THE MAKING OF D-SAT: THE DEVELOPMENT AND TESTING OF DYNAMIC SITUATION AWARENESS TASK

by

MARGO M. WOLLER

A.A.S., Michigan Technological University, 2004
B.S., Michigan Technological University, 2006

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Department of Psychology
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2010

Approved by:

Major Professor
Dr. James Shanteau

Abstract

Situation Awareness (SA) measurement takes on many forms: subjective, direct, and implicit performance, each with limitations. Subjective measures are based on self and peer reports, which allow biases to enter the measurement. Direct measures, such as SA Global Assessment Technique (SAGAT), interrupt SA in order to probe the participants' SA level using questions. Implicit performance measures are based on participants' ability to complete SA tasks, which must be created for each domain. A new approach, Dynamic – SA Task (D-SAT), was developed using a microworld wildfire fighting simulation, Networked Fire Chief (NFC). D-SAT is an implicit performance measure that can be adapted to multiple domains, for example inattentional blindness. Scenarios were developed during study one by tracking participant performance and scenario situations. Study two used the scenarios developed during study one to test D-SAT's ability to evaluate SA by comparing D-SAT performance to an established SA performance measure, situation awareness global assessment technique (SAGAT). While the manipulation used to create had an effect on D-SAT performance, it was not associated with the established SA performance measure. However, a signal detection theory (SDT) analysis showed additional promise for D-SAT being a useful SA measure.

Table of Contents

List of Figures ... vi
List of Tables ... vii
Acknowledgements ... viii
Dedication ... ix
CHAPTER 1 - Introduction ... 1
SA Measures ... 1
Subjective Measures ... 2
Direct Measures ... 3
Implicit Performance ... 3
D-SAT ... 4
Inattentional and Change Blindness ... 4
CHAPTER 2 - Method and Results ... 8
Study 1 ... 8
Participants ... 8
Materials ... 8
Procedure ... 9
Results and Discussion ... 10
Study 2 ... 12
Participants ... 12
Materials ... 12
Design and Procedure ... 13
Results ... 14
Signal Detection Theory (SDT) Analysis ... 15
Discussion ... 16
Figures and Tables ... 19
CHAPTER 3 - General Discussion ... 25
Limitations ... 25
Possible Implications and Future Research ... 26

References 28
Appendix A - SAGAT-like questions.. 31
Appendix B - Demographics Questionnaire 33

List of Figures

Figure 2.1 D-SAT Scenario Time Course.. 19
Figure 2.2 Interaction of Speed and Priority of Fire Two on D-SAT Performance 19
Figure 2.3 Pictorial Depiction of SDT... 20

List of Tables

Table 2.1 Fire Characteristics for Each Scenario Type 21

Table 2.2 D-SAT Scenarios Order and Characteristics 22

Table 2.3 Means and Standard Deviations of D-SAT Performance in Percent Landscape Remaining for Speed and Priority by Fire and Level 23

Table 2.4 SDT Scenario Classifications 23

Acknowledgements

I would like to thank my advisory committee for their help and suggestions throughout my thesis work. I would also like to thank the Cognitive Open Seminar and Decision Research Group for their input on early ideas and continued support during the process.

Dedication

I would like to dedicate my thesis to my wife, Angela, and family for their love and support during my educational and personal growth.

CHAPTER 1 - Introduction

Situation awareness (SA) has been defined differently across disciplines; however, SA is broadly defined as "knowing what is going on around you" (Endsley, 2000a). Endsley (1995) breaks the broad definition of SA into three levels. The lowest level of the SA is *perception*, which involves the intake of information from the environment and situation need for the higher levels of SA. The next level of SA, referred to as *comprehension*, involves interpreting the perceptual information from the environment and situation into consequential information. The third level of SA is *projection*, using the information gained from the first two levels of SA to anticipate future events, and allowing decisions to change the current course of action to adapt to the anticipated future states of the situation.

Numerous measures of SA are defined by the information needed to complete the specific task; those vary across situations. A variety of measures have been developed for use in many disciplines; however, the measures are often limited to the specific situation for which the measure was developed. These task-specific measures include measures for designed to test fighter pilots, air traffic controllers (ATC), and commercial airline pilots, as well as fire fighters. However, many of the SA measures currently used have been criticized for their limitations.

The present study incorporated information gained through the use of current SA measures to develop a new measure of SA, Dynamic-SA Task (D-SAT). D-SAT can be adapted to be used in multiple situations. In addition to creating a new SA measurement technique, the results have the potential to incorporate work on inattentional and change blindness into the study of SA.

SA Measures

Measures of SA can be grouped into three categories, subjective, direct, and implicit. Each type of SA measure has advantages, disadvantages and a wide range of measures that fit into each category. These three categories will be explored in more detail before discussing the D-SAT in greater detail.

Subjective Measures

Subjective measures can be collected using self-ratings, observer ratings, or a combination of the two. Jones (2000) describes four subjective measures techniques that have been used to evaluate SA: unidimensional scales, SA Rating Technique (SART), SA – Subjective Workload Dominance (SA-SWORD), and SA Rating Scale (SARS). Unidimensional scales present a line and participants are asked to rate their level of SA. Unidimensional scales are easy to administer and interpret; however, because only one scale is used, the technique may fail to capture all levels of SA and are easily altered by participants' interruptions of what is being measured.

Similar to the unidimensional scale, SART (Selcon & Taylor, 1990) asks participants to rate themselves on a line with the endpoints of low and high. The scales are related to statements linked to SA. Also, 3-D SART statements are a demand on mental resources as well as on the supply of attentional resources, and understanding. 10-D SART has also been developed with ten general statements related to SA. While SART is more informative than the unidimensional scale, it has been criticized for adding to the mental workload of the participant and in turn altering SA (see Endsley, 1996 for additional criticism).

Other unidimensional scales have been developed and show promise, however they are not being tested. The SA-SWORD (Vidulich & Hughes, 1991) is a variant of SWORD, a scale developed to assess workload, and uses mathematical algorithms to complete pair-wise comparisons of the data collected. SA-SWORD holds promise. However there are few studies utilizing the technique, which leads to a lack of evidence of effectiveness.

The SARS (Waag & Houck, 1994) is composed of 31 elements, which are rated by both participants and those close to the participant, such as supervisors and team members. SARS measures not only SA, but contains elements that are not directly related to SA, such as personality traits. SARS allows for a comprehensive assessment of SA in aircraft context i.e. pilots and co-pilots. However, the scale is closely tied to this context and thus, allows for little versatility.

While subjective measures have the potential to capture SA, most are limited to the context for which they were developed. Additionally, subjective measures do not require the participant to use SA to complete the measures, resulting in scores dependent on the individuals understanding of SA. The interpretation of the question by the participant is a variable as it introduces an uncontrollable error rate into the measure of SA. Participants' interpretation is just one of the issues of self reported data. Consequently, subjective measures of SA are influenced by the method of data collection used to measure SA. As a result subjective measures of SA are highly subjective, as the name would suggest.

Direct Measures

Direct measures are administered during jobs that require SA as well as, they must incorporate a task that evokes SA while measuring the level of SA during the task. Direct measures use accuracy and/or time to measure SA.

The most commonly used direct measure of SA is the Situation Awareness Global Assessment Technique (SAGAT; Endsley, 2000b) relies on the accuracy of the participants' responses to questions to assess SA. SAGAT uses a list of possible questions presented during artificial stoppages of the primary task. Participants are asked to complete randomly selected questions from the SAGAT list, during the stoppages, which are in turn used to assess SA. Sessions must balance SAGAT items in order to reduce anticipation effects, i.e. focusing on the information that will be needed to answer the SAGAT questions. The balancing is done by having some sessions with no SAGAT information gathered as well as never having the participants complete the same SAGAT question(s) twice.

SAGAT is most often criticized for stopping the primary task in order for the SAGAT measures to be taken (Sarter & Woods, 1995 and Burns et al, 2008). The criticism arises because SA can be easily disrupted by interference of the main task by external distracters, such as added cognitive workload of a secondary task or unexpected stoppage of the primary task. Critics feel that by stopping the primary task, any SA that was present before the stoppage is lost when the stoppage occurs (Sarter & Woods, 1995 and Burns et al, 2008). One way to avoid the disruption is to use an implicit measure of SA.

Implicit Performance

Implicit performance measures have subject matter experts (SMEs) rate the participants' SA based on their accuracy and reaction times during a SA task; which, allows SA to be measured without disruption of the SA task. The implicit nature of the measurement is that the researcher is inferring that changes in accuracy and reaction time are a result of changes in SA. The current list of implicit performance measures is limited due to most simply being a measure of reaction time. The Pilot Performance Index (PPI), described by Venturino, Hamilton, and Dvorchak (1989), equals the ratio between the number of enemies killed to the number of allies killed. Other measures include reaction times and number of errors committed, but have not been named as they were not developed to test SA. Few measures have been explicitly developed in this area and the D-SAT measure could be added to the list of implicit measures of SA.

D-SAT

In order to avoid disruption of SA, D-SAT uses a microworld simulation, to collect SA data without interruption of the task. The simulator can also be programmed to create different situations allowing for a variety of phenomenon to be investigated. In a previous study, Woller, Park, Burrell, Hilgenkamp, Vowels, and Shanteau (2008), has shown that performance scores do not differ across cover stories and displays when the task is held constant. This study shows that scenarios can be developed for many disciplines by changing the appearance of the icons in the landscape. This allows for many different contexts to be used which lead to the same performance results.

For the current study the researcher programmed situations which allowed for the investigation and assessment of SA. Preliminary research on using a microworld simulator to measure SA by Omodei (1995 and personal communication, October 18, 2006) shows promise for using a simulator for SA research in many contexts, such as expert firefighters and military. D-SAT offers an easy-to-administer and adapt SA measure which does not disrupt SA while measuring SA, as D-SAT does not require the task to be stopped to measure SA. The simulator is both the stimulus and data collection device; which scores each participant’s performance using a weighted average of terrain remaining. The participants are able to see their score while completing the task allowing them to test resource management approaches. After testing the approaches, participants will choose an approach to the problem, which allows them to gain the highest score possible. Participants must decide how to allocate resources across multiple problems. However, in order to place resources on a problem the participants must first recognize the problem.

Inattentional and Change Blindness

D-SAT scores depend on how quickly a participant reacts to a new problem in the simulation. If participants fail to perceive a change or comprehend the significance of a new fire, their SA score will be lower than participants who were able to perceive and comprehend the change. The failure to perceive a change in the environment is known as change blindness, which is a subclass of inattentional blindness, the failure to direct attention to important aspects of the environment. Inattentional blindness is a phenomena discovered during attentional studies of the 1960's and 1970's. During the late 1980's and early 1990's inattentional blindness became a focus of study. However, inattentional blindness was not popularized until the late 1990's by Simons & Levin (1997). Countless studies have been completed since this pivotal work; many dealing with failures that occur in complex environments and tasks.

For example, research on the use of cell phones while driving has made a connection between SA and inattentional blindness research. Strayer, Drews, and Johnston (2003) asked participants to engage in cell phone conversations while using a driving simulator, in order to test the effects conversations have on SA. Participants had longer reaction times and lower memory for roadside billboards while having a phone conversation than when not. Eye tracking results indicated that memory failures were due to not encoding the information even when the billboard was fixated on. The failure to encode information which has been fixated on is referred to as "looking, but not seeing," or inattentional blindness.

The research on SA and inattentional blindness was extended by McCarley, Vais, Pringle, Kramer, Irwin, and Strayer (2004) again using cell phone conversations' effect on driving simulator performance. McCarley et al. found that conversations caused more change blindness than the task of driving alone. The effect was influenced by the age of the participant. This research suggests change blindness can be associated with a failure at the perception level of SA. The failure in perception occurs when the information is looked at but not encoded or seen. A large body of literature involves cell phone use and driving failures; however this is not the only research being done on SA and inattentional and change blindness.

Durlach, Kring, and Browens (2008) investigated change blindness during a different SA task. Soldiers viewed a display of the current combat situation, which was updated on a predetermined schedule. Upon the updating displays, the soldiers were asked to report any changes that had occurred in the display. This research paradigm is similar to D-SAT; however when using D-SAT, participants do not report the occurrence of new fires verbally, but instead by reacting to the new fire. Inattentional blindness studies tend to be laboratory studies. However, research on traffic accidents can also be used to close the gap between SA and inattentional blindness such that higher levels of inattentional blindness should lead to lower levels of SA.

An analysis of 500 traffic accident reports by Koustanai, Boloix, Van Elslande, and Bastien (2008) found accidents could be caused by one of two failures. The drivers either (1) failed to see the danger, a failure of the perceptual stage of SA; or (2) failed to react to a perceived danger in time to make a difference in the outcome, a failure of either the comprehension or the projection stage of SA. The findings of complex environmental studies can be backed up by laboratory research.

Laboratory studies of inattentional blindness have increased the understanding of the cognitive processes which underlie inattentional blindness. Mitroff, Simons and Levin (2004) have shown that participants are able to report the pre and post change item even if they are unable to report a change in the item. The ability to recall both the pre and post change item implies the information is encoded, but the comparison between the items requires attention to be drawn to the object of interest. For example, if the object of interest is randomly placed in the display participants must attend to the entire display to look for changes, making it difficult for participants to detect a change. However, if the participants were told the area the object would change in they are more likely to attend to the object of interest and detect a change in the display. This may affect participants' performance on D-SAT if they fail to direct their attention to the important change of a new fire developing. While the link between SA and inattentional blindness can be seen in such studies, new methods for measuring SA are required before the gap can be bridged.

Novel SA measurement methods require the incorporation of the knowledge of cognitive psychology into the measurement and knowledge of SA. Cognitive psychology is a more mature discipline of psychology, and has knowledge that can be applied to the measurements of SA to improve the outcome of measurement techniques. SA tends to be used as a description of performance instead of a cognitive process. D-SAT allows the knowledge gained from inattentional blindness research to be incorporated into SA research, by using a change blindness paradigm to measure SA.

CHAPTER 2 - Method and Results

Two studies were conducted to develop and test D-SAT as a measure of SA. Study one was a pilot study used to determine the length of the scenarios, as well as the speed and timing of the fires; these measures were then used to develop the scenarios used in study two. Study two tested D-SAT as a SA measurement method by comparing D-SAT performance to performance on a known measure of SA.

Study 1

The purpose of study 1 was to develop the materials and scenario lengths to be used in study 2. The scenario lengths and timing of fires were manipulated to make sure participants were forced to make a choice between fire one and fire two; while simultaneously keeping participants engaged in the task.

Participants

19 participants, 7 females and 12 males with a mean age of 19.39, and a range from 18 to 25, from the psychology research pool were given class credit for participating in the study.

Materials

In this study Networked Fire Chief (NFC), a dynamic decision-making simulation, is the task performed by participants (Omodei & Wearing, 1995). NFC simulates a fire-fighting task in which participants have a limited amount of resources, and must save as much of the terrain as possible. NFC functions allow researchers to design different scenarios/tasks for the participants, which allows for assessment of SA tasks.

The simulator used in D-SAT was NFC, which was used as both for stimulus presentation and as the data gathering tool. NFC is a microworld simulator designed to research dynamic decision-making in the wildfire fighting domain. The simulator allows researchers to design tasks to answer different research questions in multiple domains, by manipulating the fires/events and the look of the landscape features. The wildfire fighting, original, domain was used as there was no difference in performance scores across domains in previous research conducted to determine if cover story/domain effected task performance in NFC (Woller et. al, 2008).

Each NFC task is referred to as a scenario. This study used scenarios begins with a primary fire that occurs soon after the simulation began. After the primary fire had time to build, a secondary fire occurs. These scenarios were a subset of the scenarios used in study two selected at random to vary the length of scenarios. By manipulating the speed, fast or slow, and priority, high or low, of the fires in the scenarios, three scenario types were created: SA, Non-SA and filler.

SA scenarios required the participant to switch from fire one to fire two to gain the highest possible performance score. Non-SA scenarios either required the participant not to switch to fire two to achieve the best performance possible, or it did not matter which fire the participant fought as the fires had the same speed and priority. Three filler scenarios were created to reduce the anticipation by hindering the ability to predict the number of fires in each scenario. These filler scenarios were comprised of 3 different formats: 1) one continuous fire with wind speed and intensity changes; 2) two fires from the start of the scenario with wind speed and intensity changes; and 3) three fires with one fire at the start of the scenario, a second added with growing relative intensity, followed by a third requiring the most attention. The participant can monitor their scores across scenarios to determine the best course of action relative to decisions made in each scenario. Non-SA and filler scenarios were included to reduce anticipation of SA measurement.

This study varied the length of the scenarios between 2, 2.5, 3, 3.5, 4, 5, and 6 minutes, in order to determine the best length to use in study two. Previous studies have used 6-minute scenarios; however, participants reported boredom during the long scenarios (Woller et. al, 2008). Additionally, study two required each participant to complete 24 scenarios in one hour, so a shorter average scenario length was required.

Procedure

Participants completed an informed consent form before completing the simulations. Participants then completed 15 to 24 scenarios in a randomized order varying in length and type. Three of the nine required cells of the fractional factorial design scenarios were randomly chosen, using random.org, to be used as test scenarios (see Table 2.2 for full design and cells tested). The three scenario types include two Non-SA scenarios; Fast Low to Slow Low and Slow High to Fast Low and one SA scenario, Slow Low to Fast High. Seven different scenario lengths, 2, 2.5, 3, 3.5, 4, 5 and 6 minutes were developed to start testing the scenario timing. In addition, three filler scenarios were developed. The number of scenarios completed increased as the specifications of the scenarios timing was finalized. Scenario timing was developed to have a consistent time of wind changes and fire intensity for each speed level, by reviewing scenario performance between each participant's sessions. Additional scenarios were developed using the scenario timing, gained through this continuous process.

After non-SA scenarios, participants were given a SAGAT-like question regarding the last scenario they completed which was used to determine if the participants noticed changes in the display. The question were similar to the type used in SAGAT experiments, however they were specific to the firefighting task. Additionally, the SAGAT-like questions were given at the end of the scenarios instead of during a stoppage of the scenario. The SAGAT-like questions were designed to probe the participants' memory of the scenario and instructions, in order to determine if the participants encoded information of the entire scenario (see Appendix A for example of the SAGAT-like questions). Participants were then debriefed and thanked for their participation.

Results and Discussion

Qualitative analysis of the participants' performance was conducted by recording if fire one was still burning when fire two started. This was required in order for participants to be faced with a choice between the fires. If either of the fires burned out prior to the end of the scenario, the scenarios were adjusted by changing the wind speed and the time of fire onset. The length and difficulty of the scenarios were varied until half the participants switched to fire two on at least half of the scenarios completed. The one half of the participants detecting one half of the changes, or the one half of one half rule, was first adopted in psychophysical experiments and was used in this study as the base level of SA is perception. These times were 2.5 and 3 minutes. Having two or more different lengths was determined to reduce anticipation of the scenario end. When one length was used participants did not work to the end of the scenario, but instead anticipated the end of the scenario and stopped reacting to the fires. The point of stopping varied between participants; which lowered the participants' performance on the scenarios and added unwanted variance to the analysis. Thus 2.5 and 3 minute scenario lengths were chosen to be used in study 2, as well as implementing scenarios created using the scenario timing standards.

Study 2

Study two used the scenarios and lengths developed during study one to test D-SAT. Performance on D-SAT was compared to performance on SAGAT-like questions as well as using a signal detection theory (SDT) analysis (Greene & Swets, 1966) in order to determine the validity of D-SAT. It was hypothesized that performance on D-SAT would be lower for participants with low SAGAT-like question performance, given that SA is required to complete D-SAT and participants reactions to the environment effect their D-SAT and is a result of their level of SA. Additionally, it was hypothesized that performance on D-SAT would be positively correlated with SDT sensitivity.

Participants

30 participants, consisting of 20 females and 10 males, with a mean age of 19.53 and a range of 18-25, from the psychology research pool were given class credit for participating in the study.

Materials

NFC was used as both the stimulus and data gathering tool for D-SAT. The D-SAT scenarios, designed to measure SA, were composed of a primary fire that occurs soon after the simulation began. After the primary fire had time to build, a secondary fire occurs. The simulator recorded the time participants made actions on the environment and what was occurring in the environment.

Multiple scenarios were created. Some scenarios measured SA, while others were used to reduce anticipation of the SA-measuring scenarios. The type of scenario, SA or non-SA was varied by manipulating the speed and priority of each fire.

Each fire can be either fast or slow moving, or either high or low priority. Priority was varied by the scores assigned to landscape elements. NFC uses a weighted average of the landscape saved to create a performance score. Participants start with a score of 100%, and as the fires burns, their score is lowered. Houses and cows were worth relatively more points, while trees and grass were less. Participants were instructed how NFC scored their performance using a weighted average of landscape remaining and told their goal was to keep the score as high as they could; by protecting high value landscape elements, houses and cows. If the participants are not attuned to the changes in the environment, they will not react correctly to the fires causing a decrease in their performance score.

Nine different scenarios were created in order to test the main effects of speed and priority on performance, as well as the first order interactions between speed and priority (see Table 2.1). The Slow Low to Fast High scenario is shown in Figure 2.1 to illustrate the time course of a scenario. Three additional filler scenarios were included to reduce anticipatory effects, one fire with a wind change (1F), two fires from the start of the scenario with a wind change (2F), and a three fire scenario, similar to the test scenarios, with the addition of another fire breaking out after the second (3F). The lengths of the scenarios were either 2.5 or 3 minutes long to reduce the anticipation of the end of the scenarios.

In addition to the NFC scenarios, the SAGAT-like questions used in study one, were asked following 12 of the non-SA and filler scenarios (see Appendix A for complete list of questions and Table 2.3 for their location within the scenarios). The scenarios the SAGAT-like questions were asked after was determined using a random permutation generator (random.org). The SAGAT-like questions used the same type of questions as the original SAGAT; however, the questions are asked after the SA test rather than during a stoppage of the SA task. The SAGAT-like questions were used to determine if participants were aware of changes in the scenario display. Additionally, the SAGAT-like questions give a known SA measure to compare D-SAT performance. If the participants were unable to answer a question, they were asked to give their best guess as to what the answer would be. The complete list of questions can be found in Appendix A.

Design and Procedure

A 2 (Speed of Fire 1, Slow and Fast) x 2(Priority of Fire 1, Low and High) x 2 (Speed of Fire 2, Slow and Fast) x 2 (Priority of Fire 2, Low and High) fractional factorial design was used (see Table 2.2 for full design and cells tested). The cells tested were selected using JMP IN (Version 4.0.4) to determine which cells were required test the main effects of fire speed and priority, as well as the first order interactions between fire speed and priority. Participants completed each scenario twice, once for each scenario length (2.5 and 3 min), which allows for the use of repeated measures to compare participants' D-SAT and SAGAT-like question performance.

Participants completed an informed consent form before completing the simulations. Participants then went through the 24 scenarios (6 SA, 12 non-SA, and 6 filler) in a predetermined randomized order, which was the same for all participants. The order made sure that a scenario was not completed after the same scenario consecutively. Additionally, SA test scenarios were not introduced until the participant had learned how to use the NFC commands (see Table 2.3 for order of scenarios).

After 9 non-SA scenarios and 3 fillers, participants were given a question regarding the last scenario they completed (see Table. 2.3 for the scenarios questions were asked after). Question responses were used to determine if the participants noticed changes in the display. The SAGAT-like questions are a known measure of SA (Burns et. al, 2008) and were used to compare D-SAT performance to. Participants also completed a demographic questionnaire, (see Appendix B) and were then debriefed and thanked for their participation.

Results

Paired t-tests of D-SAT performance scores were conducted to determine if the scenarios could be collapsed over scenario length. The analysis revealed that all pairings (t (27) $\geq$ 2.21, $p \leq$ .036) aside from Fast Low to Fast Low (t (29) = -1.09, p = .284) and Slow High to Slow Low (t (28) = 1.57, p = .127) could be combined.

Data were analyzed using a one way repeated measures analysis of variance, which allowed for the investigation of the effect of the speeds and priorities of each fire on scenario performance. Table 2.4 displays the means and standard deviations of this analysis. The main effects of speed of fire one (F (1, 29) = 162.42, $p < .001$), speed of fire two (F (1, 29) = 77.44, $p < .001$), and priority of fire two (F (1, 29) = 50.30, $p < .001$) were statistically significant, and (priority of fire one F (1, 29) = 0.26, p = .61) was not significant. In addition, a statistically significant interaction between speed of fire two and priority of fire two (F (1, 29) = 68.91, $p <$ .001) was found (see Figure 2.2). However, no other interactions were statistically significant (F (1, 29) $\leq$ 0.88, p $\geq$ .35).

The SAGAT-like questions were hand scored for correctness (see Appendix A for correct answers and number of participants answering each question correctly). One point was given for correct answers with a total possible score of 17 points (see Footnote). The scores ranged from 5 to 13 with a mean of 8.47 points. Scores were then changed to percent correct by dividing each score by 17 for ease of comparison, resulting in a range of 29.41 to 76.47 and a mean of 49.80 percent correct. The mean percent of outcome dictates that either the participant has a low level of SA, or that the test is not sensitive to SA.

In order to investigate if the scenario performance was associated with the SAGAT-like questions score, correlations were performed on each scenario type, after combining across times using the average of the two scores, and SA question score. Only one scenario performance score was significantly correlated with the SAGAT-like questions, Slow Low to Fast High (r (28) = .639, p = .000). Further investigations of the descriptive statistics of these scenarios showed no consistent differences in scenario events or participant performance between the significant and non-significant scenarios. The lack of significant results increases the evidence of the SAGAT's shortcomings, and requires additional analysis to determine the usefulness of D-SAT.

Signal Detection Theory (SDT) Analysis

In addition to comparing scenario performance with SA question performance, a SDT analysis was conducted to investigate participant performance and switching behavior. The switching behavior, the likelihood of a participant to switch, and on which scenarios the participant switched from fire one to fire two, were coded to determine the Hit, False Alarm (FA), Correct Rejection (CR), or a Miss. Scenarios were labeled as a hit if the participants switch when they should switch to gain the best possible performance score. If the participants switched on a scenario that did not require a switch to gain the best possible performance score the scenario it was labeled as FA. A CR was scored if they did not switch on a scenario that required no switch to gain the best possible performance score. A Miss was deemed to be scenario that required a switch to gain the best possible performance score but the participant failed to switch from fire one to fire two (see Table 2.5). Switching behavior was coded for all scenarios; however, only scenarios with a definitive switch behavior, i.e., participant moved at least one vehicle to the new fire prior to the end of the scenario, were included in calculations. From this information d' and β were calculated for each participant using the formulas below (see Figure 2.3 for a pictorial representation of signal detection theory). The switching behavior can be used to classify participants in three types; switchers, non-switchers, and no preference (see Table 2.6

for examples), which can be used as another indication of the cognitive processes underlying SA (Abdi, 2007).

$$d' = Z_{Hit} - Z_{FA}$$

$$C = -\frac{1}{2}(Z_{Hit} - Z_{FA})$$

$$\beta = d' * C$$

To assess the reliability of the coding independent raters coded two scenarios for all participants with 99% agreement between the raters. The mean d' score was 1.57 with a standard deviation of 0.71 and a ranged from -0.10 to 2.61. To determine if d' was associated with SAGAT-like question performance a Pearson correlation was conducted resulting in a marginally significant correlation, r (29) = .35, p = .060. The β's ranged from -1.26 to 0.96 with a mean of -0.39 and a standard deviation of 0.69.

Negative β's are associated with liberal biases, i.e., a propensity to say yes a switch is required. In order to determine if β was associated with the rate of switching on possible switch scenarios, see Table 2.3 for list of possible switch scenarios, i.e. same priority and speed for fire one and two, a Pearson correlation was conducted. The correlation between β and a possible switch scenario switch rate was statistically significant, r (29) = -0.49, p = .006. The negative correlation occurs because the liberal, negative β is associated with more switching behavior.

Discussion

These results indicated that the speed and priority of the fires did have an effect on performance of NFC scenarios, indicating that the speed and priority manipulation used to measure SA was strong enough to create the effect needed for the current study. There was a statistically significant interaction between speed and priority of fire two, adding support to an efficient manipulation of speed and priority.

Additionally, only one scenario's performance, Slow Low to Fast High, was significantly correlated with SAGAT-like question performance. The scenario with a statistically significant correlation between the scenario and SA question performance had no consistent differences from the non significant scenarios, other than the significant scenario had obvious indications of a switch being required, indicating the drastic difference between the two fires and making the choice to switch apparent to the participants.

Furthermore, performance on the SAGAT-like questions was low, which could indicate that the participants used for the study may not have had the level of SA required to test SA using D-SAT, due to a lack of fire fighting experience. Conversely, SAGAT may not be an adequate comparison instrument due to the shortcomings of the measure. Future studies using D-SAT may need to use expert participants to determine D-SAT's full potential as an SA measurement method. Additionally, the performance on the scenarios was related to the performance on the SAGAT-like questions for only two of the scenarios. This may mean that the scenarios created to test SA were actually testing another construct, such as time allocation to tasks, or the SAGAT-like questions may be too hard for novice participants. In order to determine if the SAGAT-like questions were responsible for the non-significant results with D-SAT, additional SA measures must be compared to D-SAT performance. The D-SAT performance measure was not correlated with SAGAT-like question performance; although, the SDT analysis was correlated with question performance.

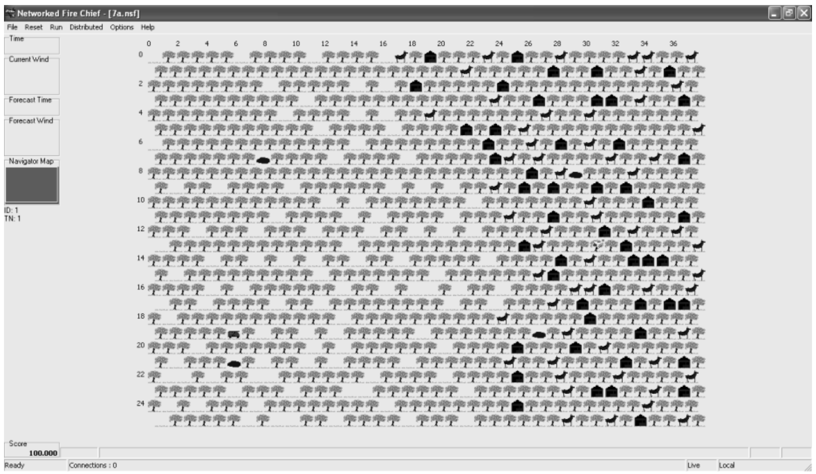
In particular, a marginally significant correlation was found between d' and SAGAT-like question performance, and a negative correlation between β and SAGAT-like question performance. The weak correlation between d' and SAGAT-like question performance illustrates weak manipulation of speed and priority. However, the negative correlation between β and possible switch scenarios' switching behavior is a reflection of how β is calculated, because a negative β means the participant has a tendency to switch. If the correlation would not have been significant an error during the coding or calculations would have occurred causing the SDT analysis to be non-significant. The mixed results suggest additional testing of D-SAT is required to determine whether D-SAT or the SAGAT variant is the cause of the current result, prior to wide-spread use.

Footnote: The analysis of the SAGAT-like questions was also completed by scoring the questions for partial correctness. A half point was given for close answers; however, there was no difference in the significance of the results between the two sets of analysis. As such the simpler scoring rule was adopted for the analysis of the SAGAT-like questions.

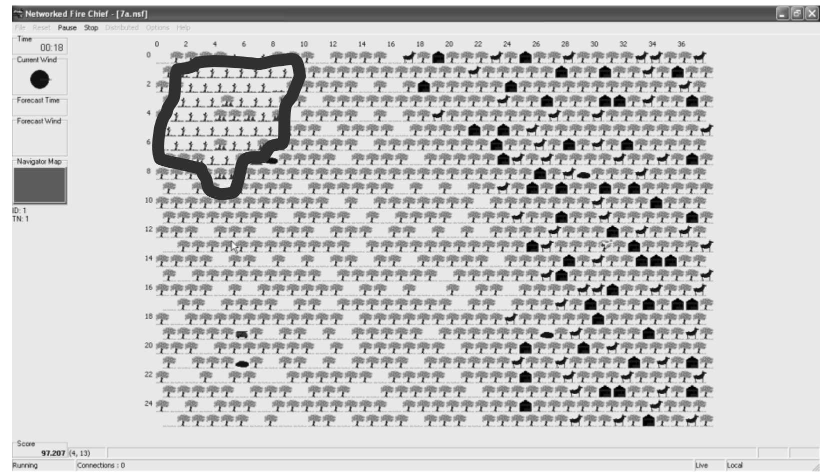
Figures and Tables

Figure 2.1 Time Course for Slow Low to Fast High D-SAT Scenario

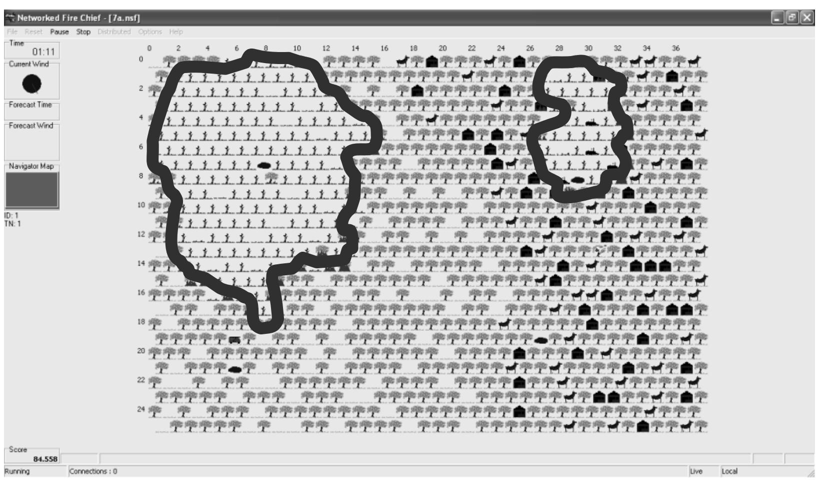
a)



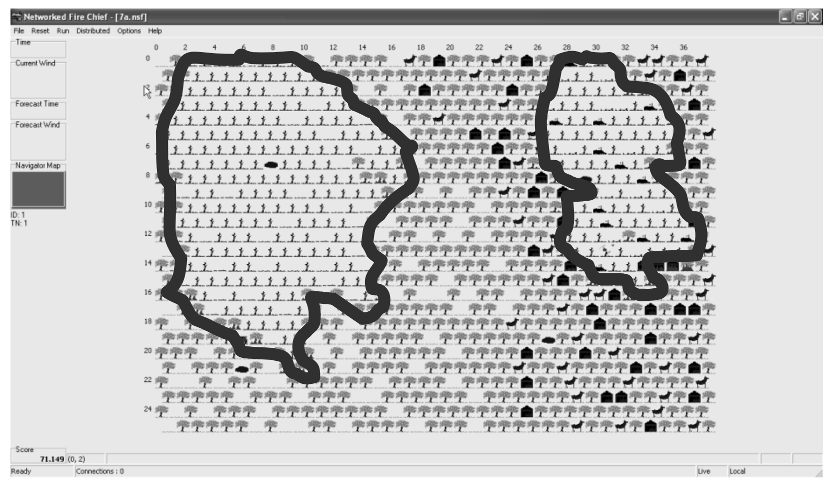
b)



c)



d)



Note: Slow low priority fire followed by a fast high priority fire at a) time equals zero, b) time equals 18 seconds, c) time equals one minute and 11 seconds, and d) time equals two minutes and 30 seconds.

Figure 2.2 Interaction of Speed and Priority of Fire Two on D-SAT Performance

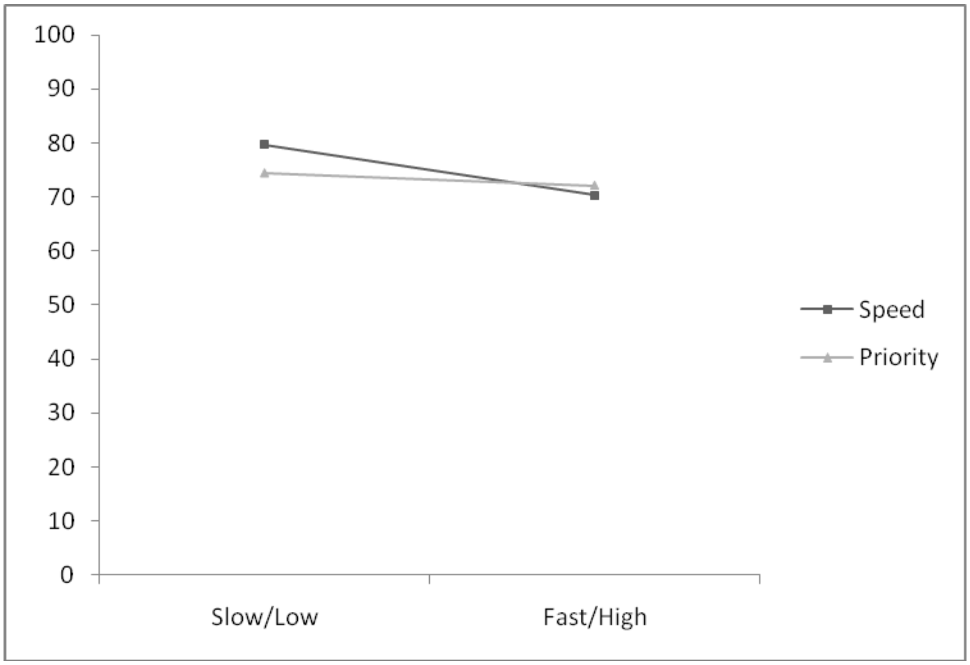


Figure 2.3 Pictorial Depiction of SDT

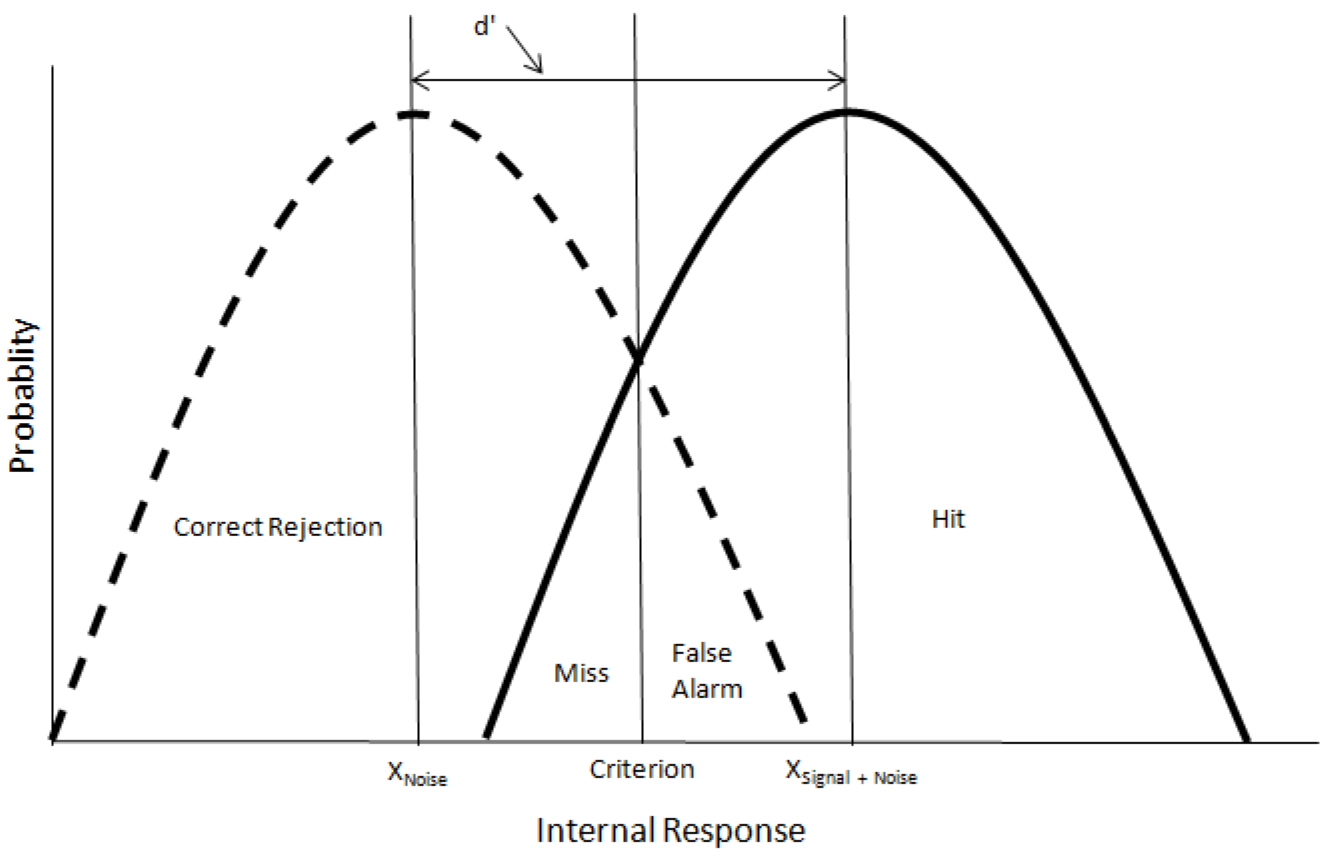


Table 2.1 Fire Characteristics for Each Scenario Type

Scenario	Speed of Fire 1	Priority of Fire 1	Speed of Fire 2	Priority of Fire 2
FLSL	Fast	Low	Slow	Low
FLFL	Fast	Low	Fast	Low
SHSL	Slow	High	Slow	Low
SHFL	Slow	High	Fast	Low
SHFH	Slow	High	Fast	High
FHFH	Fast	High	Fast	High
SLFH	Slow	Low	Fast	High
SLFL	Slow	Low	Fast	Low
SHSH	Slow	High	Slow	High

Table 2.2 Fire Characteristics for Each Scenario Type

	Fire 1		Fire 2	
Tested Study 2	Speed	Priority	Speed	Priority
Yes	Slow	Low	Fast	Low
Yes	Slow	Low	Fast	High
No	Slow	Low	Slow	Low
No	Slow	Low	Slow	High
Yes	Slow	High	Slow	Low
Yes	Slow	High	Slow	High
Yes	Slow	High	Fast	Low
Yes	Slow	High	Fast	High
Yes	Fast	Low	Fast	Low
No	Fast	Low	Fast	High
Yes	Fast	Low	Slow	Low
No	Fast	Low	Slow	High
No	Fast	High	Slow	Low
No	Fast	High	Slow	High
No	Fast	High	Fast	Low
Yes	Fast	High	Fast	High

Table 2.3 D-SAT Scenarios Order and Characteristics

Order	Scenario	Length (min)	Type	Switch	Question Number
1	1F	2.5	Filler	N/A	
2	FLFL	2.5	Non-SA	Possible	
3	SHSH	2.5	Non-SA	Possible	
4	3F	2.5	Filler	Yes	
5	FHFH	3	Non-SA	Possible	1
6	FLSL	2.5	Non-SA	No	2
7	FHFH	2.5	Non-SA	Possible	3
8	FLFL	3	Non-SA	Possible	
9	3F	2.5	Filler	Yes	
10	SHSH	3	Non-SA	Possible	
11	1F	2.5	Filler	N/A	4
12	SHSL	2.5	Non- SA	No	5
13	SLFL	2.5	Non-SA	Possible	
14	SHFL	3	Non-SA	No	6
15	2F	2.5	Filler	No	7
16	SLFH	2.5	SA	Yes	
17	SHFH	2.5	SA	Yes	
18	FLSL	3	Non-SA	No	8
19	2F	2.5	Filler	No	9
20	SLFH	3	SA	Yes	
21	SHSL	3	Non- SA	No	10
22	SHFH	3	SA	Yes	
23	SHFL	3	Non-SA	No	11
24	SLFL	3	Non-SA	Possible	12

Table 2.4 Means and Standard Deviations of D-SAT Performance in Percent Landscape Remaining for Speed and Priority by Fire and Level

Fire	Factor	Level	Mean	Std. Dev.
Fire 1				
	Speed			
		Slow	76.70	.39
		Fast	66.83	.88
	Priority			
		Low	71.19	.55
		High	75.14	.68
Fire 2				
	Speed			
		Slow	79.66	.78
		Fast	70.28	.39
	Priority			
		Low	74.42	.48
		High	72.11	.73

Table 2.5 SDT Scenario Classifications

		Need to Switch Fires	
		Yes	No
Switched Fires	Yes	Hit	False Alarm
	No	Miss	Correct Rejection

Table 2.6 Examples of Participant SDT Types with Actual Participant Data

a)

Switcher				
		Need to Switch Fires		
		Yes	Maybe	No
Switched	Yes	.90	.92	.92
Fires	No	.08	.08	.1

b)

Neutral				
		Need to Switch Fires		
		Yes	Maybe	No
Switched	Yes	.95	.50	.67
Fires	No	.33	.50	.05

c)

Non-Switcher				
		Need to Switch Fires		
		Yes	Maybe	No
Switched	Yes	.56	.17	.08
Fires	No	.92	.83	.44

CHAPTER 3 - General Discussion

D-SAT is a budding SA measurement technique; however, further adaptation and manipulations are required for D-SAT to reach its full potential. The results of the current study have the potential to add to the way SA is measured and allow for in-depth investigation of SA; by allowing research from inattentional blindness to be included in the understanding of SA. The two-fire paradigm used to measure SA in D-SAT requires the participant to overcome inattentional blindness. The inability to react to changes in the environment is a result of both inattentional blindness and low SA, requiring performance to be further investigated to determine the nature of poor performance. Unfortunately, the questions asked of participants were adapted from previous research, and did not allow for an independent investigation of inattentional blindness. All questions that probed inattentional blindness also probed SA. D-SAT can be a useful tool in the future, but there are issues that must be addressed prior to the widespread use of D-SAT as a SA measurement.

Limitations

While the theory behind D-SAT seems sound, the administration as tested in the present research has faults. The speed and priority manipulation used to measure SA did yield a significant effect on performance, indicating that D-SAT may have applications yet to be investigated and future research should delve into its possible implication and applications. However, D-SAT performance was not correlated with SAGAT-like question performance. The low number of correlations between scenario performance and performance on the SAGAT-like questions may be due to the low SAGAT-like question performance. Further research is required to determine what causes performance on the scenarios not to be correlated with SA question performance. One possibility may be the SAGAT's faults render the measure a poor comparison for D-SAT, requiring additional, different SA measures to be compared to D-SAT performance to further test D-SAT as an SA measure.

In addition to the few correlations found between D-SAT and SAGAT-like performance, the scores on the SAGAT-like questions were lower than expected. The low score may indicate the participants used for the study may not have had the expertise needed to measure SA, as the overall performance on the questions was consistent with the intended design of the study. However, participants tended to cluster towards the low end of the continuum instead of being normally distributed throughout. The skewed distribution can be seen by having a mean less than 50 percent correct and having the mean closer to the minimum than the maximum of the distribution, resulting in a floor effect for SA performance. In the future, participants may be given a longer practice time prior to completing the SA tests in order to develop a higher level of expertise on the task. As similar measures of SA have been used with experts to yield significant results in a study of nuclear power plant informational displays' effect on SA (Burns, et. al, 2008) suggesting the use of experts might also increase the scores on SAGAT-like questions.

Possible Implications and Future Research

Having an easy-to-administer test of SA, which could also be adopted to fit many domains, could also change training procedures following performance errors. For example, after a death during a wildfire the SA of the commander may come into question. If D-SAT is on hand, a quick measure of SA can be taken to determine if the death was a failure of the commander's SA or a misunderstanding of the situation.

NFC is a compact program that can be easily programmed to meet the needs of multiple disciplines. The versatility of NFC allows for the creation of D-SAT, which can be implemented to test SA in multiple contexts across a variety of situations. By allowing SA to be measured without interruption of the SA state, a better measure of SA can be made. The knowledge gained from the use of D-SAT could help establish SA as a sound research construct.

Future research will first focus on writing a computer program to extract additional data from the NFC outputs, allowing for an additional performance measure of SA. Unfortunately, the computer program could not be developed for the current study due to a lack of programming knowledge and resources of the researcher. The performance-based measure of SA will allow D-SAT to be used to determine the limits of SA, such as when failures occur and what is required for SA to occur. D-SAT will also be tested in an expert population to determine if the low D-SAT performance and SAGAT-like question performance were due to the low SA of the population used to test D-SAT. Additionally, D-SAT performance will be compared other measures of SA to determine if SAGAT or D-SAT led to the non-significant results between the measures.

References

Abdi, H. (2007). Signal detection theory (SDT). (Ed.). Salkind, N. Encyclopedia of measurement and statistics. (pp. 1-9). Thousand Oaks, CA: Sage.

Burns, C. M., Skraaning, G. A., Lau, N., Kwok, J., Welch, R., & Andresen, G., (2008). Evaluation of ecological interface design for nuclear process control: Situation awareness effects. *Human Factors, 50,* 663-679.

Durlach, P. J., Kring, J. P., & Browens, L. D., (2008). Detection of icon appearance and disappearance on a digital situation awareness display. *Military Psychology*, *20*, 81-94.

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

Endsley, M. R., (1996). Situation awareness measurement in test and evaluation. O'Brien, Thomas, G. & Charlton, Samuel, G. (Eds.). *Handbook of human factors testing and evaluation*. (pp. 159-180). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Endsley, M. R., (2000a). Theoretical underpinnings of situation awareness: A critical review. (Eds.). *Situation awareness analysis and measurement*. (pp. 3-32). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Endsley, M. R., (2000b). Direct measurement of situation awareness: Validity and use of SAGAT. (Eds.). *Situation awareness analysis and measurement*. (pp. 147-173). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Green, D. M. & Swets, J. A., (1966). *Signal detection theory and psychophysics.* Oxford, England: Wiley.

Jones, D. G., (2006). Subjective measures of situation awareness. (Eds.). *Situation awareness analysis and measurement*. (pp. 113-128). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Koustanai, A., Boloix, E., Van Elslande, P., & Bastien, C., (2008). Statistical analysis of "looked-but-failed-to-see" accidents: Highlighting the involvement of two distinct mechanisms. *Accident Analysis & Prevention, 40*, 461-469.

McCarley, J. S., Vais, M. J., Pringle, H., Kramer, A. F., Irwin, D. E., & Strayer, D. L., (2004). Conversation disrupts change detection in complex traffic scenes. *Human Factors, 46*, 424-436.

Mitroff, S. R., Simons, D. J., & Levin, D. T., (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics, 66*, 1268-1281.

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*, 303-316.

Sall, J. (2001). JMP IN (Version 4.0.4) [Computer Software]. Cary, NC: SAS Institute Inc.

Sarter, N. B. & Woods, D. D., (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors, 37*, 5-19.

Selcon, S. J. & Taylor, R. M., (1990). Evaluation of the situational awareness rating technique (SART) as a tool for aircrew systems design. In *Situational Awareness in Aerospace Operating* (AGARD-CP-478; pp. 5/1-5/8). Neuilly Sur Seine, France: NATO-AGARD.

Simons, D. J. & Levin, D. T. (1997). Change Blindness. *Trends in Cognitive Sciences, 1*, 261-267.

Strayer, D. L., Drews, F. A., & Johnson, W. A., (2003). Cell phone-induced failure of visual attention during simulated driving. *Journal of Experimental Psychology: Applied, 9,* 23-32.

Venturino, M., Hamilton, W. L., & Dvorchak, S.R. (1989). Performance-based measures of merit for tactical situation awareness. In *Situational Awareness in Aerospace Operations* (pp. 1-5). Copenhagen, Denmark: NATO - AGARD.

Vidulich, M. A. & Hughes, E. R., (1991). Testing a subjective metric of situation awareness. In *Proceeding of the Human Factors Society 35th Annual meeting,* 1307-1311. Santa Monica, CA: Human Factors Society.

Waag, W. L. & Houck, M. R., (1994). Tools for assessing situation awareness in an operational fighter environment. *Aviation, Space, and Environmental Medicine, 65*, A13-A19.

Woller, M. M., Park, J. A., Burrell, T. P., Hilgenkamp, H. A., Vowels, C. L., & Shanteau, J., (2008). The mismatch of perceptions of performance and actual performance in microworld simulations. Poster presentation, Association of Psychological Science Annual Conference, Chicago, IL.

Appendix A - SAGAT-like questions

The SAGAT-like questions below were piloted during study one and used to measure SA in study two. During the studies question were presented one at a time on 8.5” x 5.5” sheets of paper. The questions are in the order they were presented in study two, with the answers to each question highlighted. Additionally, each question is coded (in parenthesis) to indicate what concept was being probed by the question. The number of participants providing the correct answer is included (in parenthesis) following the correct answer for each question.

1. Please, rank the priority of the landscape items using a 1 (highest) to 5 (lowest) priority scale. (Understanding of instructions)

__3_(23) Tree

__2_(24) Cows/Horses

__4_(25) Grass

__1_(26) Houses

2. In order to save as much landscape as possible, what fire speed do you place a higher priority on? (Understanding of instructions)

__x_(26) Fast Moving

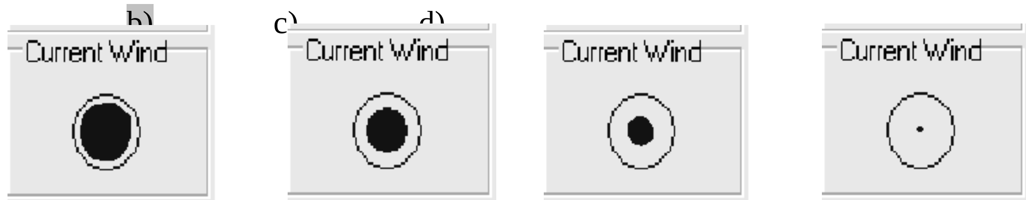
_____ Slow Moving

3. What appliance moves faster, the truck or the helicopter? (SA) ____Helicopter_(28)

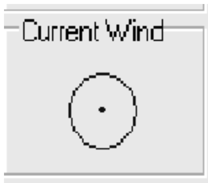
4. What appliance uses less water to fight fires, the truck or the helicopter? (SA) __Truck_(20)

5a. Which of the pictures below most closely matches the **speed** of the wind at the end of the trial? (Inattentional Blindness and SA)

a) b) c) d)



5b. On the picture, please mark the **direction** of the wind at the end of the trial. (Inattentional Blindness and SA; correct if response fell in southwest quadrant)



Southwest (12)

6. How many fire areas were there at the end of the last trial? (Inattentional Blindness and SA)

_____2_(22)

7. Please, outline the area destroyed by fire, in the last trial, on the map below. (Inattentional Blindness and SA) NOT SCORED

8a. What landscape element does not burn? (SA) ______Water/Pond/Lake_(16)

8b. How many were there on the map? (SA) _______4_(10)

9. What burns the fastest? (SA)

_______ Tree

__x (3) Cows/Horses

______ Grass

______ Houses

10. What burns the slowest? (SA)

______ Tree

______ Cows/Horses

__x_(3) Grass

______ Houses

11. How long are the trials? (SA; one point for either) _2.5 and 3 minutes_(3)

12. How many trials were there? (SA) ______24_(5)

Appendix B - Demographics Questionnaire

1. Age __________
2. Gender (1 = male, 2 = female) _______
3. Ethnicity _____________________
4. Year in school (1 = freshman, 2 = sophomore, etc.) _______
5. On an average weekday, how many hours do you play video/computer games _______
6. On an average weekend, how many hours do you play video/computer games _______
7. What is your favorite video/computer game to play? ___________________________

Please use the following scale to complete these questions:

1	2	3	4	5
Not at All				Extremely

1. How fun was the experiment you did today? ___________________
2. Overall, how difficult was it to do the entire experiment? ______________
3. Were the computer programs too difficult? _______________
4. Compared to others, do you think you performed better? ____________