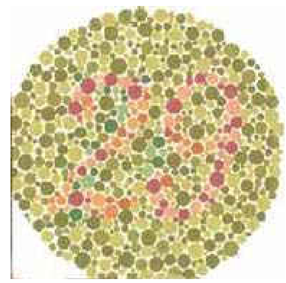
Color Vision Test

What numbers do you see revealed in the patterns of dots below? Please record the number on the answer sheet or, if you do not see a number, write "NONE."

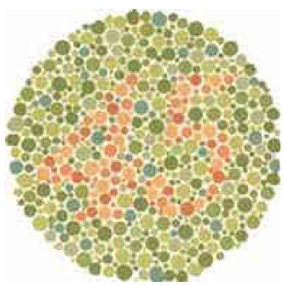
Question 1



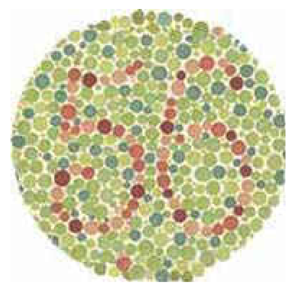
Question 2



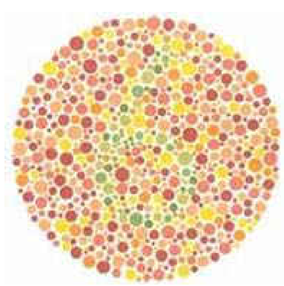
Question 3



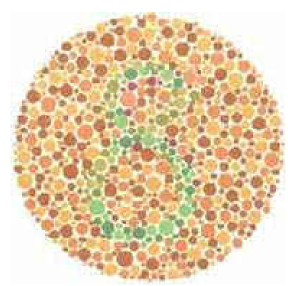
Question 4



Question 5



Question 6

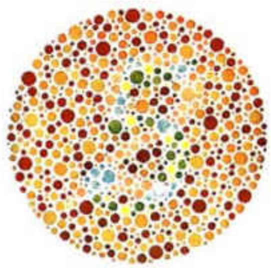


Please turn the page to continue. . .

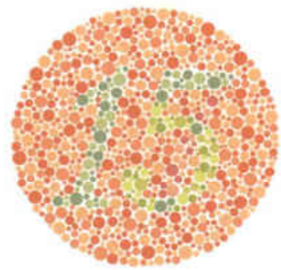
Color Vision Test

What numbers do you see revealed in the patterns of dots below? Please record the number on the answer sheet or, if you do not see a number, write "NONE."

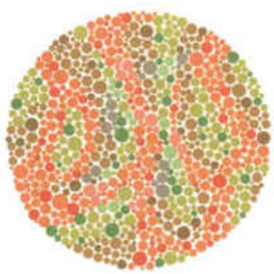
Question 7



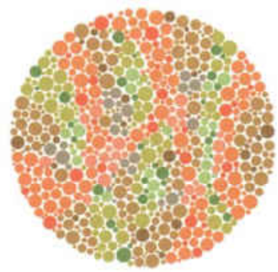
Question 8



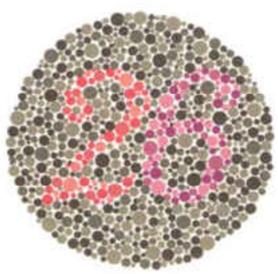
Question 9



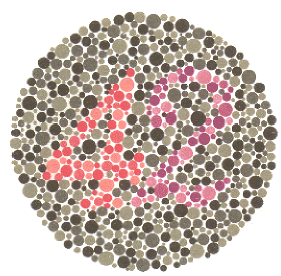
Question 10



Question 11



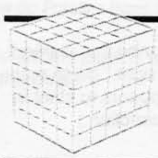
Question 12



Please stop here.

APPENDIX G: SPATIAL ORIENTATION TEST

The Guilford-Zimmerman Aptitude Survey



Part 5/Spatial Orientation

Name ________________ Date ________ Score ________ Sex: M F

INSTRUCTIONS.

This is a test of your ability to see changes in direction and position. In each item you are to note how the position of the boat has changed in the second picture from the original position in the first picture.

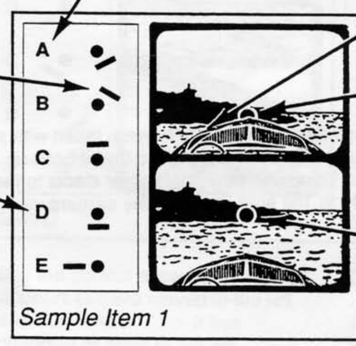
Here is Sample Item 1.

These are the five possible answers to the item.

These bars represent the boat's prow.

This is the correct answer. It shows that the prow of the boat has dropped below the aiming point.

(If the prow had risen, instead of dropped, the correct answer would have been C, instead of D.)



Sample Item 1

This is the prow (front end) of a motor boat in which you are riding.

This is the aiming point. It is the exact spot you would see on land if you sighted right over the point of the prow.

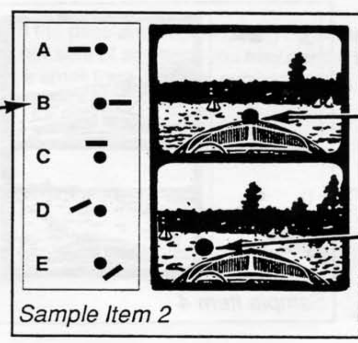
This is the same aiming point shown above. Note that the prow has dropped below it.

To work each item: **First**, look at the top picture and see where the motor boat is headed. **Second**, look at the bottom picture and note the CHANGE in the boat's heading. **Third**, mark the answer that shows the same change on the separate answer sheet.

Try Sample Item 2.

This also shows that the prow of the boat is to the right of the aiming point. So, it is the correct answer.

(If the boat had turned to the left, instead of to the right, the correct answer would have been A.)



Sample Item 2

This is the aiming point.

This is the same aiming point. The motor boat is now headed to the right of it.

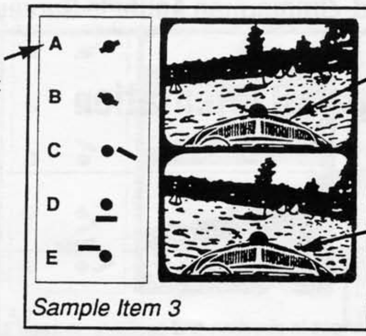
CPP Consulting Psychologists Press, Inc., 3803 E. Bayshore Road, Palo Alto, California 94303

Printed in the United States of America. 03 02 01 00 99 14 13 12 11

0039

Now try Sample Item 3.

This is the correct answer. It shows that the motor boat changed its slant to the left, but is still heading toward the aiming point.



Sample Item 3

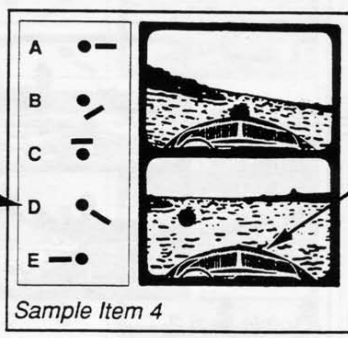
Here the motor boat is slanted slightly to the right. (Note that the horizon appears to slant in the opposite direction.)

Here the boat has changed its slant toward the left. (To become level, the boat slanted back toward the right.)

Imagine that these pictures were taken with a motion picture camera. The camera is fastened rigidly to the boat so that it bobs up and down and turns and slants with the boat. Thus, when the boat tips or slants to the left (as in the lower sample in SAMPLE ITEM 3), the scene through the camera view finder looks slanted like this.

Look at Sample Item 4.

D is the correct answer. It shows that the boat changed its heading both downward and to the right; also that it changed its slant toward the right.



Sample Item 4

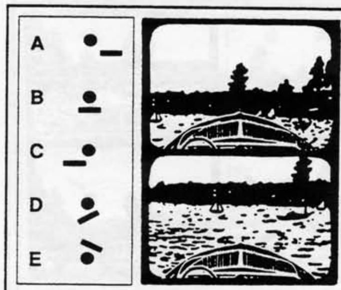
The prow of the boat has moved downward and toward the right. Also, it has changed its slant toward the right.

Now do Practice Items 5, 6, and 7. Record your answers on the separate answer sheet.

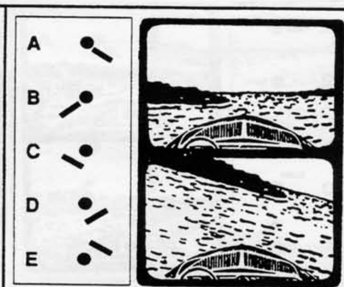
The aiming point is not marked in the test items. You must see the change in the boat's position without the aid of the dots.

To review:

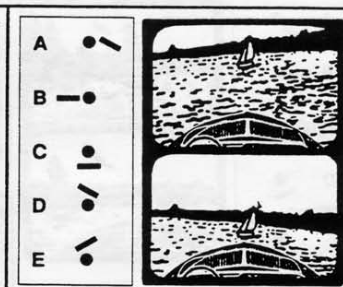
First – Look at the top picture. See where the motor boat is headed.
Second – Look at the bottom picture. Note the change in the boat's heading.
Third – Mark the answer that shows the same change (in reference to the aiming point before the change).



Item 5



Item 6



Item 7

C is the correct answer. The prow appears to have moved to the left and downward. It has not changed its slant.

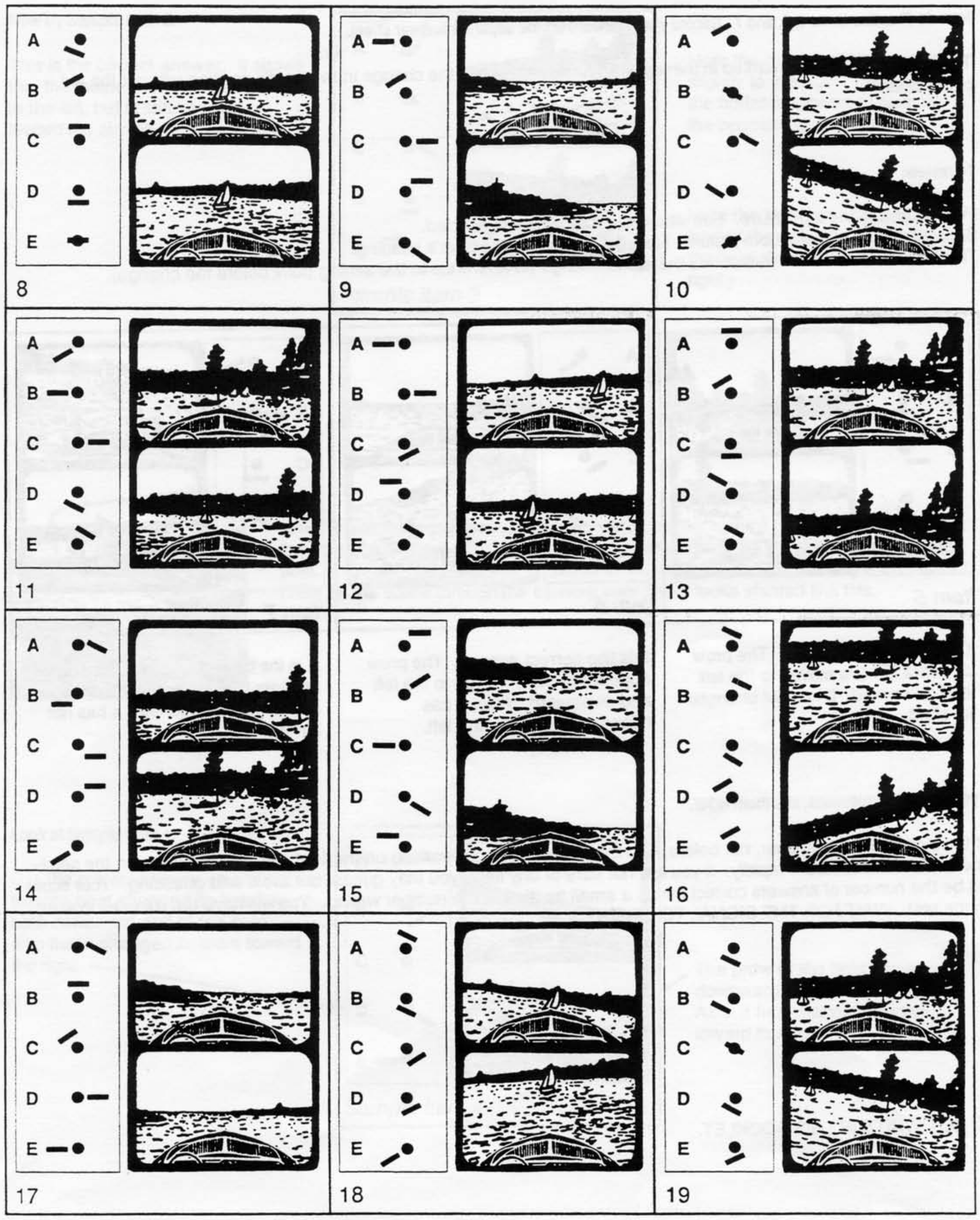
B is the correct answer. The prow appears to have moved to the left and downward. Also, it has changed its slant to the left.

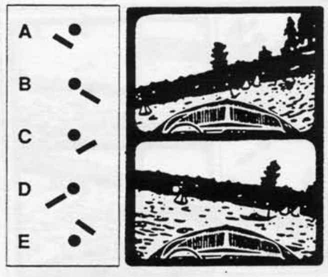
E is the correct answer. The prow appears to have moved upward, and to have tipped left. It has not turned.

If you have any questions, ask them NOW.

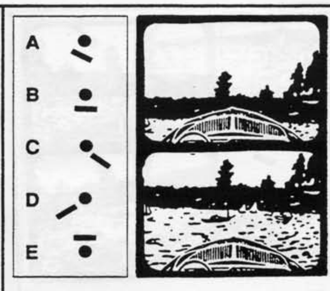
At the signal of the examiner, not before, turn the page and begin working on the test. Mark all answers on the separate answer sheet. Work rapidly. If you are not sure of any item, you may guess, but avoid wild guessing. Your score will be the number of answers correct minus a small fraction of the number wrong. You will have ten minutes to work on the test. WAIT FOR THE SIGNAL TO BEGIN.

DO NOT WRITE IN THIS BOOKLET.

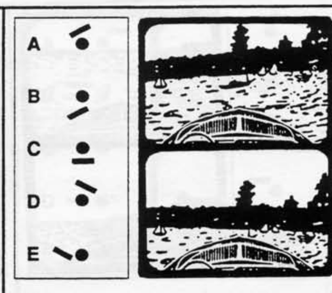

8
9
10
11
12
13
14
15
16
17
18
19
A
B
C
D
E



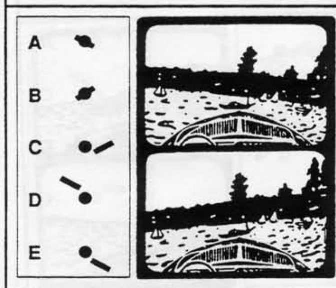
20



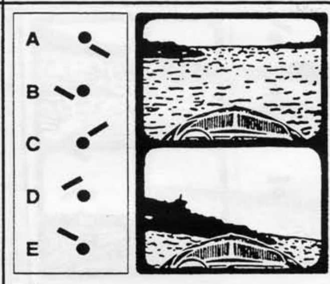
21



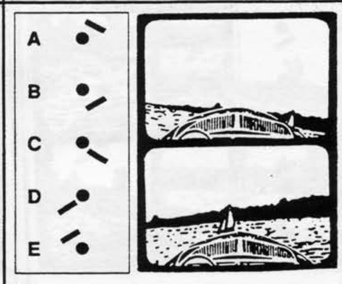
22



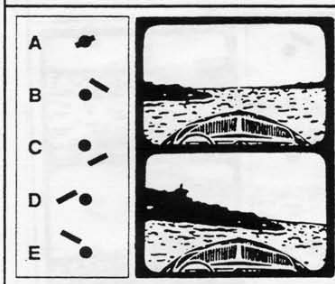
23



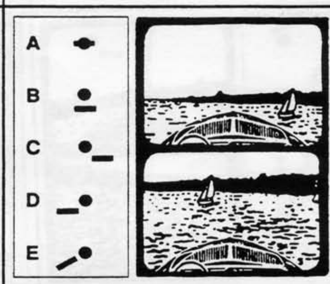
24



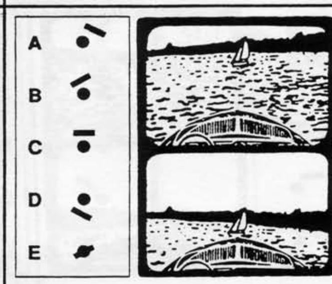
25



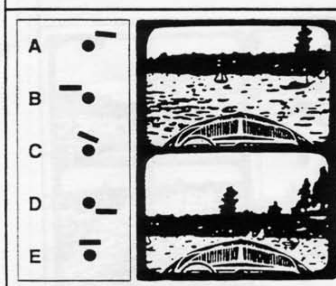
26



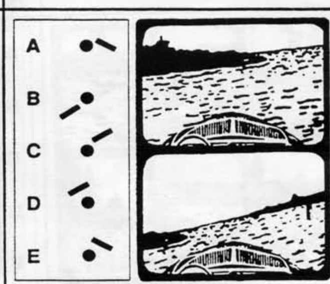
27



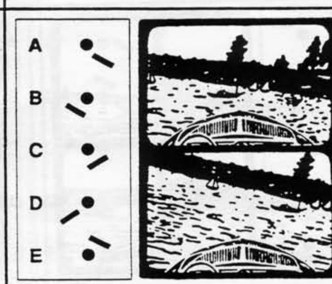
28



29

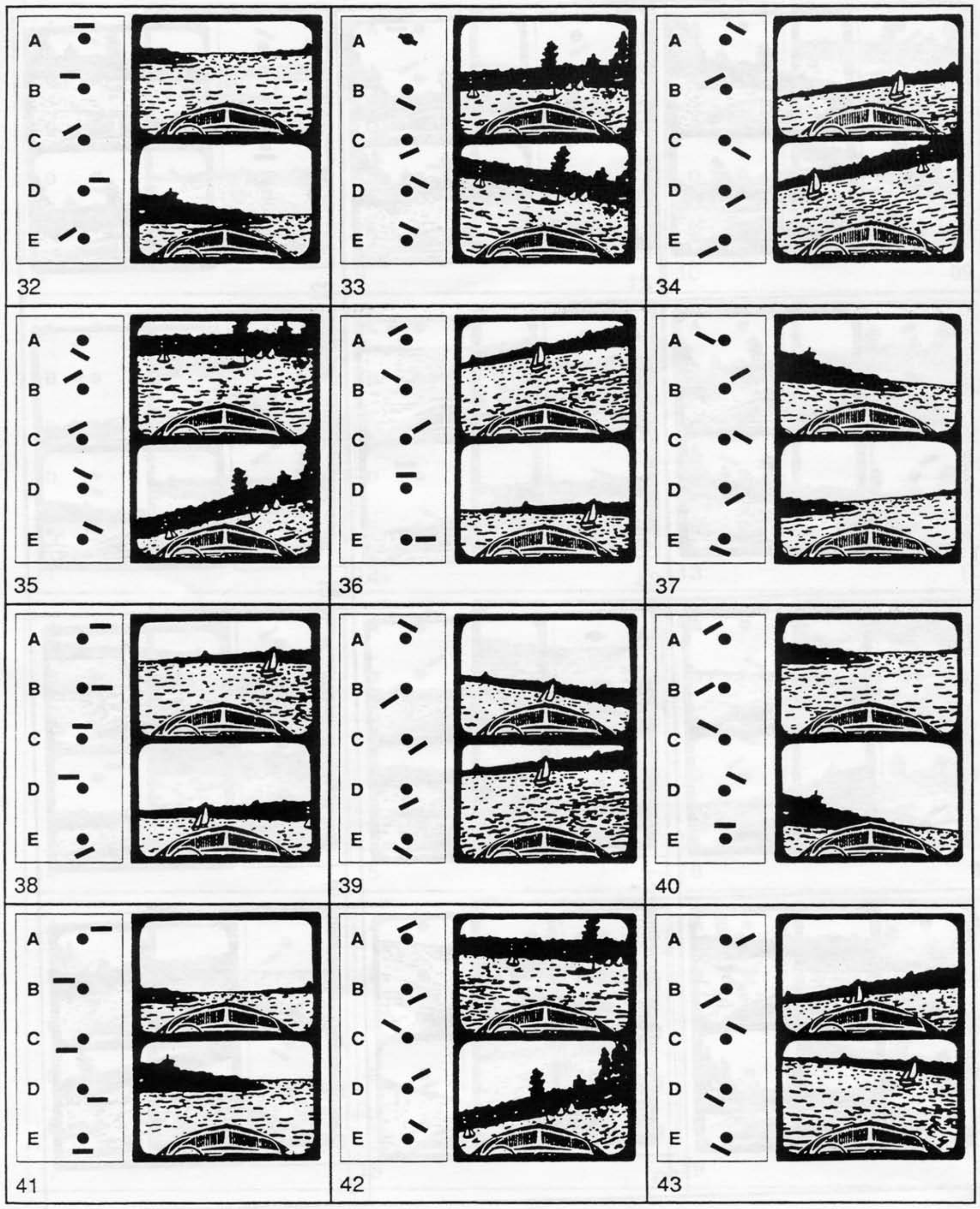


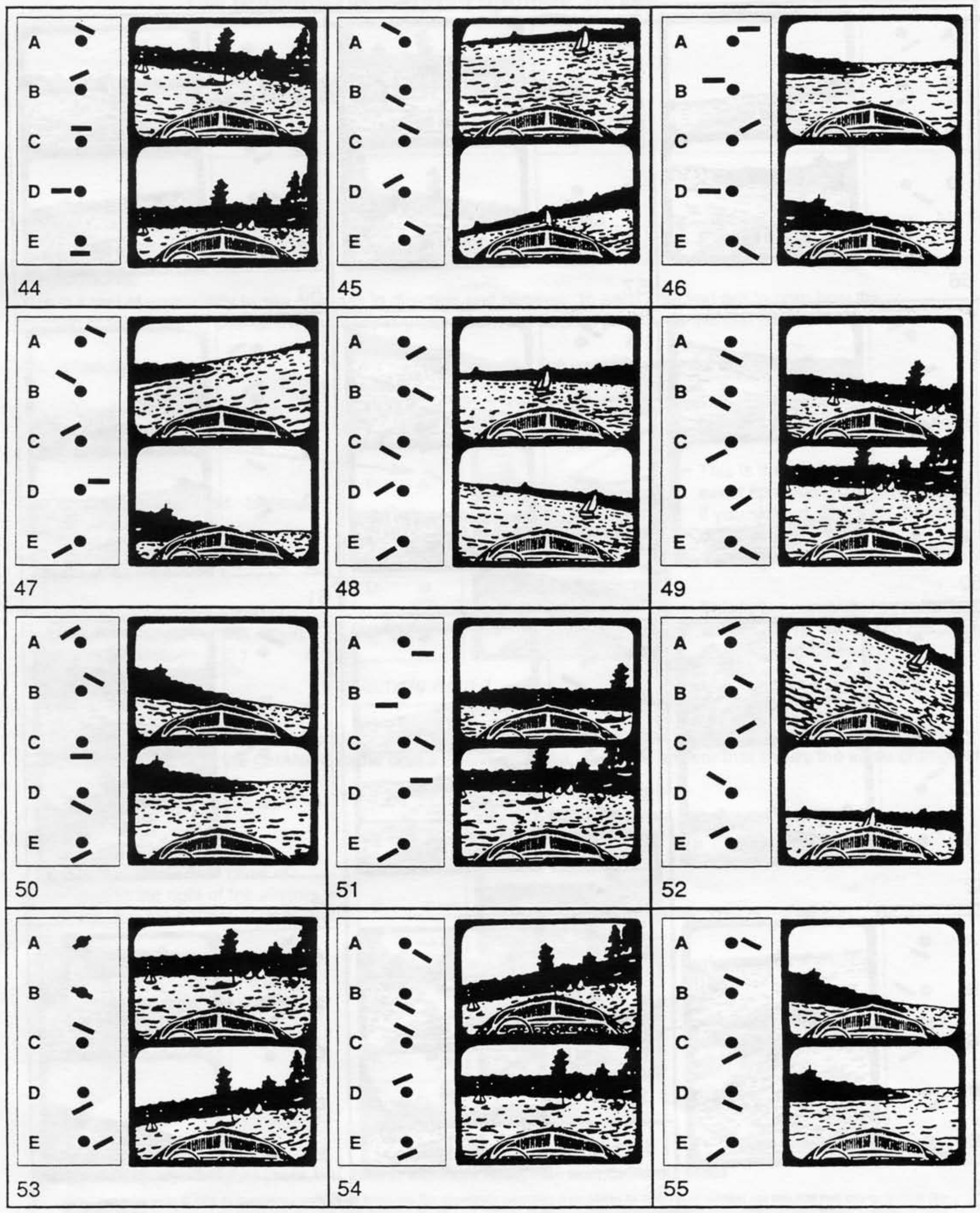
30



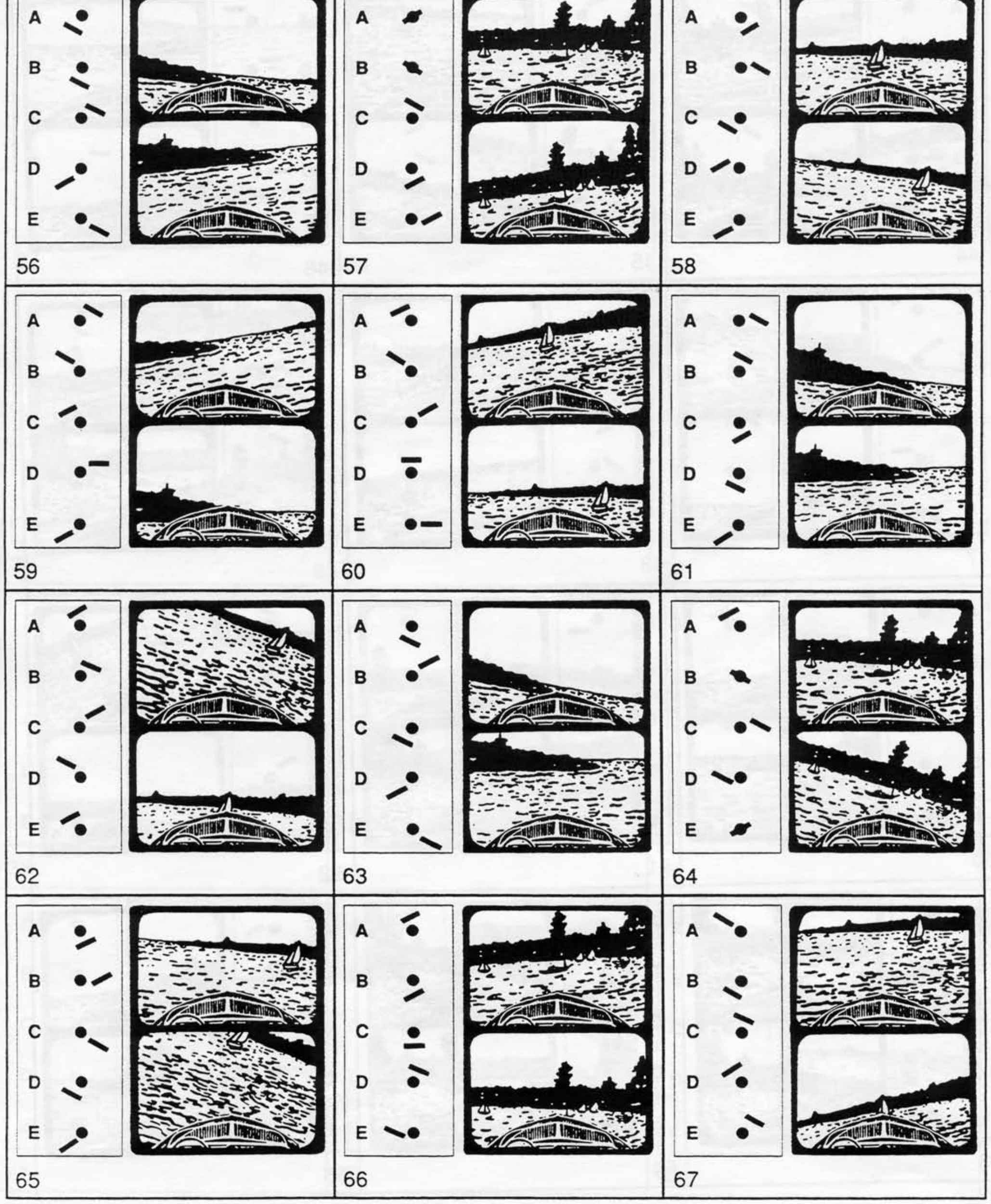
31

(Continue)


32
33
34
35
36
37
38
39
40
41
42
43



(Continue)


A B C D E
56
A B C D E
57
A B C D E
58
A B C D E
59
A B C D E
60
A B C D E
61
A B C D E
62
A B C D E
63
A B C D E
64
A B C D E
65
A B C D E
66
A B C D E
67

APPENDIX H: VISUALIZATION TEST

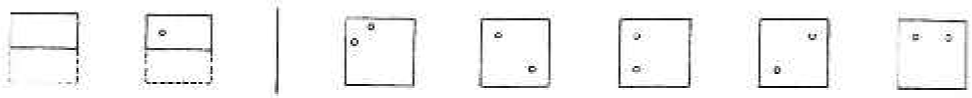
Name ______________________________

PAPER FOLDING TEST — VZ-2

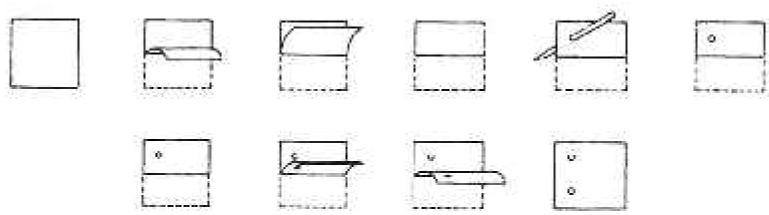
In this test you are to imagine the folding and unfolding of pieces of paper. In each problem in the test there are some figures drawn at the left of a vertical line and there are others drawn at the right of the line. The figures at the left represent a square piece of paper being folded, and the last of these figures has one or two small circles drawn on it to show where the paper has been punched. Each hole is punched through all the thicknesses of paper at that point. One of the five figures at the right of the vertical line shows where the holes will be when the paper is completely unfolded. You are to decide which one of these figures is correct and draw an X through that figure.

Now try the sample problem below. (In this problem only one hole was punched in the folded paper.)

A B C D E



The correct answer to the sample problem above is C and so it should have been marked with an X. The figures below show how the paper was folded and why C is the correct answer.



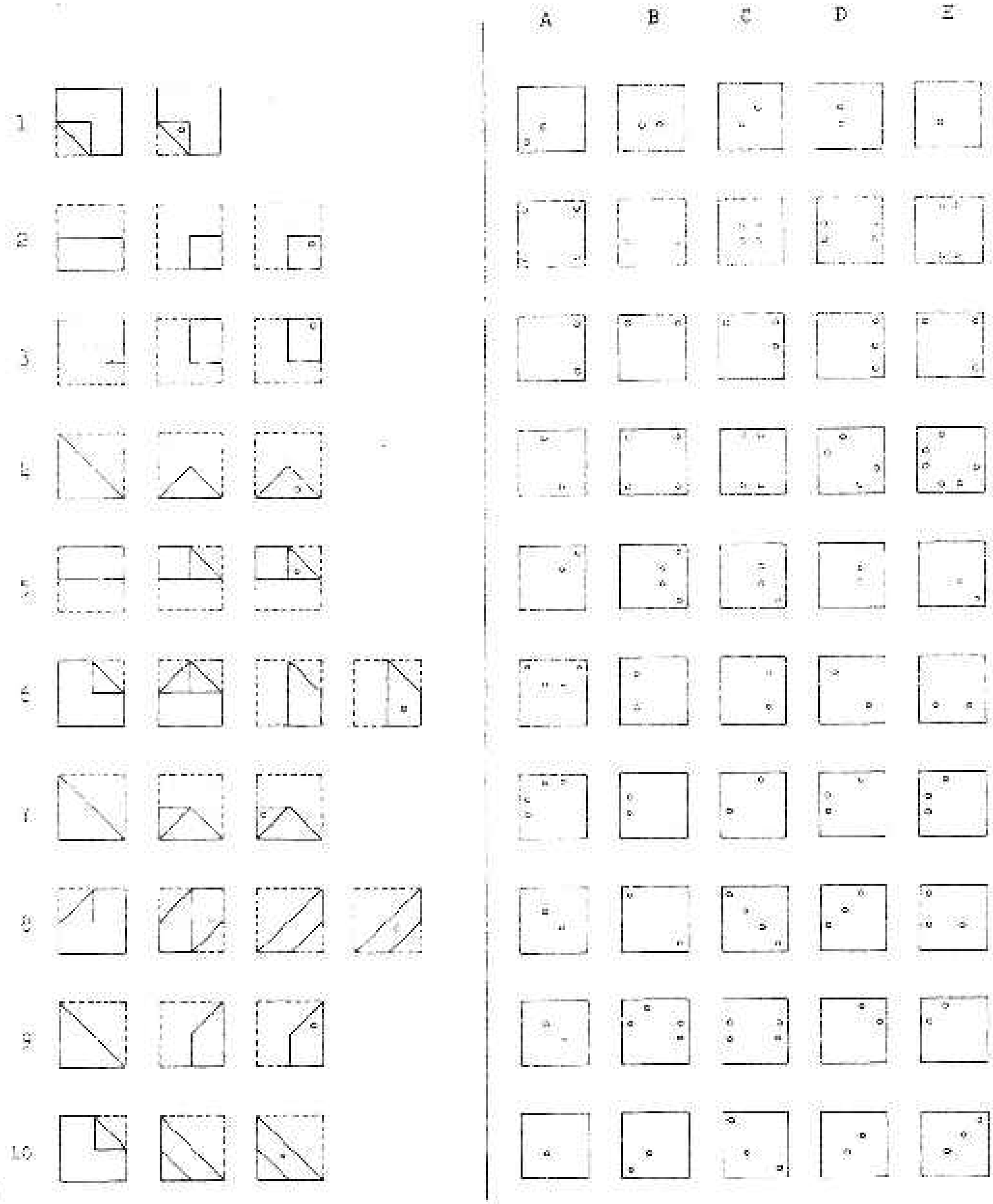
In these problems all of the folds that are made are shown in the figures at the left of the line, and the paper is not turned or moved in any way except to make the folds shown in the figures. Remember, the answer is the figure that shows the positions of the holes when the paper is completely unfolded.

Your score on this test will be the number marked correctly minus a fraction of the number marked incorrectly. Therefore, it will <u>not</u> be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

You will have <u>3 minutes</u> for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, <u>STOP</u>. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

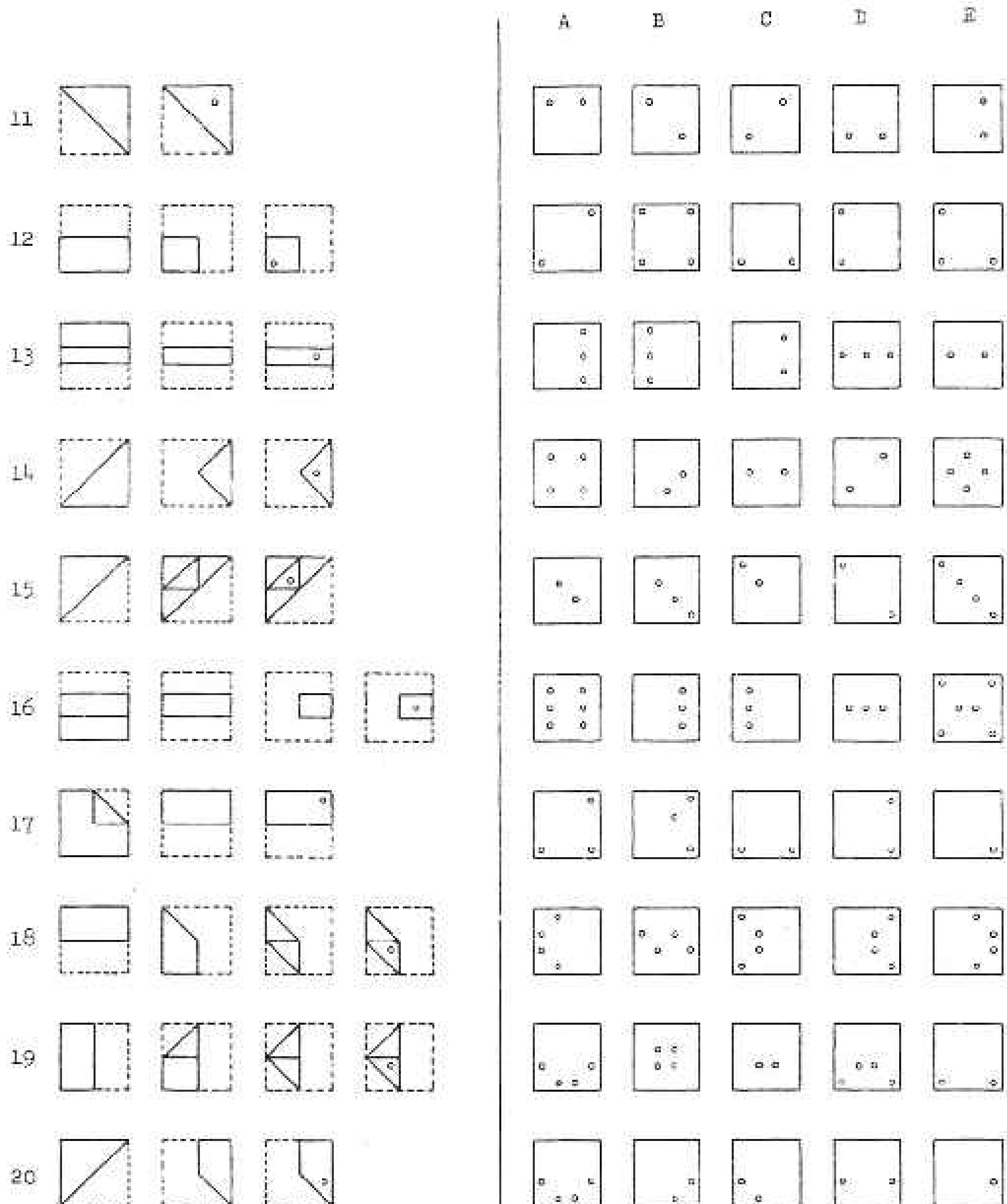
Part 1 (3 minutes)



DO NOT GO ON TO THE NEXT PAGE UNTIL ASKED TO DO SO.

STOP.

Part 2 (3 minutes)



DO NOT GO BACK TO PART 1, AND

DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO. STOP.

APPENDIX I: HIDDEN FIGURES TEST

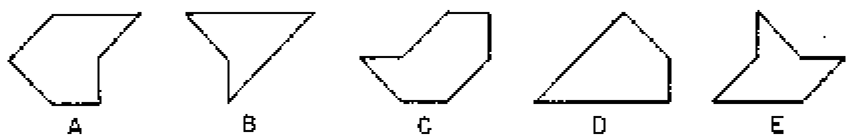
Name ______________________________

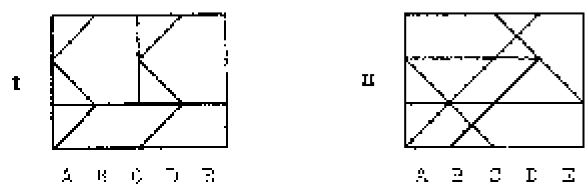
HIDDEN FIGURES TEST — CF-1 (Rev.)

This is a test of your ability to tell which one of five simple figures can be found in a more complex pattern. At the top of each page in this test are five simple figures lettered A, B, C, D, and E. Beneath each row of figures is a page of patterns. Each pattern has a row of letters beneath it. Indicate your answer by putting an X through the letter of the figure which you find in the pattern.

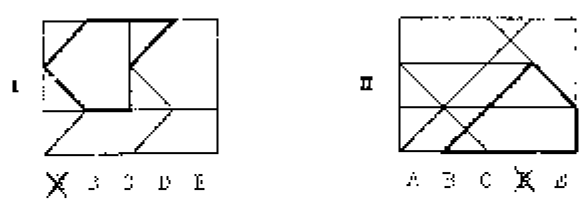
NOTE: There is only one of these figures in each pattern, and this figure will always be right side up and exactly the same size as one of the five lettered figures.

Now try these 2 examples.





The figures below show how the figures are included in the problems. Figure A is in the first problem and figure D is in the second.



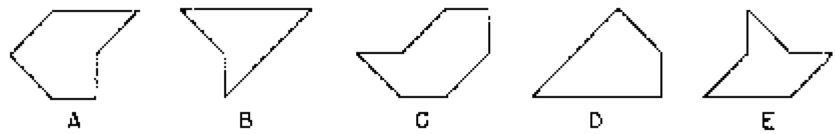
Your score on this test will be the number marked correctly minus a fraction of the number marked incorrectly. Therefore, it will not be to your advantage to guess unless you are able to eliminate one or more of the answer choices as wrong.

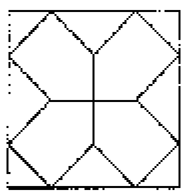
You will have 12 minutes for each of the two parts of this test. Each part has 2 pages. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

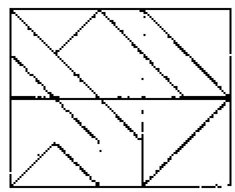
Page 2

Part 1 (12 minutes)

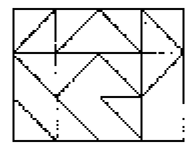


1. 

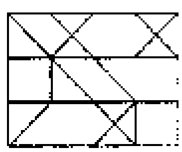
A B C D E

2. 

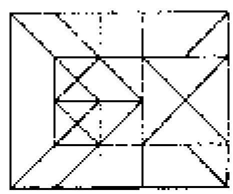
A B C D E

3. 

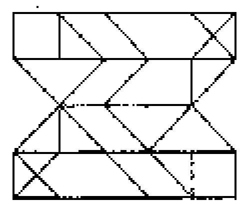
A B C D E

4. 

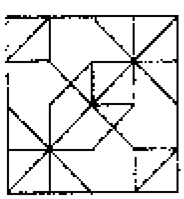
A B C D E

5. 

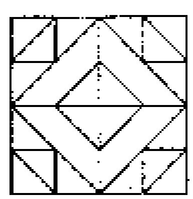
A B C D E

6. 

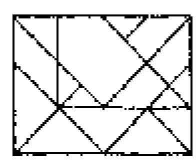
A B C D E

7. 

A B C D E

8. 

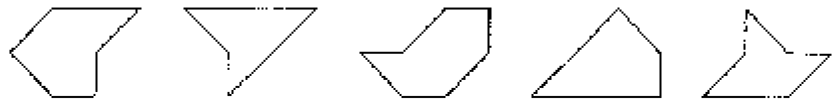
A B C D E

9. 

A B C D E

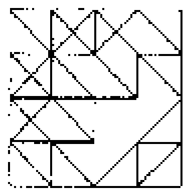
GO ON TO THE NEXT PAGE

Part 1 (continued)



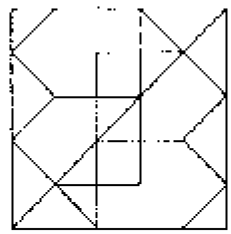
A B C D E

10.



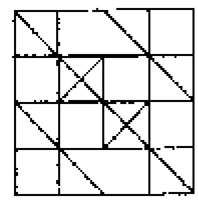
A B C D E

11.



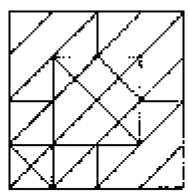
A B C D E

12.



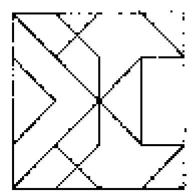
A B C D E

13.



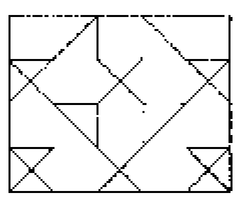
A B C D E

14.



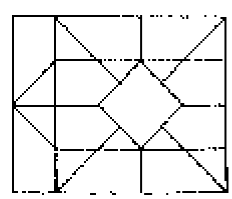
A B C D E

15.



A B C D E

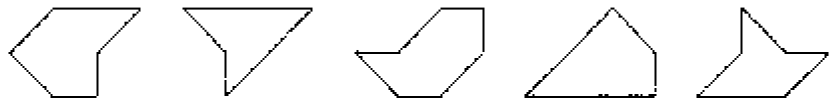
16.



A B C D E

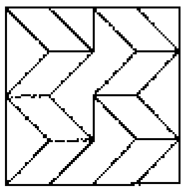
DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO STOP.

Part 2 (12 minutes)



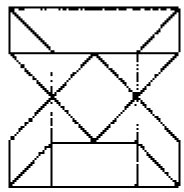
A B C D E

17.



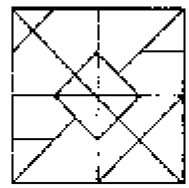
A B C D E

18.



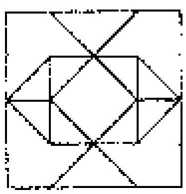
A B C D E

19.



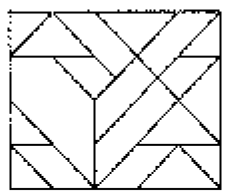
A B C D E

20.



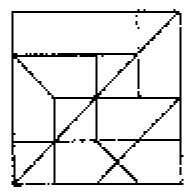
A B C D E

21.



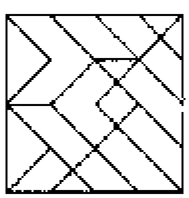
A B C D E

22.



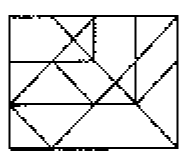
A B C D E

23.



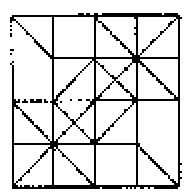
A B C D E

24.



A B C D E

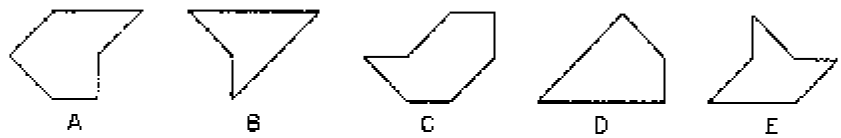
25.

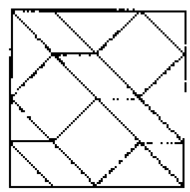


A B C D E

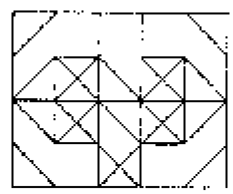
GO ON TO THE NEXT PAGE

Part 2 (continued)

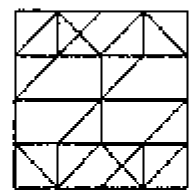


26. 

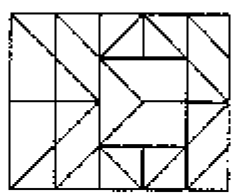
A B C D E

27. 

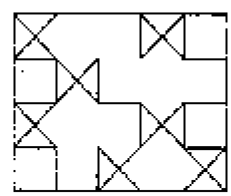
A B C D E

28. 

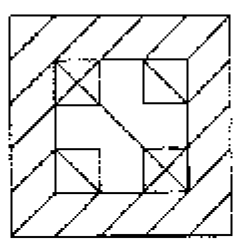
A B C D E

29. 

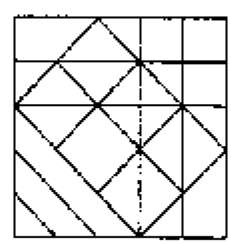
A B C D E

30. 

A B C D E

31. 

A B C D E

32. 

A B C D E

DO NOT GO BACK TO PART 1, AND
DO NOT GO ON TO ANY OTHER TEST UNTIL ASKED TO DO SO.

STOP.

APPENDIX J: COGNITIVE LOAD QUESTIONNAIRE

Cognitive Load Question

While searching for the threats, I invested (circle one only)
1. Very, very low mental effort
2. Very low mental effort
3. Low mental effort
4. Neither low nor high mental effort
5. High mental effort
6. Very high mental effort
7. Very, very high mental effort

APPENDIX K: VISUAL VERBAL LEARNING STYLE QUESTIONNAIRE

Verbal-Visual Learning Style Rating (VVLSR, Version 1.0, 2004)

In a learning situation sometimes information is presented verbally (e.g., with printed or spoken words) and sometimes information is presented visually (e.g., with labeled illustrations, graphs, or narrated animations). Please place a check mark indicating your learning preference.

○	○	○	○	○	○	○
Strongly more verbal than visual	Moderately more verbal than visual	Slightly more verbal than visual	Equally verbal and and visual	Slightly more visual than verbal	Moderately more visual than verbal	Strongly more visual than verbal

Validation of a One-Item Test of Verbalizer-Visualizer Cognitive Style
Richard E. Mayer and Laura J. Massa
University of California, Santa Barbara
2004

APPENDIX L: EXPERIMENT 2 KNOWLEDGE PRE TEST

Observation Knowledge Pre Test

Based on what you learned from the pre-training you just completed, please classify each type of object as low or high priority by circling the appropriate category. Then, indicate whether each type of object provides cover, concealment, a concealed exit or all three by checking the appropriate categories with a check mark.

Object Type	High or Low Priority	Provides Cover	Provides Concealment	Provides a Concealed Exit
Alleys	High / Low			
Vehicles	High / Low			
Trash	High / Low			
Rooftops	High / Low			
Humvees	High / Low			
Faraway Buildings	High / Low			
Women in Open	High / Low			
Doors	High / Low			
Barriers	High / Low			
Vegetation	High / Low			
Store Fronts	High / Low			
Tables & Chairs	High / Low			
Baskets	High / Low			
Windows	High / Low			
Men on Roof	High / Low			
Trashcans & Barrels	High / Low			
Balconies& Porches Without Walls	High / Low			
Fruit Stands	High / Low			
Concrete Balcony	High / Low			
Sticks	High / Low			

APPENDIX M: EXPERIMENT 2 KNOWLEDGE POST TEST

Observation Knowledge Post Test

Based on what you learned from the training you just completed, please classify each type of object as low or high priority by circling the appropriate category. Then, indicate whether each type of object provides cover, concealment, a concealed exit or all three by checking the appropriate category with a check mark.

Object Type	High or Low Priority	Provides Cover	Provides Concealment	Provides a Concealed Exit
Fruit Stands	High / Low			
Windows	High / Low			
Concrete Balcony	High / Low			
Vegetation	High / Low			
Doors	High / Low			
Women in Open	High / Low			
Rooftops	High / Low			
Trashcans & Barrels	High / Low			
Faraway Buildings	High / Low			
Sticks	High / Low			
Men on Roof	High / Low			
Alleys	High / Low			
Tables & Chairs	High / Low			
Balconies& Porches Without Walls	High / Low			
Vehicles	High / Low			
Trash	High / Low			
Baskets	High / Low			
Humvees	High / Low			
Store Fronts	High / Low			
Barriers	High / Low			

APPENDIX N: EXPERIMENT 2 PARTICIPANT FEEDBACK FORM

Feedback Questionnaire

1) Rate your confidence level in being able to perform a scan for a sniper threat after the training you just experienced

Not at all Confident	1	2	3	4	5	6	7	Very Confident

Please respond to the questions below by circling the appropriate response

2) The training taught me how to successfully scan for a sniper threat

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

3) I became tired during the training

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

4) I could easily understand the training

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

5) I understood the purpose of the training and how it could improve my scanning strategy

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

6) I found the training module to be confusing

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

7) I was easily distracted during the training module

Strongly Disagree	1	2	3	4	5	6	7	Strongly Agree

Other comments/feedback:____

APPENDIX O: EXPERIMENT 1 DEBRIEF FORM

Experiment I
Debriefing Form

In this study, we are investigating the effectiveness of eye-tracking based search performance diagnosis methods for diagnosing where in the perceptual process performance breakdowns occur (i.e., search or detection errors). It is hypothesized that this diagnosis will be effective at pinpointed why threat detection errors occur.

If you would like to find out more about the results of this study, please email meredith@designinteractive.net after May 2009.

Thank you for your participation in this study!

APPENDIX P: EXPERIMENT 2 DEBRIEF FORM

Experiment II

Debriefing Form

In this study, we are investigating the effectiveness of eye-tracking based search performance diagnosis methods for diagnosing where in the perceptual process performance breakdowns occur (i.e., search or detection errors). It is hypothesized that this diagnosis will be effective at pinpointing why threat detection errors occur. The training and feedback you experienced are targeted at correcting these types of detection errors to improve search performance.

If you would like to find out more about the results of this study, please email meredith@designinteractive.net after May 2010.

Thank you for your participation in this study!

REFERENCES

Arani, T., Karwan, M. H., & Drury, C. G. (1984). A variable-memory model of visual search. *Hum Factors, 26*(6), 631-639.

Army, Department of Defense. (1992). The Infantry Reconnaissance Platoon and Squad (Airborne, Air Assault, Light Infantry) Manual. Deparment of the Army, Washington, DC.

Bloom, B. S. (1984). The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher, 13*(6), 4-16.

Boot, W. R., Becic, E., & Kramer, A. F. (2009). Stable individual differences in search strategy: The effect of task demands and motivational factors on scanning strategy in visual search. *Journal of Vision, 9*(3), 1-16.

Brannick, M. T., & Prince, C. (1997). An Overview of Team Performance Measurement. In *Team Performance Assessment and Measurement: Theory, Methods, and Applications* (pp. 3-14). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Burgert, O., Orn, V., Velichkovsky, B. M., Gessat, M., Joos, M., Strauss, G., et al. (2007). *Evaluation of perception performance in neck dissection planning using eye tracking and attention landscapes.* Paper presented at the Medical Imaging 2007: Image Perception, Observer Performance, and Technology Assessment, San Diego, CA, USA.

Burns, N. R., Nettelbeck, T., McPherson, J., & Stankov, L. (2007). Perceptual Learning on Inspection Time and Motion Perception. *Journal of General Psychology, 134*(1), 83-100.

Cannon-Bowers, J. A., Rhodenizer, L., Salas, E., & Bowers, C. A. (1998). A framework for understanding pre-practice conditions and their impact on learning. *Personnel Psychology, 51*(2), 291-320.

Carroll, M. B., Milham, L., Champney, R., Eitelman, S., & Lockerd, A. (2007). ObSERVE Initial Training Strategies Report. Program Interim Report, Contract No. W911QY-07-C-0084, Arlington, VA: Office of Naval Research.

Chabukswar, S., Gramopadhye, A. K., Melloy, B. J., & Grimes, L. W. (2003). Use of aiding and feedback in improving visual search performance for an inspection task. *Human Factors and Ergonomics in Manufacturing, 13*(2), 115-136.

Chapman, P., Underwood, G., & Roberts, K. (2002). Visual search patterns in trained and untrained novice drivers. *Transportation Research Part F: Psychology and Behaviour, 5*(2), 157-167.

Chun, M. M., & Jiang, Y. (1999). Top-Down Attentional Guidance Based On Implicit Learning of Visual Covariation. *Psychological Science, 10*(4), 360-365.

Chung, W., Chen, H., Chaboya, L. G., O'Toole, C. D., & Atabakhsh, H. (2005). Evaluating event visualization: a usability study of COPLINK spatio-temporal visualizer. *International Journal of Human-Computer Studies, 62*(1), 127-157.

Cleeremans, A., & McClelland, J. L. (1991). Learning the structure of event sequences. *Journal of Experimental Psychology, 120*(235-253).

Cox, B. D. (1997). The rediscovery of the active learner in adaptive contexts: A developmental-historical analysis of transfer of training. *Educational Psychologist, 32*(1), 41-55.

Crowley, R., Medvedeva, O., & Jukic, D. (2003). SlideTutor: A model-tracing Intelligent Tutoring System for teaching microscopic diagnosis. In U. Hoppe, F. Verdejo & J. Kay (Eds.), *Artificial Intelligence in Education: Shaping the Future of Learning Through Intelligent Technologies*: IOS Press.

Davis, W. D., Carson, C. M., Ammeter, A. P., & Treadway, D. C. (2005). The Interactive Effects of Goal Orientation and Feedback Specificity on Task Performance. *Human Performance, 18*(4), 409-426.

Doane, S. M., Alderton, D. L., Sohn, Y. W., & Pellegrino, J. W. (1996). Acquisition and transfer of skilled performance: Are visual discrimination skills stimulus specific? *Journal of Experimental Psychology: Human Perception and Performance, 22*(5), 1218-1248.

Doll, T. J., & Home, R. (2001). Guidelines for developing and validating models of visual search and target acquisition. *Optical Engineering, 40*(9), 1776-1783.

Eccles, D. W., Walsh, S. E., & Ingledew, D. K. (2006). Visual attention in orienteers at different levels of experience. *Journal of Sports Sciences, 24*(1), 77-87.

Eddy, D. M. (1998). Performance measurement: problems and solutions. *Health Affairs, 17*(4), 7.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191.

Fiore, S. M., Scielzo, S., & Jentsch, F. (2004). Stimulus Competition During Perceptual Learning: Training and Aptitude Considerations in the X-ray Security Screening Process. *International Journal of Cognitive Technology, 9*(2), 34-39.

Ford, J. K., Smith, E. M., Weissbein, D. A., Gully, S. M., & Salas, E. (1998). Relationships of goal orientation, metacognitive activity, and practice strategies with learning outcomes and transfer. *Journal of Applied Psychology, 83*(2), 218-233.

Fowlkes, J., Dwyer, D. J., Oser, R. L., & Salas, E. (1998). Event-based approach to training (EBAT). *International Journal of Aviation Psychology, 8*(3), 209-221.

Fowlkes, J. E., Salas, E., Baker, D. P., Cannon-Bowers, J. A., & Stout, R. J. (2000). The Utility of Event-Based Knowledge Elicitation. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 42*(1), 24-35.

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65*(7), 999-1010.

Goodman, J. S., Wood, R. E., & Hendrickx, M. (2004). Feedback Specificity, Exploration, and Learning. *Journal of Applied Psychology, 89*(2), 248-262.

Gramopadhye, A. K., Drury, C. G., & Sharit, J. (1997). Feedback strategies for visual search in airframe structural inspection. *International Journal of Industrial Ergonomics, 19*(5), 333-344.

Hagemann, N., Strauss, B., & Canal-Bruland, R. (2006). Training Perceptual Skill by Orienting Visual Attention. *Journal of Sport & Exercise Psychology, 28*(2), 143-158.

Hale, S. K., Fuchs, S., Berka, C., Cowell, A., Baskin, A., Jones, D., et al. (2007). Revolutionary Accelerated Processing Image Detection (RAPID) System. Technical Paper for CASE PI meeting in Dallas, TX, March 12-16.

Hillstrom, A. P., & Logan, G. D. (1998). Decomposing visual search: Evidence of multiple item-specific skills. *Journal of Experimental Psychology: Human Perception and Performance, 24*(5), 1385-1398.

Hoffman, J. E. (1999). Stages of processing in visual search and attention. In B. Challis & B. Velichovsky (Eds.), *Stratification in cognition and consciousness* (Vol. , pp. 43-71). Philadelphia: John Benjamins.

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience, 2*(3), 194-203.

Jarodzka, H., Scheiter, K., Gerjets, P., & van Gog, T. (2009). In the eyes of the beholder: How experts and novices interpret dynamic stimuli. *Learning and Instruction, 20*(2), 146-154.

Jiang, Y., & Chun, M. M. (2001). Selective attention modulates implicit learning. *The Quarterly Journal of Experimental Psychology Section A, 54*(4), 1105-1124.

Jodlowski, M. T., & Doane, S. M. (2004). A knowledge scoring engine (KSE) for real-time knowledge base generation used in intelligent tutoring systems. *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, 131-139.

Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology, 8*(4), 441-480.

Kasarskis, P., Stehwien, J., Hickox, J., Aretz, A., & Wickens, C. (2001). *Comparison of expert and novice scan behaviors during VFR flight.* Paper presented at the 11th International Symposium on Aviation Psychology, Columbus, OH: The Ohio State University.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. *Decision making in action: Models and methods*, 138-147.

Klein, G. A. (1998). *Sources of Power: How People Make Decisions*: MIT Press.

Leach, D. J., & Conto, H. (1999). The additional effects of process and outcome feedback following brief in-service teacher training. *Educational Psychology, 19*(4), 441-462.

Leonards, U., Rettenbach, R., Nase, G., & Sireteanu, R. (2002). Perceptual learning of highly demanding visual search tasks. *Vision Research, 42*(18), 2193-2204.

Levine, M. W. (2000). *Levine and Shefner's Fundamentals of Sensation and Perception* (Third ed.). New York: Oxford University Press.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide (2nd ed.).* Mahwah, NJ US: Lawrence Erlbaum Associates Publishers

Manning, D., & Ethell, S. (2002). *Insights into Sources of Error in Diagnosis of Lung Cancer from Chest Radiography.* Paper presented at the Medical Image Understanding and Analysis Conference, University of Portsmouth, Portsmouth, UK.

Manning, D., Ethell, S., Donovan, T., & Crawford, T. (2006). How do radiologists do it? The influence of experience and training on searching for chest nodules. *Radiography, 12*(2), 134-142.

Manning, D. J., Leach, J., & Bunting, S. (2000). A comparison of expert and novice performance in the detection of simulated pulmonary nodules. *Radiography, 6*, 111-116.

McCarley, J. S., Kramer, A. F., Wickens, C. D., Vidoni, E. D., & Boot, W. R. (2004). Visual Skills in Airport-Security Screening. *Psychological Science, 15*(5), 302-306.

MCI 03.35c Infantry Patrolling. (1996). Washington DC: Marine Corps Institute.

Mehta, P., Sadasivan, S., Greenstein, J. S., Gramopadhye, A. K., & Duchowski, A. T. (2005). Evaluating Different Display Techniques for Communicating Search Strategy Training in a Collaborative Virtual Aircraft Inspection Environment. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 49*, 2244-2248.

Mello-Thoms, C., Dunn, S., Nodine, C. F., Kundel, H. L., & Weinstein, S. P. (2002). The Perception of Breast Cancer: What Differentiates Missed from Reported Cancers in Mammography. *Academic Radiology, 9*(9), 1004-1012.

Mello-Thoms, C., Ganott, M., Sumkin, J., Hakim, C., Britton, C., Wallace, L., et al. (2008). Different Search Patterns and Similar Decision Outcomes: How Can Experts Agree in the Decisions They Make When Reading Digital Mammograms? In *Digital Mammography* (pp. 212-219).

Mello-Thoms, C., Nodine, C. F., & Kundel, H. L. (2002). What attracts the eye to the location of missed and reported breast cancers? *Proceedings of the symposium on Eye tracking research & applications*, 111-117.

Mory, E. H. (2004). Feedback Research Revisited. In D. H. Jonassen (Ed.), *Handbook of Research on Educational Communications and Technology* (2nd ed., pp. 745-783). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Nalanagula, D., Greenstein, J. S., & Gramopadhye, A. K. (2006). Evaluation of the effect of feedforward training displays of search strategy on visual search performance. *International Journal of Industrial Ergonomics, 36*(4), 289-300.

Nodine, C. F., Krupinski, E. A., & Kundel, H. L. (1990). A perceptually-based algorithm provides effective visual feedbackto radiologists searching for lung nodules. *Visualization in Biomedical Computing, 1990., Proceedings of the First Conference on*, 202-207.

Nodine, C. F., Kundel, H. L., Mello-Thoms, C., & Weinstein, S. P. (2001). Role of computer-assisted visual search in mammographic interpretation. *Proceedings of SPIE, 4324*, 52.

Nodine, C. F., Mello-Thoms, C., Kundel, H. L., & Weinstein, S. P. (2002). Time course of perception and decision making during mammographic interpretation. *American Journal of Roentgenology, 179*, 917–923.

Parasuraman, R., Boff, K. R., Kaufman, L., & Thomas, J. P. (1986). Vigilance, monitoring, and search. In *Handbook of perception and human performance, Vol. 2: Cognitive processes and performance.* (pp. 1-39). Oxford, England: John Wiley & Sons.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*(1), 107-123.

Pike, G., & Edgar, G. (2005). Perception. In N. Braisby & A. Gellatly (Eds.), *Cognitive psychology*. Milton Keynes, England: Oxford University Press.

Raab, M., & Johnson, J. G. (2007). Expertise-based differences in search and option-generation strategies. *Journal of Experimental Psychology: Applied, 13*(3), 158-170.

Rezec, A. A., & Dobkins, K. R. (2004). Attentional weighting: A possible account of visual field asymmetries in visual search? *Spatial Vision, 17*(4/5), 269-293.

Rooney, J. J., & Heuvel, L. N. V. (2004). Root cause analysis for beginners. *Quality progress, 37*(7), 45-56.

Sadasivan, S., Greenstein, J. S., Gramopadhye, A. K., & Duchowski, A. T. (2005). *Use of eye movements as feedforward training for a synthetic aircraft inspection task*. Paper presented at the Proceedings of the SIGCHI conference on Human factors in computing systems.

Salas, E., & Cannon-Bowers, J. A. (2001). The science of training: A decade of progress. *Annual Review of Psychology, 52*, 471-499.

Salas, E., Rosen, M. A., Burke, C. S., Nicholson, D., & Howse, W. R. (2007). Markers for Enhancing Team Cognition in Complex Environments: The Power of Team Performance Diagnosis. *Aviation, Space, and Environmental Medicine, 78*(1), B77-B85.

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*(4), 207-217.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review, 84*(2), 127-190.

Sidani, T. A., & Gonzalez, A. J. (1994). Learning situational knowledge through observation of expertperformance in a simulation-based environment. *Southco 94. Conference Record*, 65-70.

Sireteanu, R., & Rettenbach, R. (2000). Perceptual learning in visual search generalizes over tasks, locations, and eyes. *Vision Research, 40*(21), 2925-2949.

Smith-Jentsch, K. A., Johnston, J. H., Payne, S. C., Cannon-Bowers, J. A., & Salas, E. (1998). Measuring team-related expertise in complex environments. In *Making decisions under stress: Implications for individual and team training*. (pp. 61-87): American Psychological Association.

Smith-Jentsch, K. A., Zeisig, R. L., Acton, B., McPherson, J. A., Cannon-Bowers, J. A., & Salas, E. (1998). Team dimensional training: A strategy for guided team self-correction. In *Making decisions under stress: Implications for individual and team training*. (pp. 271-297). Washington, DC, US: American Psychological Association.

Smith, C. A. P., Johnston, J., & Paris, C. (2004). Decision Support for Air Warfare: Detection of Deceptive Threats. *Group Decision and Negotiation, 13*(2), 129-148.

Sowden, P. T., Davies, I. R. L., & Roling, P. (2000). Perceptual learning of the detection of features in X-ray images: A functional role for improvements in adults' visual sensitivity? *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 379-390.

Sowden, P. T., Rose, D., & Davies, I. R. L. (2002). Perceptual learning of luminance contrast detection: Specific for spatial frequency and retinal location but not orientation. *Vision Research, 42*(10), 1249-1258.

Togami, H. (1984). Affects on visual search performance of individual differences in fixation time and number of fixations. *Ergonomics, 27*(7), 789 - 799.

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

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114*(3), 285-310.

Treisman, A., Vieira, A., & Hayes, A. (1992). Automaticity and preattentive processing. *American Journal of Psychology, 105*(2), 341-362.

Ucko, D. (2007). Countering Insurgents through Distributed Operations: Insights from Malaya 1948–1960. *Journal of Strategic Studies, 30*(1), 47-72.

Underwood, G. (2007). Visual attention and the transition from novice to advanced driver. *Ergonomics, 50*(8), 1235 - 1249.

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. (2003). Visual attention while driving: Sequences of eye fixations made by experienced and novice drivers. *Ergonomics, 46*(6), 629-646.

Verghese, P. (2001). Visual Search and Attention A Signal Detection Theory Approach. *Neuron, 31*(4), 523-535.

Wang, H., Chignell, M., & Ishizuka, M. (2006). *Empathic Multiple Tutoring Agents for Multiple Learner Interface.* Paper presented at the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology.

Wang, M. J., Lin, S., & Drury, C. G. (1997). Training for strategy in visual search. *International Journal of Industrial Ergonomics, 20*(2), 101-108.

White, K. P., Jr., Hutson, T. L., & Hutchinson, T. E. (1997). Modeling human eye behavior during mammographic scanning: Preliminary results. *IEEE Transactions on Systems, Man, & Cybernetics Part A: Systems & Humans, 27*(4), 494-505.

Wickens, C. D., & Flach, J. M. (1988). Information processing. In E. Wiener & D. Nagel (Eds.), *Human Factors in Aviation.* San Diego, CA: Academic Press.

Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance (3rd ed.).* Upper Saddle, NJ: Prentice-Hall Inc.

Williams, A. M., Ward, P., Knowles, J. M., & Smeeton, N. J. (2002). Anticipation skill in a real-world task: Measurement, training, and transfer in tennis. *Journal of Experimental Psychology: Applied, 8*(4), 259-270.

Wolfe, J. M. (1998). Visual search. In H. Pashler (Ed.), *Attention* (pp. 13–73). London, UK: University College London Press.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology, 15*, 419–433.

Yi, D.-J., Olson, I. R., & Chun, M. M. (2006). Shape-specific perceptual learning in a figure-ground segregation task. *Vision Research, 46*(6), 914-924.