THESIS

INDIVIDUAL DIFFERENCES IN WORKING MEMORY AFFECT SITUATION AWARENESS

Submitted by

Robert S. Gutzwiller

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2011

Master's Committee:

Advisor: Benjamin A. Clegg

Edward DeLosh
Stephen Hayne

ABSTRACT

INDIVIDUAL DIFFERENCES IN WORKING MEMORY AFFECT SITUATION AWARENESS

Situation awareness (SA) is a construct that brings together theories of attention, memory, and expertise in an empirical effort to showcase what awareness is and how it is acquired by operators. Endsley (1995a) defined SA in a way that includes many theoretical associations between awareness and specific memory and attention mechanisms. Work characterizing these relationships has been sparse, however, particularly with regard to the influence of working memory (WM) on SA in novices. An experiment was devised which principally investigated novice SA as a theorized function of WM across two distinct tasks; one in which operator attention and perception (Level 1 SA) was assessed, and one in which an operator's ability to respond to events in the future (Level 3 SA) was implicitly assessed. Factors analysis was used and resulting outcomes from three WM tasks loaded well onto one overall WM factor. Findings from 99 participants indicate that WM does have a correlative and predictive relationship with Level 3, but not Level 1 SA. Results reported here contribute to ongoing theory and experimental work in applied psychology with regard to SA and individual differences,

showing WM influences awareness in novice performance even in the case where SA measures are not memory-reliant.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION

What is Situation Awareness?

Situation awareness (SA) is an empirical description of human operator awareness for the performance of complex dynamic tasks. A few key publications in this domain have developed a widely accepted conceptualization of awareness (e.g., Endsley 1995a; 1995b), with Endsley (1995a) in particular setting forth a clear definition as "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (36). Each of these main ideas is described as a "Level" of situation awareness, thus *Level 1* SA is the perception of critical elements of information in the environment, *Level 2* represents the comprehension and integration of the information elements, and *Level 3* SA represents the mental prediction of the environment or system state into the future. The accuracy of operator awareness, then, can be broken down into these same tractable portions for further understanding.

While there have been arguments about the viability of situation awareness by a small number of researchers (see Dekker & Hollnagel, 2004), many others have made it clear that SA assessment is not simply a buzzword, but is a useful evaluative measure for analyzing the behavior of humans operating in complex systems (Durso & Sethumadhavan, 2008; Parasuraman, Sheridan, & Wickens, 2008; Wickens, 2008). The trend in the field has been to examine the SA-specific requirements for certain tasks (Durso & Sethumadhavan, 2008) while remaining theoretical at times with regard to the contributions of underlying cognitive processes which contribute to SA.

Though SA has been shown to reflect more than the individual summation of processes (in other words there is a gestalt benefit of SA above any singular contributions of attention, memory, or any other individual process; Durso, Bleckley, & Dattel, 2006), this finding does not rule out the importance of individual differences. Individual differences in attention and memory have been shown to exhibit important and empirical relationships to awareness (Endsley & Bolstad, 1994; Gonzalez & Wimisberg, 2007; Gugerty & Tirre, 2000; O'brien & O'hare, 2007; Sohn & Doane, 2004) and SA has also been shown to differ between operators as a function of task expertise (Sohn & Doane, 2004) which suggests a strong memory component. It is not well known how these relationships contribute to awareness specifically within the context of the SA model posited by Endsley (1995a). Endsley (1995a)'s three level model has been applied effectively to many settings though, and is applicable for dissociating the contribution of cognitive processes, as influence can be summed at each Level, allowing for more sensitivity (Endsley & Bolstad, 1994). The remaining section will highlight the empirical work and theory surrounding each Level in Endsley's model.

Levels of awareness and associated empirically supported theory.

Endsley's model contains three levels of SA corresponding to perception, integration and comprehension, and finally prediction.

Level 1 situation awareness.

Level 1 SA embodies the concept of critical information noticing. Level 1 SA is considered perception-based awareness, and the allocating or directing of attention is the main influencer of whether information is perceived by an operator (Endsley, 1995a). These attention mechanisms influence where, when, and how well the operator can generally gather information (Endsley, 1995a). Past work, however, has shown only a small relationship between SA and attention and this effect appears to be modulated by the degree of difficulty of any attention task, with more difficult tasks showing a slightly more significant relationship (Endlsey & Bolstad, 1994).

Level 1 SA is also considered the building block of further SA understanding. By gathering the critical information in the environment together, an operator fosters understanding of elemental relationships and then interprets the information. Thus, Level 1 SA builds to allow Level 2 SA to form.

Level 2 situation awareness.

Level 2 SA is considered by Endsley (1995a) to be reliant on memory, though attention is certainly involved. It represents the integration and the interpretation of critical elements (information), which relies on both long term memory and working memory (Endsley, 1995a). Long term memory helps operators classify and understand information in the environment via schemas, while working memory is primarily used for manipulating, combining, or keeping information available to the operator. Recent conceptualizations of working memory (WM) consider it to be a limited capacity storage system used to keep currently focused information available over short term periods of time, and to allow for the immediate manipulation of that information in the mind of the operator (Baddeley, 1986). Thus SA theory has been developed to incorporate this understanding of cognitive ability, but also its limitation; in other words, at a certain level, awareness might be constrained by an operator's limited WM capacity.

Integrating knowledge relies on WM involvement – however in experts and highly experienced operators, this may be represented by a special subset, called long-term working memory (LT-WM) as proposed by Ericsson and Kintsch (1995). The theory proposes that an operator acquires extensive representations of task information in long term memory with large amounts of practice, which increases the functional capacity of WM via templates. This gives expert operators an easily accessed context in which to place information for storage and retrieval (Ericsson & Kintsch, 1995), and this ability has been shown to contribute significantly to awareness in experts distinct from other measures of WM (Sohn & Doane, 2004).

WM and the interactions with long term memory are important for integrating, interpreting, and storing information. Endsley (1995a) postulated that these cognitive processes are also taxed heavily during the final level of awareness, Level 3.

Level 3 situation awareness.

Level 3 SA represents the prediction component of awareness. Here, a conceptualization of how an operator engages in attempts to predict events in the environment is presented by Endsley. An accurate prediction of an event is only achievable in most cases by possessing a mastery of the complexities of the system being evaluated. Operators who are able to predict changes well anticipate future events and appreciate any upcoming changes in the system, increasing their awareness of the near-future (Endsley, 1995a) and their ability to take appropriate actions.

The ability to conduct effective perception of the elements of information (Level 1 SA accuracy) continues to play a role at this later Level as well. For example Bellenkes, Wickens, and Kramer (1997) showed novice pilots tend to be less proficient in anticipating future aircraft states, in part because of their visual scan patterns, which are characteristically inflexible. Expert pilots exhibited more anticipatory behavior, primarily as a function of their flexible visual scanning to more predictive elements in the cockpit. Thus their control over perception of information (Level 1) and their interpretation of that information (Level 2) aid them in prediction (Level 3) above novices who are limited in initial information acquisition.

Expertise also influences the impact that WM may exert at these levels, with experts relying more on LT-WM than novices (Sohn & Doane, 2004). Only a handful of studies have looked at the contributing cognitive factors specifically at the third Level of SA (Gonzalez & Wimisberg, 2007; O'brien & O'hare, 2007; Sohn & Doane, 2004) and past research makes the case that novice participants may differ significantly in the contributing constructs driving their awareness.

The following section covers major ways that SA has been assessed, potential pitfalls in different techniques, and the methods used in a later experiment.

Situation awareness measurement techniques

Techniques determine our ability to accurately measure awareness and discriminate among the Levels. There are two overarching types of SA measurement; subjective and objective.

Subjective SA Measures

A number of subjective rating techniques, including SART (Situation Awareness Rating Technique) and SA-SWORD (Situation Awareness-Subjective Workload Dominance) exist to measure what an operator believes about their own SA (see Jones, 2000 for an extensive review).

Subjective SA assessments possess several shortcomings; in particular, an operator's knowledge about what should be perceived or noticed may be inadequate due to inexperience (Jones, 2000) and any self-reports may be poorly diagnostic of awareness since the operator may feel that they are aware despite being unaware of critical information. Additional evidence (Venturino, Hamilton, & Dvorchak, 1989 as cited in Jones, 2000) also suggests that subjective SA ratings are highly influenced by the quality of performance. If performance during a test session is perceived as good, operators typically rate their awareness as "high", and if performance is poor they rate SA as "low" – in other words, subjective ratings can anchor on task performance instead of personal levels of awareness (Venturino et al., 1989).

Objective SA Measures

Two widely used objective measures of awareness are the Situation Awareness Global Assessment Technique, or SAGAT (Endsley, 1988; 1995a) and the Situation Present Assessment Method, or SPAM (Durso et al, 1998).

SAGAT is an interruption-based technique where operators are asked questions about critical components along all three Levels of SA mid-task, or in some cases can be administered immediately post-task. Each question is typically developed via a cognitive task analysis of the system, which identifies all of the critical task information needed for both performance and awareness. The accuracy of responses to these questions determines how aware the operator is, and responses can be categorized by each distinct Level of SA (1, 2, or 3).

One key aspect of SAGAT is that it requires a temporary pause to probe performance in order to answer administered questions. While some researchers object to the interruption of the task, research has validated the use of the technique by showing that interruptions from questioning do not interfere with SA or performance (Endsley, 2000); however it does represent a deviation from the conditions of real-world task performance.

Other methods exist for objective SA assessment including SPAM. This method queries operators during *ongoing* task performance without any pausing of the activity. Instead of relying solely on answer accuracy, reaction times to queries are used to determine awareness. Shorter reaction times therefore indicate the information is currently in the operator's awareness (Durso et al., 1998). The SPAM measure, however, can be negatively influenced by operator mental workload which slows question answering even when the answer may be known at the time. The technique also makes answering questions about spatial relationships and location difficult to assess due to the verbal nature of the inquiry; in other words, operators can not indicate verbally where some object or information is located with sensitivity compared to metrics where the indication can be made using actual spatial representations and operator responses, such as a map where an operator can physically indicate where an object is.

Another measure of SA uses indirect task performance to specifically assess awareness (Brickman et al., 1999; Pritchett & Hansman, 1997; Yanco & Drury, 2004). An important distinction of this method from SAGAT and SPAM is that it is unobtrusive. This method has been used to infer operator awareness by monitoring behavior in response to a potential hazard which required action from the operator – in this case, a runway incursion (Pritchett & Hansman, 1997). The delay between onset of the hazard and appropriate behavioral action is a measure of awareness, which differs from performance because even if the hazard is avoided (successful performance) the relative delay in action may be characteristic of lower overall awareness to the hazard.

An important advantage of this indirect task methodology stems from the ease of implementation by avoiding other confounds as mentioned for SAGAT and SPAM - there is no task interruption, which while having been shown to be innocuous may still interrupt operator cognition, and questions are not administered verbally or mid-task which leaves the operator free to allocate full mental resources to the task at hand which is contextually consistent with real-world performance.

One goal of the current study is to conduct assessment of specific cognitive mechanisms' contributions which are theorized to be underlying awareness. The avoidance of subjective measures (and thus potentially low construct validity) is paramount. An indirect SA measurement technique represents a particularly nice case for evaluating individual differences impacting SA, since these effects may be small and disrupted by injecting an increased load or interruptions via other assessment techniques, particularly in novices who are already dealing with unfamiliar situations and tasks. The following section highlights the empirically supported individual differences that contribute to general awareness ability.

Individual differences and the influence on situation awareness

Endsley (1995a; 1995b) theorized that many cognitive processes, including memory and attention, were implicated in the ability to form SA and maintain awareness in complex tasks. In the previous year, a seminal paper by Endsley and Bolstad (1994) examined a variety of these influences by measuring the spatial ability, attention, perception, memory, and analytical skills of their participants. After completing tests, expert pilots were asked to engage in several simulated air-to-air combat scenarios. SAGAT was administered at different points during the simulation, or not at all, and participant answers were compared to recordings from the simulation and scored on the accuracy of their responses. Several spatial measures correlated well with SA accuracy; however attention and perception measures correlated with SA accuracy only when the tasks were at their highest difficulty, suggesting that in order to attribute these constructs to general SA ability in experts, it is necessary to tax the participants heavily. Because there was no condition in which novice pilots were similarly tested, however, the characterization of novice awareness was not explored and therefore represents a case in which processes contributing to awareness may be differing in their contribution.

It is therefore conceivable that awareness in experts may be differentially influenced by perceptual processes and WM. In addition, it has been shown that during dynamic task performance the ability of the operator to focus attention improves information perception. WM contributes to performance on Level 1 style tasks (Bleckley, Durso, Crutchfield, Engle, & Khanna, 2003). High WM capacity participants in the study showed characteristics of flexible allocation of attention in a location identification task. This flexibility aided performance, while those participants with low working memory capacity remained fixed in their attention allocation and performed worse than high capacity participants (Bleckley et al., 2003). Other research suggests attention is controlled through WM. Lavie and De Fockart (1995) notably showed that by increasing WM load, the attention capture rate from a non-target stimulus also increased. Thus, the ability of an operator to attend selectively to information in an environment while ignoring competing information depends on WM and this may show up in WM-SA relationships as well.

Timing involved in the direction of attention has been tagged as potentially driving individual differences in SA (Endsley, 1995a) and interestingly has also been related to WM (Baddeley, Chincotta, & Adlam, 2001). Baddeley et al. (2001) found that the completion time of a list of simple math problems that required task switching was significantly increased when participants were concurrently engaged in a verbal trail task (theorized to tax executive processes involved in attention allocation and the control of WM). And in a more SA-relevant example, Boldstad, Endsley, Howell, and Costello (2003) showed that training on a task switching task does not improve real-world task switching performance in pilots, suggesting this is a stable ability (much as WM is considered to be stable within participants).

The ability to task switch seems to rely significantly on attentional control, which clearly implicates WM. Therefore, although attention is involved in awareness at a basic level (attending to information in the environment), attention may be directed by WM (choosing where to attend, and when) (Bleckley et al., 2003; Bolstad et al., 2003; Lavie & De Fockart, 1995). Examining the relationship between WM and SA more fully should help answer the question of exactly how WM may impact distinct Levels of awareness, and where the theory needs to be updated or revised

Working Memory

It seems clear that WM plays a role in awareness in dynamic tasks. The driver of a vehicle must maintain a myriad of critical information including knowledge of travel directions, traffic conditions, awareness of the location of the vehicle in space, and knowledge about the state of the vehicle itself, all of which is held within memory (Gugerty & Tirre, 2000). Individual variance in WM translates to individual differences in the amount of critical information and number of relationships between elements an operator is capable of maintaining at any given time – not to mention their ability to anticipate collisions (aiding avoidance). WM is theoretically important for accurate formation of Level 2 SA and Level 3 SA (Endsley, 1995a) and research supports this (Gugerty & Tirre, 2000; O'brien & O'hare, 2007; Sohn & Doane, 2004). However, others have shown WM to be related to principles of Level 1 SA formation as well (Bleckley et al., 2003; Bolstad et al., 2003; Lavie & De Fockart, 1995). The following sections highlight research which considers individual differences in WM and their impact on SA.

Research highlighting the relationship between WM and SA.

Various measures of WM have been used to examine the influence WM may have on performance and SA (Carretta, Perry, & Ree, 1996; Gonzalez & Wimisberg, 2007; Gugerty & Tirre, 2000; O'brien & O'hare, 2007; Sohn & Doane, 2004); however these measures differ significantly from each other in many ways. While some researchers (Gonzalez & Wimisberg, 2007; Sohn & Doane, 2004) choose to use traditional measures of WM from previous studies such as the *letter rotation* task from Shah and Miyake (1996), others have used complex batteries of tests that include measures of WM (usually span-based tasks) mixed with other metrics and use a combined score for participants. For example, Gugerty and Tirre (2000) used the CAM battery (Cognitive Abilities Measurement), ASVAB (Armed Services Vocational Aptitude Battery), and the AFOQT (Air Force Officer Qualifying Test) to assess participants and compiled a WM measure; and O'brien and O'hare (2007) used a battery test called WOMBAT which also contains a WM measure. These batteries of tests make results harder to interpret for pure WM influence due to potential interference from other tasks that are included in the test.

Engle and colleagues (Engle, Tuholski, Laughlin, & Conway (1999), Kane et al., 2004; Unsworth, Schrock, & Engle, 2004) have supported the idea that WM is one single construct. Other recent work has shown span-based WM tasks (specifically reading span, operation span, and counting span) load well onto a single construct of WM, separate from other tasks that load more appropriately onto a short term memory construct (Engle et al., 1999). Additional work has also suggested factor analytical methodologies are generally successful in showing relationships between WM and SA (Gugerty & Tirre, 2000). Therefore including one single test of WM, such as reading span, is less than ideal, given that it alone will not represent the construct of WM as well as a factor analytical model with additional task loadings. The most ideal approximation of WM could therefore be gained by eliciting several WM measures from participants and loading them onto a WM factor which is then used for later statistical analyses.

The hypothesized relationship between WM and SA has been previously assessed in various ways, from correlational analyses (Gonzalez & Wimisberg, 2007) to predictions of awareness or performance gained by examining measures of WM across various tasks and environments (Carretta et al., 1996; Gugerty & Tirre, 2000; O'brien & O'hare, 2007; Sohn & Doane, 2004). Carretta et al. (1996) examined military fighter pilots using a subjective awareness rating survey. Pilots were evaluated by peers and superiors on 31 behavioral survey items representative of individual traits and job-related tasks (such as tactics and communication). After controlling for flight experience, six different significant predictors of subjective SA emerged which included a measure of verbal WM, a measure of spatial WM, and a measure of spatial reasoning. Although WM was a significant predictor, the subjective rating of SA may be reflective of the quality of information integration or prediction that the operators are *perceived* to be capable of doing. It is unclear how accurate the subjective judgments were and if they fall victim to the pitfalls of subjective SA measures. More objective work is needed to decipher how individual differences in WM contribute to awareness.

A step in this direction was taken by O'brien and O'hare (2007) who looked at how individual differences affect SA accuracy when performing an air traffic control task. Participants were measured using the WOMBAT battery, a domain independent measure of an operator's ability to handle various demands of complex task performance under changing priorities (O'brien & O'hare, 2007). WOMBAT consists of four subcomponents which include a WM task and several tracking and spatial processing tasks. Using the resulting scores from WOMBAT, participants were categorized into low and high SA ability groups, and high SA ability participants were shown to perform better than low SA ability participants (correctly planning more flight patterns). In a second experiment, SAGAT technique queries were administered during pauses of the ATC simulation. The accuracy of answers to Level 2 SA and Level 3 SA queries were also found to be significantly positively correlated with scores on the WOMBAT battery. Thus over two experiments O'brien and O'hare (2007) showed that differences in WM-related measures were able to explain differences in operator awareness at later Levels (2 and 3). One issue with using O'brien and O'hare's (2007) results to support the WM-SA relationship is their use of the WOMBAT measure since it is potentially confounding spatial processing and tracking with more standard measures of WM.

To truly examine WM's pure contribution to SA, one might parse SA into levels that are influenced by separate processes, and assess the relationship at each level thus lending discriminate validity to the examination. Gugerty and Tirre (2000) used a similar methodology to show individual differences in WM are predictive of separate awareness levels within a driving task. In their study, participants completed multiple ability assessment batteries (CAM and ASVAB) and two driving trials in a simulator. During the driving simulation, participants responded to hazards (such as a merging vehicle that would result in a collision if no action was taken to avoid it) and were assessed on their ability to remember the locations of cars which occupied their blind spots. The potential *Level* of awareness being measured can be inferred, such that hazard detection reflects a Level 1 or Level 2 SA process because it is reliant on perception, and also on integrating (because identifying a hazard requires merging information together, not simply the perception of the vehicle). Blind spot car detection would instead be reliant on the prediction of the location of the car, which represents a dynamic process that takes place as a part of Level 3 SA.

A factor analysis resulted in a measure from each ability assessment which represented a combined WM and fluid *g* score. Each metric was shown to correlate significantly with performance on the hazard detection task (more WM/g ability related to greater hazard detection), and the blind spot memory task (more WM/g related to less error). Perhaps the correlation between the WM/*g* factor and blind spot accuracy is not surprising, given the role WM likely plays in any memory-based task. What is new here is the association between WM/*g* and the implicit measure of Level 1/2 SA measured by the hazard detection task.

During a second experiment, participants completed different cognitive ability measures (the AFOQT and a subset of CAM 4.1) which combined three measures of WM, including a complex spatial task, a number reasoning task, and a complex verbal task. These measures of WM were also found to be significant predictors of hazard detection responses, and explained 11% of the variance in detection of a vehicle in the driver's blind spot. These results are much easier to interpret because the WM measures used are not inclusive of additional task performance such as that used by WOMBAT, though they are tied with fluid *g*. This suggests that some relationship between WM and Level 1 SA is present, but could be tied to memory-requiring questions.

Gonzalez and Wimisberg (2007) similarly assessed the relationship of a pure WM measure and SA; operators in their experiment controlled a water purification plant simulation and SA was assessed using the SAGAT technique under two different conditions. In the first condition, the simulation display was covered during queries and in the second, the simulation display was left uncovered. The *letter rotation task* from Shah and Miyake (1996) was used as a measure of operator WM, which was shown to be a significant predictor of SA accuracy in the covered condition but not in the uncovered condition, suggesting a reliance on WM when the display was covered during questioning.

As in Gugerty and Tirre (2000) it should not be surprising that a WM measure is related to performance on essentially a memory test, and the differential strength of the relationship between WM and SA during the covered and uncovered conditions alludes to this. However, Gonzalez and Wimisberg (2007) showed the methodological efficacy of keeping WM measures separate from extraneous tasks, as only the WM task assessed is involved in the relationship although only one span task was used. As discussed previously, other WM measures should be included in a factor analytical method, and this may come to be important in assessing and determining the most influential approximation of WM and how the WM construct is tapped in dynamic environments.

In general, WM has been shown to be predictive of awareness in some way at every Level; however these relationships are dependent on whether the awareness metric is memory invoking, and may change over the time spent performing a task. The following section covers this possibility from the perspective of skill learning and finds limited evidence that relationships between WM and SA behave in a similar way.

Theory on the Time Course of the Ability, WM, and SA Relationship

Research into the contribution of WM and cognitive ability as related to task performance and skill acquisition suggests it is important to consider the time course of any effects observed between WM and SA. For instance, training on a task has been shown to influence the correlation or contribution of WM to performance in several studies (Ackerman, 1988; 1989; Rabbitt, Banerji, & Szymanski, 1989). Two studies by Ackerman (1988; 1989) suggest individual ability has the largest correlation to task performance early on in training, but that this relationship decays over time spent on the task.

In contrast, results from a study by Rabbitt, Banerji and Szymanski (1989) suggest that correlation between task performance and cognitive ability actually *increases* with practice. Participants in their study played a dynamic game, Space Fortress, and the correlation between a general intelligence metric (using two intelligence tests, the AH-4 and AH-5) and performance within the game was measured. The correlation was initially low after the first hour of game play, but significantly increased by the fifth hour. Rabbitt et al. explained these results by suggesting increased intelligence allows participants to both attend more to the important elements of the task (echoing a similar argument for the contribution of WM to effective attention allocation and accurate Level 1 SA) and to learn rules for the game more quickly, leading to greater performance (Rabbitt et al., 1989).

While the ability-performance relationship has been examined extensively and characterized over time, the ability-SA relationship is largely unknown. Gonzalez and Wimisberg (2007) examined this relationship between WM and SA accuracy and found that the correlation between WM and SA decreased with practice on their task. These results are in support of some of Ackerman's learning theories which predict that contribution of ability to performance should decrease over time (Ackerman, 1989).

The decay of the WM-SA relationship has been investigated indirectly as well by using samples of experts who have spent considerable time in performance of the task (Sohn & Doane, 2004) and comparing expert awareness to novice awareness. Theoretically, if the relationship between WM and SA decays over the time spent on the task, there should be no predictability of expert awareness through WM, while novice awareness should be predicted well by WM. Participants in Sohn and Doane (2004) saw two consecutive screens depicting changes in aircraft instrument states, in conjunction with a goal for the future state of the aircraft, and were asked if the goal state would be reached within a certain time period. Accuracy in awareness was measured as the correct identification of the goal state as being either reachable or unreachable. Significant positive correlations were found between WM spatial span and accuracy in the SA task, but importantly this varied by expertise; novice SA relied on WM, but expert SA was best predicted by a measure of LT-WM which took previous underlying knowledge of aircraft and instrumentation experience into account (Sohn & Doane, 2004).

These data show a clear differentiation between novices and experts in the influence of WM regarding an accurate SA prediction, and suggest that WM plays a large role in novice awareness with this role falling away at the expert level (Sohn & Doane, 2004). These results also confirm that utilizing novices instead of experts may yield different levels of WM influence, and that relatively memory-unrelated SA assessments (such as these prediction tasks) are still able to be significantly related to traditional WM measures.

Intermixed in task performance (shown to be related to individual differences), is the relationship between performance and SA. Endsley (1993) characterized good SA as related but not determinant of good performance and vice versa; in other words, an operator can be aware of all critical information but apply the wrong concept to the data, or be completely unaware but "luck" into the right decisions and perform well.

Because both WM and SA are tied to performance it will be critical to show that both are contributing in unique ways.

Scaled Worlds and Awareness

To assess SA in any task is challenging, and to have sensitive experimental control within the task is perhaps even more so. Researchers have used simulations or in some cases programmable microcosms of real-world tasks such as driving and flight simulations, air traffic control tasks, and firefighting tasks to investigate the ability to be aware in our environment and chart human decision making performance. Past work shows these are in general fruitful experimental means that allow a great deal of manipulation while staying in close approximation of real world activities and conditions which provides needed ecological validity. One task which has been shown useful is Networked Fire Chief (NFC; Omodei & Wearing, 1998), a firefighting simulation. In a seminal paper, Omodei and Wearing (1995) showed that NFC can be used for assessing a wide variety of questions, many about decision making.

The task in NFC is complex and in many ways speaks not only to decision making, but also to prototypical elements of SA; operators are given control of multiple fire engines to put out fires that appear on a grid-based area of land by moving a fire engine to the location of the fire and then activating the engine to spray water. Operators also monitor the amount of water left in the fire engine, and move engines to lake areas on the map whenever they are in need of a refill. Wind is a factor within the simulation and determines the direction a fire spreads, indicated by a compass direction visible at all times on the sidebar of the simulation. The wind element makes the firefighting task additionally dynamic – for instance, if the wind blows from the south the fire will spread to the north, but the wind may also shift causing the fire's progressing direction to change.

At any given point during the active simulation, the operator can only see a portion of the overall land area that exists. It is only possible to view the other areas by moving a viewing box, located on the sidebar of the simulation to view the other areas. This can be accomplished when the operator chooses to do so, at the assumed cost of losing view of the other areas of the map.

NFC scores participants (performance) on the area of the map that is burned out of a possible perfect score of 100 (no land burned), all the way down to 0 (all land burned) and different land types, such as houses or forest areas, can be weighted as more or less important to protect from fire. In addition to this overall performance measure, the amount of water used during trials can also be tracked as a more process-based measure of performance.

Overall the NFC simulation is an opportune task for assessing awareness due to the constant need to track and refresh knowledge of the location of fire engines, and fires; the necessity of taking the prevailing wind direction, a fire engine's water amount, and the locations of water refill areas into account, keeping the overall goal of limiting fire spread in mind, and executing all tasks in a timely and as efficiently a manner as possible.

Summary and methods

WM has been shown to exhibit strong relationships to awareness, and has also accounted in some ways for other individual differences clearly related to SA such as timing and attention allocation.

Evidence suggests individual differences in WM may impact all Levels of SA (Gugerty & Tirre, 2000) and WM may represent a general underlying process. This process may contribute more to novice awareness than to expert SA (Sohn & Doane, 2004) and depend differentially on time spent performing the task (Gonzalez & Wimisberg, 2007) and the nature of the WM tasks used for predicting and correlating the relationship. The potential impact of operator WM ability on SA in general has only been investigated in part (Durso, Bleckley, & Dattel, 2006; Gonzalez & Wimisberg, 2007; Gugerty & Tirre, 2000; O'brien & O'hare, 2007; Sohn & Doane 2004). What has not been adequately addressed is the explicit contribution of WM to novice SA across Levels. Instead, these ideas have been spread out over several different studies, or examined exclusively in experts. Additionally, the relationship between WM and SA has relied primarily on investigations that involved clear memory-based tasks, rather than more implicit SA measures, and the effects may differentiate between these two metrics.

The relatively unconfirmed nature of WM contributions to awareness accuracy at each of the Levels of SA conceptualized by Endsley (1995a) is also troubling, as the trend of using SA and Endsley's (1995a) definition in real-world assessment continues. It remains imperative to understand the cognitive processes influencing operator's ability to perceive, understand, and predict environments, and this pressure is echoed in the need to complement current theory with empirical evidence.

CHAPTER 2: EXPERIMENT

The following experiment was conducted as a way of assessing the relationship between WM and two distinct Levels of SA (Level 1 and Level 3) as defined by Endsley (1995a). These relationships can be examined through assessment of the correlations between measures of WM and measures of implicit operator SA taken within performance of a simulation. By implicit scenarios with different SA requirements in a simulation, I addressed specific Levels of awareness in novice participants. Using implicit measures of SA allowed investigation of awareness while limiting interference with operator cognition, which is particularly important to novice performance.

Several hypotheses were tested using these methods. WM was tested as a significant predictor of SA in trials that focused on Level 1 SA, as well as in more complex trials that were designed to assess Level 3 SA. As previous work has shown, the WM-SA relationship may change across time and this was also an issue of investigation. Finally, the relationship between WM and SA with performance was examined, as SA may contribute to performance in the tasks in unique ways that WM may or may not account for.

Method

Participants

118 students participated for optional, partial course credit. Data from 19 participants were lost from equipment or software errors, leaving 99 participants reported in these analyses.

Materials and Procedure

All WM tasks were run using the E-Studio software package on a Dell 4600 with a standard LCD display 12.5" and 9.5" in dimension.

Working memory tasks.

In the *arrow span task* (a modified version of the arrow span task from Shah & Miyake, 1996), participants were instructed to keep track of the orientation of arrows presented one at a time and to later recall the order and the orientation of the arrows they were shown during each trial. Participants recorded responses using a circular response grid with 8 locations corresponding to each of the 8 possible presentation directions. The size of the set of arrows shown to participants varied randomly between three and seven presentations. No directions were repeated within a set, and each set size size (three, four, five, six, and seven presentations) was selected randomly for a total of three times overall for each size.

In the automated *symmetry span task*, (Kane et al., 2004; Unsworth, Heitz, Schrock, & Engle, 2005) participants were shown images and asked to judge whether the image was symmetrical along the vertical axis. Following each image judgment, a number was presented. After a series of presentations, participants were asked to recall the numbers shown in order for that series of trials. The sizes of the set of images shown varied randomly between two, three, four, five, and six images during the trials and in total three trials of each set length were presented.

In the automated *reading task* (Daneman & Carpenter, 1980; Unsworth et al., 2005) participants were asked to read sentences presented one at a time. After the sentences were read, participants judged whether the sentence made semantic sense, after which a number was presented for a brief period of time, and then the next sentence appeared. After each sentence set was completed, participants were asked to recall the numbers presented after each sentence in serial order. The sizes of the set of sentences varied randomly between two, three, four, five, and six sentences, and in total three trials of each set length were used.

Networked fire chief (NFC) simulation.

The simulation used in this experiment is a dynamic fire-fighting simulation program Networked Fire Chief (NFC; Omodei & Wearing, 1998). A typical operator-perspective view of the simulation can be seen in Appendix A.

NFC scenarios.

One basic *Practice* scenario was used for NFC training trials. This was designed to give participants experience in every aspect of the simulation, from controlling fire engine movement, to putting out fires, filling engines with water, monitoring wind direction, and assessing fire locations. Two additional types of scenarios were developed in NFC to address the theorized components of Level 1 SA and Level 3 SA in novice operators. For each scenario type, aspects of the operator's performance were hypothesized to implicitly reflect awareness.

Detection scenario trials were developed to assess Level 1 SA. During *Detection* scenarios, participants were asked to locate fires currently burning, as well as fire trucks and current wind direction on a display of the land over a brief time period, and then post-trial recall the location of fires, fire engines, and the wind direction without consulting the simulation. These trials ranged in difficulty from the easiest trials (*4x4 trials*), where participants had to locate 4 fire engines and 4 fires, to the most difficulty conditions, in which participants had to locate 8 fire engines and 8 fires (*8x8 trials*). A mixed condition was also generated for each item, such that participants also completed 4 trials of *4x8* (4 fire engines, 8 fires) and 4 trials of *8x4* (8 fire engines, 4 fires). A total of 16 *Detection* trials were developed, 4 of each difficulty type. This was expected to give a sense of whether people may differ in their ability to perceive important elements of the situation, and the task implicates processes shown to underlie Level 1 SA, in that fires, engines, and wind direction are critical to the task of putting out fires and require operator attention and perception in order to be noticed. Importantly, this is a more memory-based test of awareness.

Prediction scenario trials were developed to assess Level 3 SA. During *Prediction* scenarios, participants were given several fire engines to control and told to prevent critical areas (houses) from being burned from fires started at varying locations. These housing locations were different during each trial, providing an element of randomness that precluded the participants from being cued via context as to the location of initial fires. The critical areas themselves differed in spatial location between trials, and in addition several areas that looked perceptually similar were included in each simulation, forcing participants to carefully distinguish critical from non-critical areas. Each scenario also contained a built-in event set to occur at a time within 5 seconds of the median of the simulation trial duration. In the event, wind direction was programmed to shift directions to fuel a new fire which started in a location bordering a critical area on the map. If left unchecked the fire progressed through critical house areas, therefore the new goal that arises in the simulation should be to predict this change and allocate resources appropriately (view and react to the fire).

Awareness was implicitly assessed in each trial by tracking how long it took participants to view fire engines at the beginning of the trial (V*T-F*), the time it took participants to view a new fire event (*VT-E*) and how long it took to respond to this new event (*RT-E)* by moving fire engines. For *RT-E,* a response was defined as the time at which moving a fire engine to the immediate vicinity of the new event area occurred. A flow diagram can be found in the results section and should aid understanding the time course of these measurements (see Figure 4).

In conjunction, additional implicit measures were recorded during *Prediction* trials including *idle time* (in seconds) of the fire trucks in the simulation. This was defined as any time in which the fire truck was not engaged in a task such as moving, refilling, or fighting a fire. This gives an indication about the operator's ability to multitask with fire engines and be generally aware of what each truck's state is to allow efficient planning.

It was theorized that the participant who demonstrated good SA in these *Prediction* scenarios would control the fire which starts closest to the critical land areas first before progressing to stop other fires, and would remain aware and recognize when a new dangerous fire begins and act accordingly to stop the spread of the fire. In addition, an increased ability to utilize resources (minimizing *idle time*) represents an awareness constraint such that operators who remain aware of their truck locations should be able to send them to more fires (and would have less idle time and higher performance in comparison to lower awareness participants).

Procedure

All participants completed three separate WM assessment tasks (*symmetry task*, *arrow task*, and *reading task*) which were counterbalanced between participants.

Following completion of the WM tasks, a summary packet of information explaining the NFC simulation was given to participants (see Appendix B for a copy of the instructions). Participants then completed 2 trials of the *Practice* NFC task each lasting 4 minutes, the goal of the training trials being to familiarize participants with the basic skills necessary to operate in the simulation and put out fires using fire engines provided on the map to the best of their ability. Questions about operating NFC were not answered after the training phase.

After completing 2 *Practice* NFC trials, participants completed the 16 randomized *Detection* trials. During *Detection* trials, participants were instructed to find fire engines and in-progress fires on the displayed NFC map as well as attend to the wind direction. After a short presentation of the NFC map (5 seconds), the screen was blanked and participants recalled the locations of fires and fire engines as well as indicated the wind direction on a paper-based grid space provided.

Detection scenario trials were followed immediately by 6 unique *Prediction* trials. In *Prediction* trials, participants were told by the experimenter and reminded in the packet to attend to two key aspects of the simulation; 1) protect houses in the simulation at all costs, and 2) remain aware of any new fires that may or may not occur during the scenarios. During *Prediction* trials, participants attempted to predict fires and fire spread to save important areas from burning. Each trial lasted 5 minutes.

Results

Networked Fire Chief Performance

Performance in the six *Prediction* trials was investigated as a measure of showing task learning. As previously discussed, though good and poor SA is not a *guarantee* of good or poor performance, it is related (Endsley, 1993) and because of the impact of WM to performance it remains important to show that both WM and SA are separate contributors.

A repeated measures ANOVA was calculated to test for a main effect of time on performance for the *Prediction* trials. Because Maunchly's test indicated the data violated sphericity ($X^2(14) = 76.12$, $p<.01$), the Greenhouse-Geisser correction ($\epsilon = .775$) was used to correct degrees of freedom. There was a significant effect of time on performance in the simulation ($F(3.87, 379.53)= 355.64$, $p<.01$) which suggested that performance differed significantly between *Prediction* trials as a function of time spent performing the task (see Figure 1).

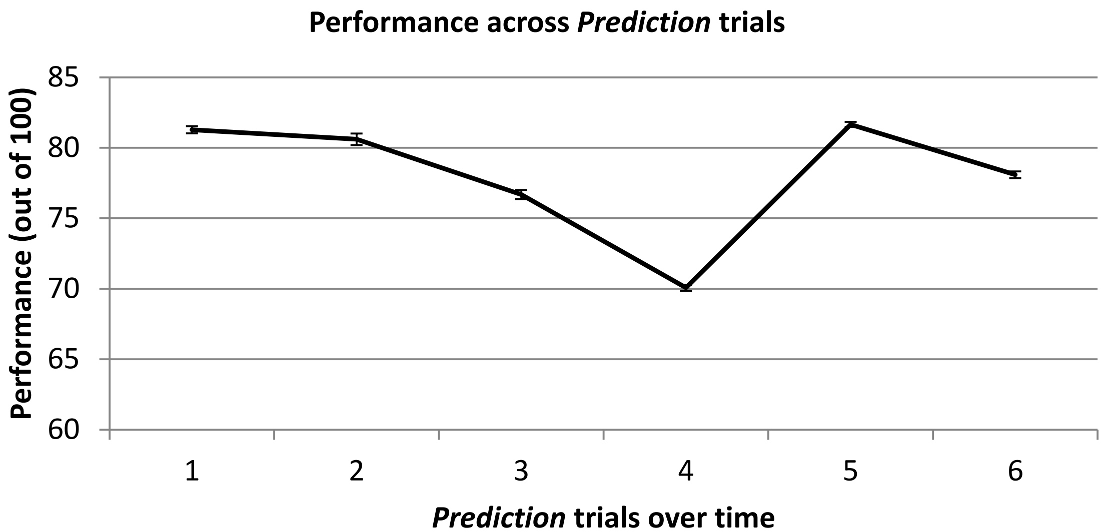


Figure 1. Performance across *Prediction* trials.

An additional, more process-based measure (*water use*) was assessed within the context of the *Prediction* trials. *Water use* was recorded as the amount of water used during the trials, and thus serves as a within-task measure of how often participants attempted to put water on fires which is thought to reflect an aspect of prospective memory for the task, since in addition to moving an engine, participants have to remember to click to put water on a fire.

A repeated measures ANOVA was calculated to test for a main effect of time on *water used* for the *Prediction* trials. Mauchly's test indicated the data violated sphericity ($X^2(14) = 67.91$, $p<.01$) so the Greenhouse-Geisser correction ($\epsilon = .727$) was used to correct degrees of freedom and a significant effect of time on *water used* in the simulation was found ($F(3.64, 356.32)= 14.22$, $p<.01$). This suggested *water used* differed significantly between *Prediction* trials as a function of the amount of trials completed (see Figure 2).

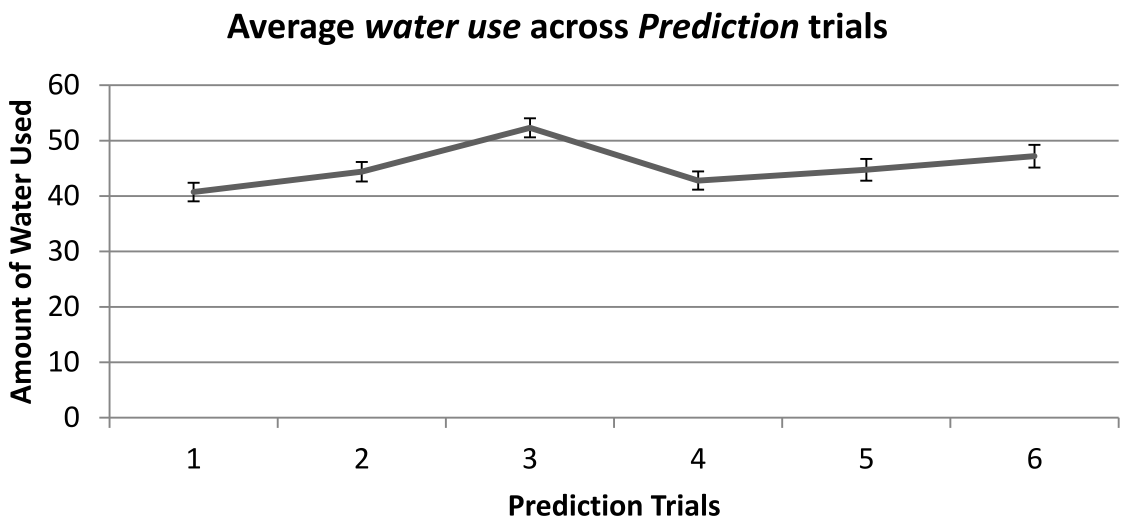


Figure 2. Average water used across each *Prediction* trial.

SA measure summary

In addition to characterizing performance over time, it is imperative to show similar data for each implicit SA measure for both *Detection* and *Prediction* trials. *Detection* trials will be examined first (see Figure 3).

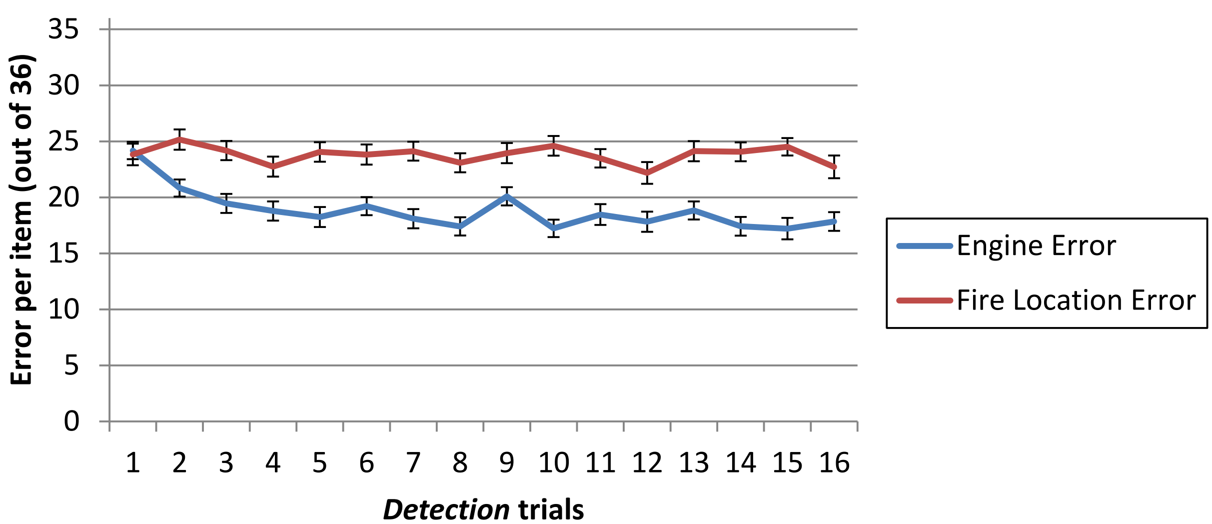


Figure 3. Per-item error across *Detection* trials for fire engine and fire location.

Error for individual fire engines and fire locations was calculated as the geometric distance between the actual location in the scene presented, and the participant-indicated location. If no fire engine or fire location was indicated as being present by participants, maximum error (36) was assigned for that item.

It is clear that while participants were not performing optimally, there is a reasonable range of performance in *Detection* trial SA measures for WM measures to account for (see Figures 3). For instance, error in truck location and fire location has a potential minimum of 0 but a maximum of 36 per item located, and participant means per item in each trial are much closer to maximal (however, not at ceiling).

A repeated measures ANOVA was conducted for fire location error across time in the *Detection* trials, but was not significant indicating that performance was relatively uniform across this measure ($F(3, 294) = 0.20$, $p>.05$). An additional repeated measures ANOVA was conducted on fire engine location error across trials in the *Detection* trials as well as wind direction accuracy. A significant effect was found for engine location ($F(3,294) = 11.53$, $p<.01$), suggesting that error varied in this task across time. A significant effect was not found for wind direction accuracy ($F(3, 288)=2.482$, $p>.05$).

Prediction trial SA measures were measured implicitly over time (see Figure 5 and 6) and are examined next.

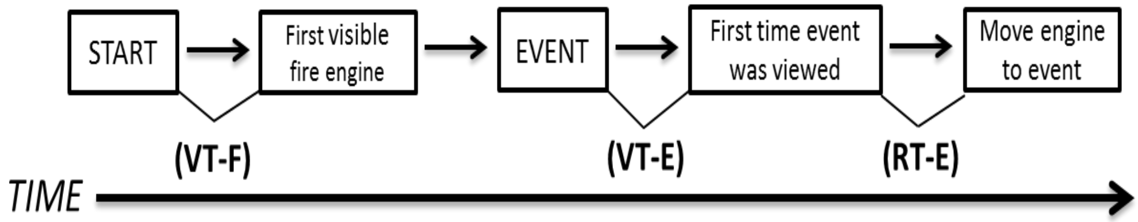


Figure 4. Linear depiction of the measurement of *VT-F*, *VT-E*, and *RT-E* in relation to critical events within *Prediction* trials.

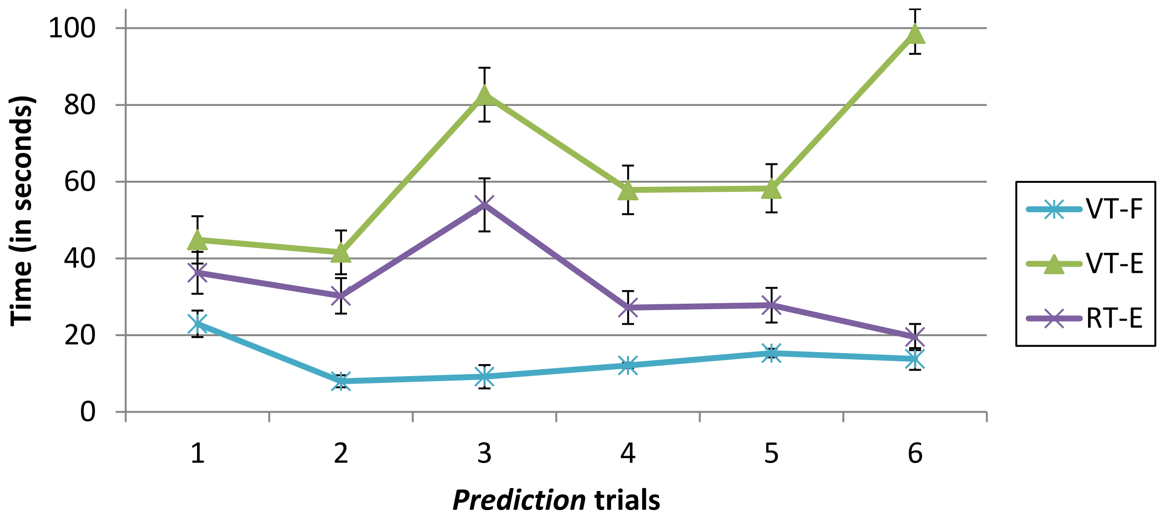


Figure 5. Average *VT-F*, *VT-E*, and *RT-E* across *Prediction* trials.

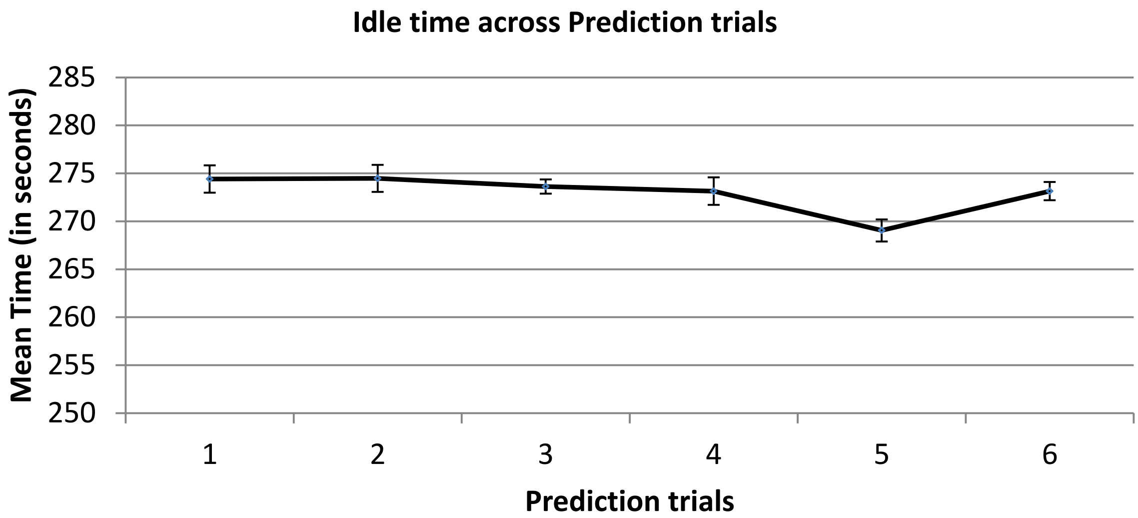


Figure 6. Average idle time of fire engines across *Prediction* trials.

A repeated measures ANOVA was conducted for *VT-F* across time in the *Prediction* trials and after correcting for a violation of sphericity (X^2 (14) = 200.01, *p*<.05) using the Greenhouse-Geisser correction (ϵ =.560), *VT-F* was found to significantly vary across time (*F*(2.802, 274.589)=13.06, *p*<.01). A similar procedure was conducted for V*T-E*, *RT-E*, and *idle time* across time in the *Prediction* trials; *VT-E* was found to significantly vary across time (*F*(4.58, 448.786)=14.62, *p*<.01), as was *RT-E* (*F*(4.19, 410.54)=5.64, *p*<.05) and *idle time* (*F*(2.58, 252.85) = 4.63, *p*<.05). Thus SA measures significantly change across trials and relationships, for both early task performance and late task performance, and in both *Detection* and *Prediction* trials.

Factor Scoring of WM Measures

A factor analytic approach is founded on the assumption that these task measures are related to one another but not perfectly, as they are expected to be accounting for unique variance in WM. Pearson correlation coefficients were calculated between each task (see Table 1) and the pattern supported this approach as the measures show some but not complete correlation.

Table 1.

	Task	Correlation		
		1	2	3
1	Arrow Span	1	-	-
2	Symmetry Span	.243*	1	-
3	Reading Span	0.114	.342**	1

n=99; ** = p<.05; ** = p<.01*

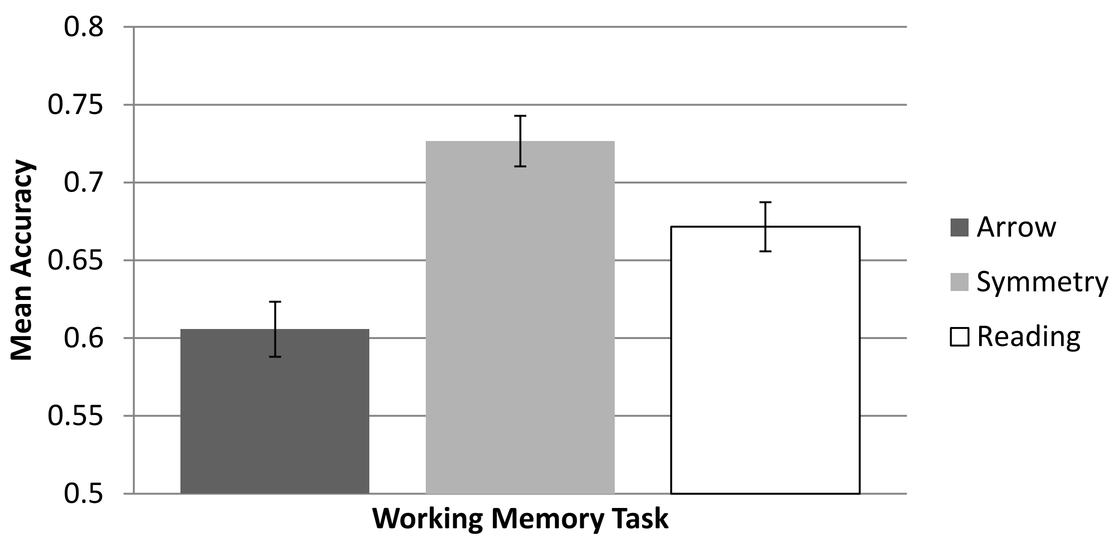


Figure 7. Working memory task performance across *Arrow span*, *Symmetry span*, and *Reading span*.

Factor scores for each participant were calculated using unrotated mean accuracy rates from each of the three WM tasks (results of each task shown in Figure 7). All three WM tasks loaded well onto a single factor model (see Table 2).

Table 2.

Component Matrix	1
Arrow span	0.529
Symmetry span	0.795
Reading span	0.731

Unrotated loadings onto a single factor

WM Relationship with SA

Of primary interest in this study was whether WM is related to measures of SA. This question was addressed under two conditions; one in which Level 1 SA was primarily assessed (*Detection* trials), and one in which Level 3 SA was primarily assessed (*Prediction* trials). The results are broken into these two levels for clarification.

***Detection* Trials (Level 1 SA) and WM**

For *Detection* trials which were designed to assess Level 1 SA, error was calculated for each measure as previously stated. Per-engine error overall (M=18.82, SD=5.03) was calculated as the total error for engine locations divided by the number of engines present in each trial. Per-fire location error overall (M=23.79, SD=5.50) was calculated similarly. Finally wind direction accuracy (M=64.5, SD=28.66) was calculated using a binary scale (1 being correct, 0 being incorrect) over all *Detection* trials. Pearson correlation coefficients were calculated for comparisons between WM and error in participant responses in the *Detection* trials. It was predicted that WM would correlate with errors in Level 1 SA. WM was not significantly correlated with per-fire location error (r= 0.16, p>.05), per-engine location error (r= -0.07, p>.05), or wind direction accuracy (r= 0.08, p>.05). However, because SA is rarely thought of as a singular item but rather a broad concept, a composite measure was calculated. Because no one SA variable was more important than another, the simplest solution was to use standardized scores for each variable and sum them. Even in this case, WM did not correlate significantly with the combined score (M=0.00, SD=1.92; r=-.005, p>.05). The null results here, while contrary to predictions, do show that memory-based measures of SA do not necessarily have to substantially reflect WM.

Although WM was not correlated with mean SA errors in *Detection* trials, perhaps the relationship is subject to change across time as previously discussed. Given the results of earlier repeated measures analysis showing error changes significantly over time, I split results into early (trials 1-8) and late (trials 9-16) performance. Pearson correlations were calculated for examining each SA variable's relation to WM.

Within early trials, WM was not significantly correlated with fire location error (r=.141, p>.05), fire engine location error (r=-.03, p>.05), or wind direction accuracy (r=-.077, p>.05). The relationship did not change when looking at results from the late trials as well, and no significant correlations were found between fire location (r=.156, p>.05), engine location (r=-.09, p>.05), or wind (r= -.138, p>.05). Therefore the WM-SA relationship was not shown to be significant in the *Detection* trials.

WM and SA as a function of difficulty.

One additional question may be whether the WM relationship with SA differed according to task difficulty in *Detection* trials. Previous work by Endsley and Bolstad (1994) suggested especially for attention or perceptual tasks the relationships with SA are tenuous, and within their experiment perceptual tasks only related to SA for the most difficult trials. To look at this in *Detection* trials I examined error for each measure averaged for the difficulty of the trial.

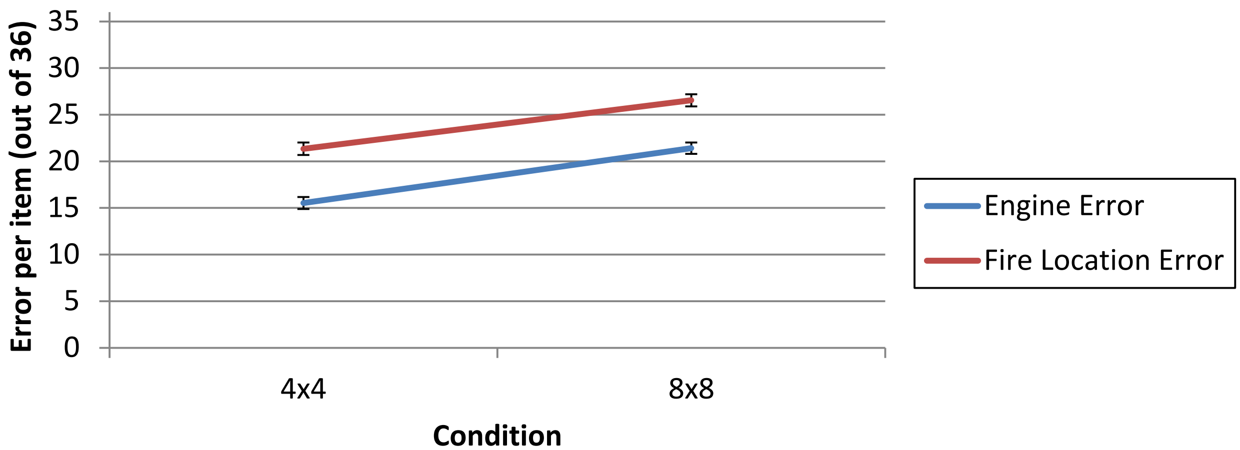


Figure 8. Fire engine and fire location error for easy (4x4) and difficult (8x8) conditions during *Detection* trials.

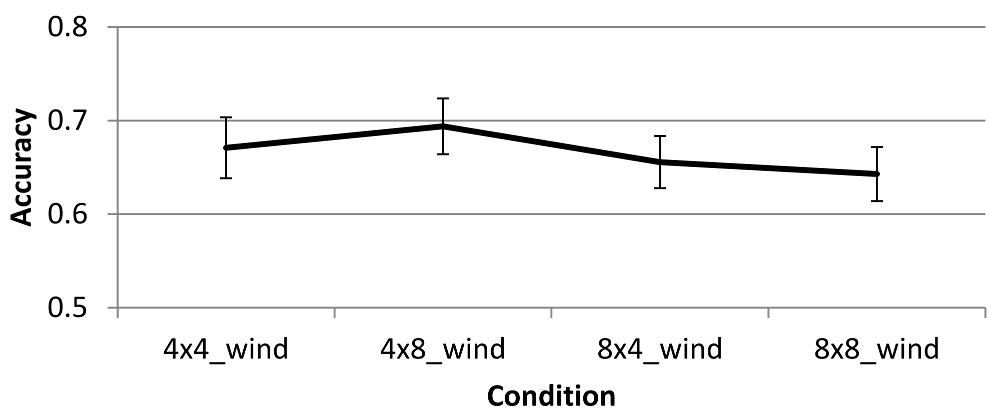


Figure 9. Wind accuracy across task difficulty for *Detection* trials.

Easier 4x4 *Detection* trials resulted in less total engine locating error than 8x8 trials ($t(98)= -28.45$, $p<.05$) and fire location error showed the same pattern, with significantly less error in the 4x4 trials versus the 8x8 trials ($t(98)=-31.40$, $p<.05$). Wind direction did not significantly change between easier 4x4 trials and more difficult 8x8 trials ($t(98)= .759$, $p>.05$).

Given that task difficulty affected error in *Detection* trials, additional Pearson correlations were calculated using error based on task difficulty and examining how this related to WM. No significant correlations were found between WM and the easier 4x4 *Detection* trials for fire location error ($r= .09$, $p>.05$) nor engine location error ($r= -.11$, $p>.05$) nor wind accuracy ($r= .084$, $p>.05$). At higher difficulty 8x8 *Detection* trials no significant correlations were found either; fire locations ($r=.08$, $p>.05$) nor engine ($r= -.10$, $p>.05$) nor wind ($r=-.009$, $p>.05$) was significant. Thus while the difficulty in the *Detection* trials was shown to influence the error rates themselves, it was not found to influence the relationship between Level 1 SA information and WM.

Level 3 SA (*Prediction* Trials)

Another aim of this study was to examine the relationship between WM and Level 3 SA. Pearson correlations were calculated between WM factor scores and the dependent SA measures averaged across scores during *Prediction* trials (*VT-F, VT-E, RT-E,* and *idle time*).

As previously explained for Level 1 SA, an overall SA composite score was calculated as a summation of all standardized SA measures from the *Prediction* trials ($M= 0.00$, $SD=1.91$). The combination was significantly negatively correlated with WM ($r= -.291$, $p<.01$) such that higher WM was related to higher overall SA, via less time for each measure. This provides some evidence that WM is related to a composite measure of SA, and subsequent individual analyses were then used to examine the different elements.

Pearson correlations showed WM was negatively correlated with *VT-F* (r=-.212, p<.05) and mean *idle time* (r= -.249, $p<.05$), thus higher WM ability was significantly related to lower time overall to view fire engines (*VT-F*) and lower *idle time* of fire engines. This also supports the idea that awareness in the *Prediction* task is related to WM ability. However WM was not significantly correlated with mean *RT-E* (r=-.002, p>.05) or mean *VT-E* (r=-.088, p>.05).

As both of these event-related SA measures also do not appear to be improving over time, one possibility is that participants may have cognitively tunneled in the initial firefighting at the beginning of each trial and thus failed to check the surrounding non-viewable area in the simulation for new fires. Tunneling may occur for a multitude of reasons, but most primarily is due to high levels of interest or engagement and saliency within tasks (Alexander, Wickens, & Hardy, 2006; Thomas & Wickens, 2001). However, examination of the data offered no direct evidence to support such a hypothesis[1].

To explore any potential change in the correlative relationship between WM and SA over time, Pearson correlations between WM and each SA measure were examined for early and late *Prediction* trial performance. For early *Prediction* trials, WM did not correlate significantly with any of the implicit SA measures; *VT-F* (r=-.188, p>.05), *VT-E* (r=-.09, p>.05), *RT-E* (r=-.01, p>.05), or *idle time* (r=-.19, p>.05).

[1] By splitting participants into groups on their viewing of and responding to the critical event (non-maximum times) or not over the course of the 6 *Prediction* trials, I was able to address a possible tunneling explanation. Such a split gave no significantly useful pattern of findings between the groups.

However for later *Prediction* trials, *idle time* was significantly negatively correlated with WM (r=-0.27, p<.01); the other SA measures were not (*VT-F*; r=-.12, p>.05, *VT-E*; r=-.06, p>.05, *RT-E*; r=.02, p>05). This shows how the correlation between WM and implicit SA measures may be subject to change over time in performing a task.

Performance: WM and SA Correlations

Finally, the correlation between WM and performance, and SA and performance in *Prediction* trials was assessed.

A Pearson correlation was calculated for comparing WM to performance overall in *Prediction* trials (M =78.07, SD=1.81) and was found to be significantly positive (r=.331, p<.05), such that a higher WM ability was related to better performance overall in *Prediction* trials.

However given that performance changed significantly over time, it was important to investigate the WM relationship across time as well, and performance was split into early (trials 1-3) and late (trials 4-6) performance to examine this. Within early performance, WM was significantly positively correlated with performance in Prediction trials (r=.261, p<.01), and this relationship existed in a stronger form when later trials were examined (r=.345, p<.01), suggesting that higher WM ability related to higher performance on *Prediction* trials. These results are congruent with those of Rabbitt et al. (1989) suggesting ability correlates with performance in a task increase over time, rather than decreasing.

Similar analyses were calculated with the process measure of performance (*water used*) overall (M= 45.36, SD= 15.10). A Pearson correlation was calculated and showed WM significantly related to this measure (*r* = .202, *p*<.05), suggesting higher WM ability enabled participants to deploy more water in the simulation.

As this measure also changed over time, *water used* was split into early (trials 1-3) and late (4-6) usage amounts. Within both early *water usage* (*r*= .186, *p*>.05) and late (*r*= .194, *p*>.05) WM was not significantly related, though both of these effects were marginal. Therefore within process measures no change was observed in the relationship with WM over time in the *Prediction* trials.

SA and performance in *Prediction* trials.

SA may play a role in driving performance (Endsley, 1993). The SA Level 3 composite score was found to significantly negatively correlate with overall performance (*r* =-.652, *p*<.01) and with water used during the simulation (*r*=-.421, *p*<.01). This suggests that decreased time across several SA measures results in increases in performance *and* in the amount of water used in the simulation.

Examining the individual SA measures and relating them to performance may help shed light on whether any one particular measure is driving this relationship and whether this relationship changes across time. Pearson correlation coefficients showed that for early trials (again, 1-3) *VT-F* was negatively correlated (*r*=-.234, *p*<.05); *VT-E* was negatively correlated (*r*=-0.241, *p*<.05), and *idle time* (*r*=-.479, *p*<.01) was also negatively correlated with performance, suggesting that decreases in time to view or react was associated with increased performance in early *Prediction* trials. *RT-E* was the only SA measure not significantly correlated (*r*=-.002, *p*>.05).

In the later *Prediction* trials, *VT-F* was correlated with performance (r=-.318, p<.01), as was *VT-E* (r=-.273, p<.01) and *idle time* (r=-.604, p<.01). This again suggests that a decrease in the time to gather awareness information or orchestrate fire engine movements and activities is related to increased performance in later *Prediction* trials. *RT-E* was again the only measure that did not significantly relate to performance (r=-.128, p>.05).

Because of the related nature of WM and SA in their potential to predict performance in the simulation, tests for a mediating relationship were conducted with the idea that SA is a potential mediating variable between WM and performance. This relationship is in part suggested by Endsley (1993) by showing that performance does not directly depend on SA, but SA may contribute to good or bad performance.

SA was tested for mediation according to the steps outlined by Baron and Kenny (1986). WM (*beta*= .237, t(97)=2.40, p<.05) was first shown to significantly predict average performance (R^2=.056, F(1, 97) = 5.76, p<.05); then WM (*beta*=-.249, t(97)=-2.53, p<.05) was shown to was shown to significantly predict SA (R^2=.06, F(1,97) = 6.40, p<.05); and finally WM and SA standardized variables were added to a stepwise model predicting performance in which WM was no longer a significant predictor (*beta*=.13, t(96)=1.43, p>.05) after adding SA (idle time; *beta*=-.415, t(96) = -4.46, p<.001) to the model (R^2=.218, F(1,96)=19.85, p<.001); see Figure 10 for the mediation model.

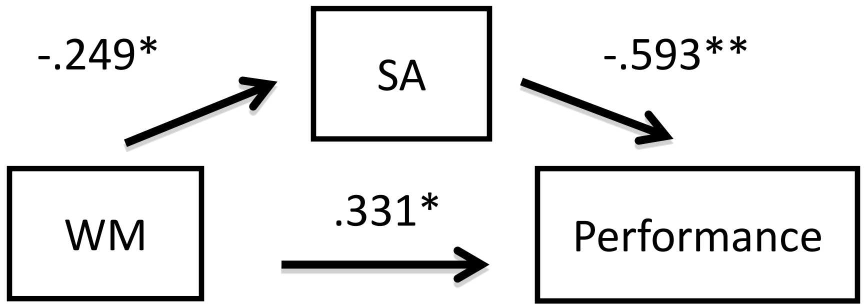


Figure 10. Mediation model for WM, SA (idle time) and average performance in *Prediction* trials.

Because the beta associated with WM was not zero, the relationship shows partial mediation (Baron & Kenny, 1986). This suggests that WM clearly influences SA and perhaps those with ability in this area are able to develop a higher level of awareness, which then furthers their performance above others. Importantly these data do not speak to the rate of development of any SA ability, simply that participants appear to be learning to be generally aware in the task.

SA and *Water Usage* in *Prediction* trials.

In addition to examining relationships between SA and performance, the relationships between SA and a process-based measure of performance, *water usage*, were also tested. Pearson correlation coefficients showed for early trials, all implicit SA measures were correlated to *water usage* except for *VT-F* : r=-.011, p>.05. *VT-E* was positively correlated (r=.248, p<.05), *idle time* was negatively correlated (r=-.741, p<.01) and *RT-E* (r=-.204, p<.05) was negatively correlated with early *water usage*, suggesting that more water used by the participant during the early simulation trials, the slower they viewed the event (but reacted faster to it), and the less *idle time* they had.

In the later *Prediction* trials, both *VT-F* ($r=-.154$, $p>.05$) and RT-E ($r=-.059$, $p>.05$) were not correlated with *water usage*. However, *VT-E* still correlated with usage ($r=.210$, $p<.05$) as did *idle time* ($r=-.855$, $p<.01$), suggesting again that the more water used, the slower participants were to view the event and the less overall idle time they had during *Prediction* trials.

CHAPTER 3: GENERAL DISCUSSION

WM and SA: Related?

It was hypothesized WM would be able to predict participant's error in locating elements of information in Level 1 based trials in NFC, as well as differences in implicit measures of Level 3 SA in later trials. However, WM did not correlate with error during Level 1 Detection trials even when a composite measure was used, and this is especially surprising given the memory-based nature of the task.

While no relationship was found for Level 1 SA, significant relationships were found between implicit Level 3 SA measures and a factor-analyzed WM component, and these represent the unique findings in this experiment. By showing implicit measures of SA are related to WM, a door opens to further study of componential SA using unobtrusive methods and also importantly alludes to WM as a true cognitive influence on SA and not a potential artifact of memory-based assessments. These results also support Endsley's (1995a; 1995b) seminal work by showing clear separation between Level 1 and Level 3 SA, especially as it relates to influence from WM ability.

Also of interest was the nature of any correlative relationship between WM and SA over time. In the case of Level 3 SA, this correlative relationship does not appear to decline over time in the task, and suggests in novices the influence of WM on SA may persist as expertise develops.

One additional distinctive aspect of this experiment relates to the WM construct used, which is uniquely devoid of other unrelated task influence. Only measures of WM previously shown to load well onto a strictly WM construct were used (Engle et al., 1999). However, one may attribute the strength of the relationships shown here as low, due to WM being a separate component of general or fluid intelligence.

The consensus of researchers on how much relation exists between WM and fluid intelligence remains mixed at best. Engle and colleagues have shown a strong relationship (.49) exists between WM and fluid intelligence and they argue a relationship exists because of the common use of the central executive (Engle et al., 1999) for both WM and fluid intelligence tasks. Ackerman, Beier and Santacreu (2005) additionally showed support in their meta-analysis that a relationship exists between WM and fluid intelligence.

However others (Colom, Rubio, Shih, & Santacreu, 2006) have shown the relationship between fluid intelligence and WM remains even after removing the variance due to an executive control component, suggesting this relationship may exist for other reasons. It is thus quite impossible to draw direct comparisons between relationships shown with SA and task performance, and those which may result from using a measure of fluid intelligence in place or in addition to WM, other than to say they may be similar.

One aspect of this assumed similarity could be found within correlations between WM and early, versus late performance shown here. Learning is assumed to be most heavily influenced by g during early learning, though the correlations with WM here increase, not decrease, over time, suggesting these two concepts are not equivalent.

Future work in this area may be interested in exploring these differential relationships, as fluid intelligence remains a mainstay in selection tests.

WM-SA and Performance

The relationship between WM and performance in Prediction trials appeared to be reflective of a similar result from Rabbitt et al. (1989) in which the association between task performance and WM ability actually increases over time in the task.

One way this may be occurring and is supported by data from the current experiment is through development of an SA “skill” in which participants with higher WM are learning to be more aware of upcoming changes in the environment compared to those with lower WM. In other words, WM may become more related to performance over time via SA development. Additionally, the relationship between SA itself and performance increases with time, suggesting that whatever awareness skill is being developed, it is aiding performance. And finally, SA was shown as a mediating variable between WM and performance in *Prediction* scenarios, which supports the postulation that WM is driving the development of awareness skills contributing to effective task performance.

These results are most likely not due to any calibration to the task or to SA measures (e.g., a repeated question may prompt users to focus on the future answering of the question, rather than the task at hand; arguably this results in something different than SA). *Prediction* trials possessed inherent complexity and varied between trials, which discourages any calibration to events and the lack of explicit SA queries meant participants had a much lower chance of anchoring on administered SA questions and then biasing responding.

Implications of Implicit SA Measures

As mentioned before, by assessing SA implicitly it is possible to avoid the large number of confounds which can be present in other methodologies (such as lack of real-world conditions, reliance on purely memory-based queries, extrinsic workload influences, and subjective responses which have shown to be biased by non-SA information.

All SA measures used in Prediction trials were implicitly assessed and successful at measuring awareness. These results show in conjunction with evidence from Gugerty and Tirre (2000) and Vidulich, Stratton, Crabtree, and Wilson (1994) that implicit SA measures are useful for assessing SA. Furthermore, finding significant correlations with WM makes a strong case for the use of implicit measures for examining cognitive relationships with SA. It fact may require this type of experimental methodology to observe WM and other cognitive construct relationships to SA in such a clear manner.

Conclusion and Future Implications

In sum, WM clearly relates to the proposed implicit measures of SA for trials that targeted Level 3 awareness. Like the relationship between performance and WM, which increased over time in NFC, SA was consistently related to performance in *Prediction* trials and this relationship was shown to increase as well, though the WM-SA relationship only grew significant in later trials. SA was additionally shown to be a partial mediator between WM and performance in this relationship.

One methodologically unique contribution to this research was utilizing factor analysis to build a WM construct out of several traditional WM tasks. By combining scores from multiple WM tasks into one construct, we can be sure the construct is well represented, and indeed in investigations where only small amounts of variance are accounted for in SA, this is a principal issue. Future work in this area may find that adopting this method results in being able to build a stronger case for WM influences.

Additional consideration may be given to the way performance is measured for tasks that involve SA. Measures of performance may reflect the awareness attained especially for novices, but they are also influenced by WM. Perhaps SA metrics can serve as an additional indicator of another operator skill, especially for novice. It remains to be seen if similar effects would be found with expert participants, though the evidence available certainly suggests experts would not be taxed to the amount that would elicit differences based on WM ability even in *Prediction*-style trials and would be more likely to differentiate as a function of experience (e.g., flight time, years of experience with the domain, etc; Sohn & Doane, 2004), not WM.

REFERENCES

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117*(3), 288-318.

Ackerman, P. L. (1989). Individual differences in skill acquisition. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences: Advances in theory and research* (pp. 165-217). New York: Freeman.

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30-60.

Alexander, A. L., Wickens, C. D., & Hardy, T. J. (2005). Synthetic vision systems: The effects of guidance symbology, display size, and field of view. *Human Factors*, *47*, 693-707.

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

Baddeley, A., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: evidence from task switching. *Journal of Experimental Psychology: General, 130*(4), 641-657.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *Journal of Personality and Social Psychology, 51*, 1173-1182.

Bellenkes, A. H., Wickens, C. D., & Kramer, A. F. (1997). Visual scanning and pilot expertise: the role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine, 68*(7), 569-579.

Bleckley, M. K., Durso, F. T., Crutchfield, J. M., Engle, R. W., & Khanna, M. M. (2003). Individual differences in working memory capacity predict visual attention allocation. *Psychonomic Bulletin & Review, 10*(4), 884-889.

Bolstad, C. A., Endsley, M. R., Howell, C. D., & Costello, A. M. (2003). The effect of time-sharing training on pilot situation awareness. Paper presented at the 12th International Symposium on Aviation Psychology, Dayton, Ohio.

Brickman, B.J., Hettinger, L. J., Roe, M. M., Stautberg D.K., Vidulich, M. A., Haas, M. W., & Shaw, R.L. (1995). An assessment of situation awareness in an air combat task: the global implicit measure approach. In D. J. Garland, & M.R. Endsley (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 339-344). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Carretta, T. R., Perry, D. C., & Ree, M. J. (1996). Prediction of situational awareness in F-15 pilots. *The International Journal of Aviation Psychology, 6*(1), 21-41.

Colom, R., Rubio, V. J., Shih, P. C., & Santacreu, J. (2006). Fluid intelligence, working memory and executive functioning. *Psicothema*, *18*(4), 816-821.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.

Dekker, S., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology, and Work, 6,* 79–86.

Durso, F. T., Bleckley, M. K., & Dattel, A. R. (2006). Does situation awareness add to the validity of cognitive tests? *Human Factors, 48,* 721-733.

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J., Nikolic, D., & Manning, C. A. (1998). Situation awareness as a predictor of performance for en route air traffic controllers. *Air Traffic Control Quarterly, 6*(1), 1-20.

Durso, F. T., & Sethumadhavan, A. (2008). Situation awareness: Understanding dynamic environments. *Human Factors, 50*(3), 442-448.

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). In *Proceedings of the National Aerospace and Electronics Conference* (NAECON) (pp. 789-795). New York: IEEE.

Endsley, M. R. (1993). Situation awareness and workload: Flip sides of the same coin. In R. S. Jensen & D. Neumeister (Eds.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (pp. 906-911). Columbus, OH: Department of Aviation, The Ohio State University.

Endsley, M. R. (1995a). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*(1), 32-64.

Endsley, M. R. (1995b). Measurement of situation awareness in dynamic systems. *Human Factors, 37*(1), 65-84.

Endsley, M. R. (2000). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley, & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 147-173). Mahwah, NJ: Lawrence Erlbaum Associates.

Endsley, M. R., & Bolstad, C. A. (1994). Individual differences in pilot situation awareness. *The International Journal of Aviation Psychology, 4*(3), 241-264.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309-331.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*(2), 211-245.

Gonzalez, C., & Wimisberg, J. (2007). Situation awareness in dynamic decision-making: Effects of practice and working memory. *Journal of Cognitive Engineering and Decision Making, 1*(1), 56-74.

Gugerty, L. J., & Tirre, W. C. (2000). Individual differences in situation awareness. In M. R. Endsley, & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 249-276). Mahwah, NJ: Lawrence Erlbaum Associates.

Jones, D. G. (2000). Subjective measures of situation awareness. In M. R. Endsley, & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 113-128). Mahwah, NJ: Lawrence Erlbaum Associates.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working-memory capacity: A latent variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*, 189-217.

Lavie, N., & De Fockart, J. (1995). The role of working memory in attentional capture. *Psychonomic Bulletin & Review, 12*(4), 669-674.

O'brien, K. S., & O'hare, D. (2007). Situation awareness ability and cognitive skills training in a complex real-world task. *Ergonomics, 50*(7), 1064-1091.

Omodei, M. M., & Wearing, A. J. (1995). The Fire Chief microworld generating program: An illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision-making behavior. *Behavior, Research Methods, Instruments, & Computers*, *27*(3), 303-316.

Omodei, M. M., & Wearing, A. J. (1998). Network Fire Chief (Version 1.0) [Computer Program]. La Trobe University.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*(2), 140-160.

Pritchett, A. R., & Hansman, R. J. (1997). Use of testable responses for performance-based measurement of situation awareness. In M. R. Endsley, & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 189-209). Mahwah, NJ: Lawrence Erlbaum Associates.

Rabbitt, P., Banerji, N., & Szymanski, A. (1989). Space Fortress as an IQ test? Predictions of learning and of practiced performance in a complex interactive video-game. *Acta Psychologia, 71,* 243-257.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*(1), 4-27.

Sohn, Y. W., & Doane, S. M. (2004). Memory processes of flight situation awareness: Interactive roles of working memory capacity, long-term working memory, and expertise. *Human Factors, 46*(3), 461-475.

Thomas, L. C., & Wickens, C. D. (2001). Visual displays and cognitive tunneling: Frames of reference effects on spatial judgments and change detection. *Proceedings of the 45th Annual Meeting of the Human Factors and Ergonomics Society*, Santa Monica, CA.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498-505.

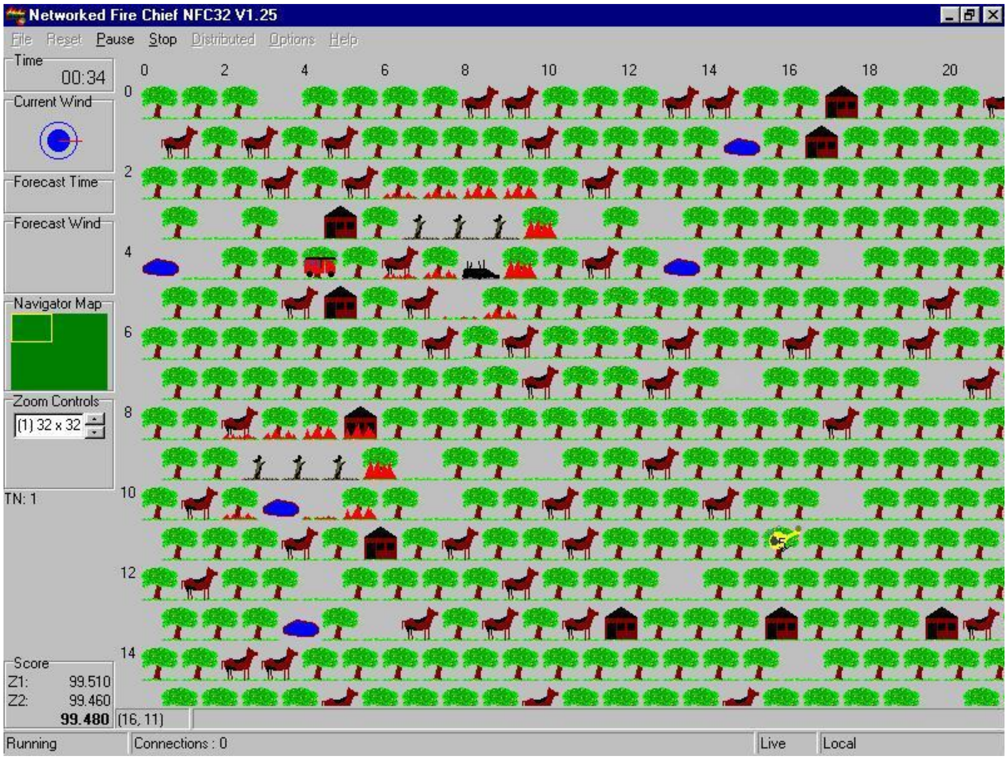
Unsworth, N., Shrock, J. C., & Engle, R. W. (2004). Working memory capacity and the antisaccade task: Individual differences in voluntary saccade control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1302-1321.

Vidulich, M. A., Stratton, M., Crabtree, M., & Wilson, G. (1994). Performance-based and psychological measures of situational awareness. *Aviation, Space and Environmental Medicine, 65*(5, Suppl.), A7-A12.

Wickens, C. D. (2008). Situation awareness: Review of Mica Endsley's 1995 articles on situation awareness theory and measurement. *Human Factors, 50*(3), 397-403.

Yanco, H. A., & Drury, J. (2004). "Where Am I?" Acquiring situation awareness using a remote robot platform. In *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics.*

Appendix A: Networked Fire Chief Example



Appendix B: Networked Fire Chief Instructions

Using Networked Fire Chief

Goal: use resources (fire truck, helicopter, and bull dozer) to control or extinguish fires as much as possible

- Navigating: navigate map using green "Navigator Map" on left side of the screen
    - Click and hold down left mouse button while dragging yellow rectangle to desired location. The yellow rectangle encompasses the area that is currently visible.

Icon Key

Forest Pasture House Clearing Burnt Forest Burnt Pasture Burnt House Burnt Clearing

- To move an appliance: when an open hand appears, left click on the appliance and drag to the desired cell
- Appliances are full at the beginning of the simulation
- To treat the cell (with water): left click the appliance in the desired cell
    - The appliance cannot be moved when it is treating a cell, and it can only be moved when the open hand is present when hovering over the appliance.
- To refill the appliance with water: drag the appliance to the desired dam  and left click. After the appliance is refilled, it must be moved to the desired location.
- It is possible to use all the water in a dam. It will appear  as more water is used. When no water is left, the dam will appear completely brown.
- Fire intensity is signaled by flame size. As flame appears larger, intensity is greater.
- Useful Information
    - Time: located on the left
    - Current Wind: updated continuously throughout the simulation
    - Information about appliance is displayed at the bottom of the screen. The following types of information are displayed
        - Water level
        - If fire is too intense for the appliance
        - If the appliance is moving