Situation Awareness in Dynamic Decision Making: Effects of Practice and Working Memory

Cleotilde Gonzalez
Carnegie Mellon University
Jacob Wimisberg
Israeli Defense Forces

ABSTRACT: The goal of this research is (a) to investigate the effects of task practice on situation awareness (SA); (b) to investigate how cognitive ability – in particular, working memory – moderates this practice effect on SA; and (c) to investigate effects of the SA measurement procedure (covering or uncovering the display while queries are answered). Task practice and working memory influence SA. However, the dynamics of the relationship between working memory and SA over time are not well understood, particularly with regard to using different SA measures. This research reports an experiment in which different SA measurement methods were used while participants played a computer simulation over several days. SA was measured using two query methods: a covered method in which questions were asked while the display was blanked out, and an uncovered method in which questions were asked while the display was shown. A working memory measure was collected from participants. Results indicate that the relationship between working memory and SA diminishes with task practice and differs between a covered and uncovered method. We also found that SA improvement with practice depends on the way in which SA is measured. We discuss the implications of these findings for understanding SA development, SA measurement, and the relationship of working memory in this process. The results from this study have clear implications for systems design, for the design of learning aids, and for SA measurement.

Introduction

SITUATION AWARENESS (SA), OR PEOPLE'S UNDERSTANDING OF WHAT IS HAPPENING AROUND them and what could happen in the future, is a complex construct that plays a key role in the performance of dynamic tasks, such as flying an airplane, driving an automobile, and conducting military operations. Common definitions of SA include "the ability to extract and integrate information in a continuously changing environment and to use such information to direct future actions" (Wickens, 2000) and "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" (Endsley, 1988). Often, SA is divided into three levels: *perception,* the awareness of cues and information in the environment; *comprehension,* the retention, interpretation, and combination of perceptual information to provide meaning; and *projection,* the ability to anticipate future events (Endsley, 2000b).

Researchers have assumed that task practice influences SA; much of the literature derives optimal SA from interviews with subject matter experts (Endsley, 2001; Endsley, Bolte, & Jones, 2003; Shebilske, Goettl, & Garland, 2000). Experimental studies on the effects of task practice on SA often rely on the comparison between experts and novices (Durso et al., 1995; Sohn & Doane, 2004) rather than on the process by which novices improve their SA (Endsley, 2001; Shebilske et al., 2000). Task practice may enable people to develop "the mental models, schema, and goal-directed processing that are critical for SA in most domains" (Endsley et al., 2003), but unfortunately, little is known about how task practice contributes to SA development. We believe that an important first step in designing more effective displays and training interventions is to understand how individuals naturally improve their SA through practice; thereafter, methods to advance this skill may be designed accordingly.

Like task practice, cognitive mechanisms such as working memory are commonly assumed to influence SA (Endsley, 1995; Endsley & Robertson, 2000). Working memory is a limited human cognitive capacity that must be divided between temporarily holding information and strategically manipulating that information for future action (Baddeley, 1992). Although there is clear evidence regarding the influence of working memory on SA (Durso, Bleckley, & Dattel, 2006), there is only a weak understanding of the relationship between working memory and experience, which is that working memory plays a larger role in SA among novices than among experts (Ericsson & Kintsch, 1995; Ericsson & Staszewski, 1989; Sohn & Doane, 2004).

The role of working memory in SA is dependent on the way in which SA is measured (Durso et al., 2006). A common method to measure SA is query. In this method a task simulation is stopped at random points and a participant answers a set of questions about the situation. Queries may be answered while the simulation display is not visible (Endsley, 1995) or while the display is visible (Durso et al., 1995). Durso and colleagues (1995) found that adding SA to a battery of cognitive tests (including working memory tests) improved prediction of performance only when SA was measured with a query method in which information was visible. When SA was measured when information was not visible, SA did not improve the prediction of the cognitive tests. This suggests that the relationship between cognitive tests and SA depends on the way SA is measured.

The goals of this paper are threefold. The first goal is to investigate the effects of task practice on SA. The second goal is to investigate how cognitive ability – in particular, working memory – moderates this practice effect on SA. The third goal is to investigate effects of the SA measurement procedure (covering or uncovering the display while queries are answered).

Relevant to describing the potential importance of working memory to specific levels of SA during task training, we also aim to conduct exploratory analyses on measures of perception, comprehension, and projection. Differential effects for the three levels of SA may help explain and understand more thoroughly the process by which individuals improve SA with practice and how working memory is related to SA.

The Role of Working Memory on SA Development

Working memory is a bottleneck for the development of SA (Endsley & Robertson, 2000). For example, forgetting information reportedly leads to approximately 9% of SA errors (Jones & Endsley, 1996). Therefore, we expect a correlation between working memory and SA – specifically, increased working memory (higher scores on a working memory test) is expected to lead to better SA.

The relationship dynamics between working memory and SA over time (practice) is less obvious. Researchers have investigated the relationships between cognitive abilities and performance over task practice. Ackerman (1988) predicted that initial task performance demands general and content abilities, but the need for these abilities diminishes with task practice. However, the change in correlations between cognitive abilities and performance through practice depends on the complexity of the task. For example, the strength of the relationship between cognitive ability and task performance increases as the complexity of a task increases (Kyllonen, 1985).

Ackerman (1992) also found that for simple tasks, the ability-performance correlations may decrease with task practice, but for complex tasks these correlations may remain consistent across task practice. We have found consistent task performance and working memory correlations through practice in complex dynamic decision-making tasks like that used in the present study (Gonzalez, Thomas, & Vanyukov, 2005).

Current research on the correlation between ability and performance leads us to believe that given the complexity of dynamic tasks, the correlation between SA and working memory will remain constant through practice.

Hypothesis 1. Correlation between scores in a working memory test and SA scores in a complex decision task will remain consistent and not increase throughout task practice.

The Role of SA Measurement

Any study of the role of working memory on SA will be influenced by how SA is measured (Gugerty, 1997). There are multiple ways to measure SA, directly and indirectly, subjectively and objectively. Direct measures often involve asking participants to recall or recognize events from their experience (Gugerty, 1997). Indirect measures infer SA from the individual's performance. For example, SA has been measured as the time it takes decision makers to detect irregularities in an environment (Sarter & Woods, 1992).

Subjective measures rely on a user's self-assessment of SA (Jones, 2000), whereas objective measures often involve querying participants to recognize a situation and then comparing their views of the situation with reality (Endsley, 2000a). The Situation Awareness Global Assessment Technique (SAGAT), created and validated by Endsley (1987, 1990, 1995), is an example of a direct and objective measure of SA. We chose to use SAGAT in the current study because it uniquely combines objectivity and directedness, and it is a well-documented measure of SA.

With SAGAT, a simulation of a dynamic task is paused at random times and the

display is blanked out. Individuals are queried concerning the current state of the simulation (e.g., "What is the current simulation time?"). SAGAT does not require user self-assessment or any inferences of user behavior. It is seemingly unintrusive because of the short (usually less than 1 min) and random interruptions it employs (Bolstad & Endsley, 1990). Further, no performance differences have been found with pauses of up to 5 min (Endsley, 1995).

SAGAT's requirement of blanking the display while the operator answers the queries demands that individuals use *memory recall* to answer the SAGAT queries (Endsley, 2000a; Gugerty & Tirre, 2000). Conversely, a method in which one allows individuals to view a display while they answer SA queries may make the answers too obvious (Durso et al., 1995; Gugerty & Tirre, 2000). Thus, the effects of working memory on SA would depend on how SA is measured.

We planned an experiment using two versions of SAGAT employing both covered and uncovered displays, similar to the way Durso et al. (1995) investigated novice and expert chess players. In the covered display condition (the standard version of SAGAT), the task is paused and the display is blanked while participants answer queries on perception, comprehension, and projection. In the uncovered display condition, all the information remains in view while the participant answers the queries. Consequently, in addition to the empirical investigation of widely accepted assumptions of the experience and working memory relationships over SA, in this paper we also attempt to shed light on a methodological controversy of how different SAGAT versions (blanking the screen or not while answering queries) put demands on working memory.

Like Durso et al. (1997), we expect higher SA scores in the uncovered condition, because all the information an individual needs in order to answer the queries is within view when SA is measured. We expect that individuals may use query information to improve their SA with practice. Similarly, we expect lower SA scores and weaker effects of practice on SA for the covered condition, as exposure to queries with uncovered displays may contribute to practice.

Hypothesis 2. Any change in SA with practice will depend on the way SA is measured. Increases in SA attributable to practice in the covered condition will be less than increases in SA in the uncovered condition.

We also expect a stronger relationship between working memory and SA in a covered versus uncovered display because individuals must recall information from memory while answering the queries rather than using the information available on the screen.

Hypothesis 3. The correlation between working memory and SA will be greater in the covered than the uncovered condition.

Further, it seems that different effects of working memory in the covered and uncovered conditions would depend on the level of the SA queries. For example,

working memory will be essential to answer perception queries (Level 1) under a covered display because responses to perception queries depend particularly on the visibility of information (Endsley, 2000a). In contrast, working memory may be less important to answer the same perception queries under an uncovered display because there is no need to remember information (the display is visible).

In addition, the need for working memory in responding to queries on Levels 2 and 3 of SA (comprehension and projection) is less obvious. For example, it is possible that less working memory is necessary to answer projection rather than perception queries because correct responses to these projection queries may depend more on the individual's experience with the task (long-term memory) than on holding and manipulating current information (i.e., working memory). However, it is also possible that projection queries will demand working memory (in both the covered and uncovered displays) because of the need for mental simulation and pattern-matching required to respond to those queries (Endsley, 2000b).

In this paper, we tested these three hypotheses and explored the effects of working memory on the three levels of SA for the covered and uncovered conditions.

Methods

Water Purification Plant: A Dynamic Decision-Making Task

The research reported here involved a computer simulation of a dynamic decision-making task called *water purification plant* (WPP). This simulation has been used in many empirical investigations (Gonzalez, 2004, 2005; Gonzalez, Thomas et al., 2005). WPP is a resource allocation and scheduling task that is representative of many real-world tasks. As reported in previous research (Gonzalez, Vanyukov, & Martin, 2005), WPP features the characteristics of dynamic decision-making environments as well as a simplified interface that helps in both explaining the task and learning it more quickly.

In WPP, the decision maker plays the role of an operator whose goal is to distribute water to different locations before given deadlines (see Figure 1). The operator activates the pumps assigned to each tank, letting water flow into adjacent tanks in an effort to pump the water through several chains of tanks before the deadlines. Water flows only from left to right in the system. The system's 23 tanks are interconnected in a treelike structure. Each tank has two pumps, and only five pumps can be active at any given time (participants are told that electricity constraints prevent them from using more than five pumps concurrently). Consecutive tanks (e.g., Tank 3, Tank 15, and Tank 21) compose a chain and have the same deadline (in this case, 8:00 p.m.).

The simulation provides an indicator that tracks the number of pumps in use (shown in the upper left corner of Figure 1). Each of the pumps delivers water at the rate of 1 gallon every 2 min; thus, when two pumps are active within one tank, the delivery rate is 1 gallon/min.

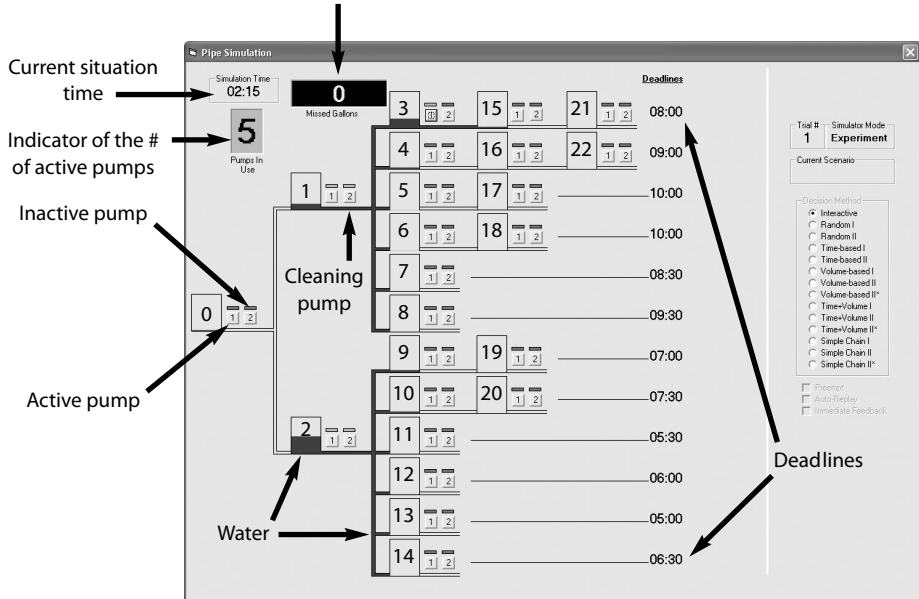Each pump may be operating in one of four states, identifiable by the color of

Figure 1. Layout of the WPP simulation. The simulation time is 2:15 p.m., the operator has missed 0 gallons of water, and five pumps – the maximum number of pumps – are active at this time. Water comes from outside the system and moves continuously from left to right through the active pumps and toward the deadlines. The operator decides when to activate and deactivate pumps while the simulation time is running. The numbers in the tanks have been added for reference in the text only; they do not appear during the simulation. For example, Tank 3 has water, whereas Tank 16 is empty.

the indicator bar above each pump: off (red), on (green), cleaning (yellow), or in queue (purple). Pumps switch to cleaning mode after they are turned off (either by the operator or by the system when there is no more water to pump in the tank), and each pump requires 10 min of simulation time to finish cleaning. While pumps are cleaning, the user can select and queue other pumps (within the 5-pump limit), which become active as soon as a cleaning pump turns red.

In WPP, exogenous events (i.e., the arrival of water at different times throughout the simulation) harm the status of the system if no action is taken (e.g., water will remain in the chains after the deadline expires). The user's actions, which are restricted by limited resources, also affect the status of the system.

WPP is a complex task because it involves multiple variables (e.g., pumps, water, and deadlines) and because some of the relationships among these variables are nonlinear. For example, pumping water out of a tank before the deadline reduces the number of gallons missed, but pumping water out of a tank after the deadline has expired is ineffective. Also, the more often a pump is opened, the more time it spends in cleaning mode and the less time it is available for pumping water through the chain.

Finally, WPP is "opaque" in the sense that many of the system's characteristics are not visually represented and therefore are identifiable only by user inference. For example, later deadlines have longer chains, causing more water to accumulate in the tanks, a situation that may affect the task prioritization suggested by the deadlines.

The main performance measure in WPP is the cumulative number of gallons of water remaining in the tanks after the expiration of all deadlines (shown in the upper left corner of the screen in Figure 1). The system has a total capacity per trial of 1,080 gallons of water.

Experimental Design

The design was a 2×18 mixed design. Participants were randomly assigned to one of two conditions (covered or uncovered display) and were asked to run the simulation 18 times (trials). The condition was used as a between-subjects variable in our analyses, and the trial was used as a within-subjects variable.

Individuals were asked to answer SA queries while the simulation was paused. The display was either blanked (for the covered screen condition) or visible (for the uncovered condition). Figures 2a and 2b present examples of queries in the covered and uncovered conditions. The difference between the covered and uncovered conditions was simply the availability of information in the display; there was no difference in WPP goals and activities.

All participants ran WPP 6 times per day (1 h per day) on 3 consecutive days, for a total of 18 trials. A preset scenario of water distribution dictated the arrival time, the amount of water dispersed, and the destination tank. For example, an entry in the WPP scenario may indicate that 10 gallons of water are to be assigned to Tank 2 at 2:02 p.m. In this experiment, the scenario was unknown to the participants and was kept constant during the 18 trials to accelerate learning.

Participants

Thirty-six volunteers – 18 females and 18 males ages 18–47 ($M = 25.1$, $SD = 7.0$) – responded to ads on electronic bulletin boards and in local newspapers. Each participant received $50 at the end of the experiment. Nineteen participants – 7 females and 12 males ages 18–47 ($M = 27.8$, $SD = 8.7$) – were assigned to the uncovered condition and 17 participants – 11 females and 6 males ages 19–27 ($M = 22.1$, $SD = 2.1$) – were assigned to the covered condition.

A Working Memory Measure

Modern measures of working memory evaluate an individual's ability for both storing and processing information in real time (Baddeley, 1992; Baddeley & Hitch, 1974). Two common measures analogous in terms of simultaneous processing and storage requirements are the Reading Span Test (Daneman & Carpenter, 1980) and the Visual Span Test, or VSPAN (Shah & Miyake, 1996). VSPAN measures the ability to process visual or spatial information; whereas Reading Span measures the ability to process linguistic information. Our research has tested the relationship and predictive ability of these two common tests of working memory on the performance
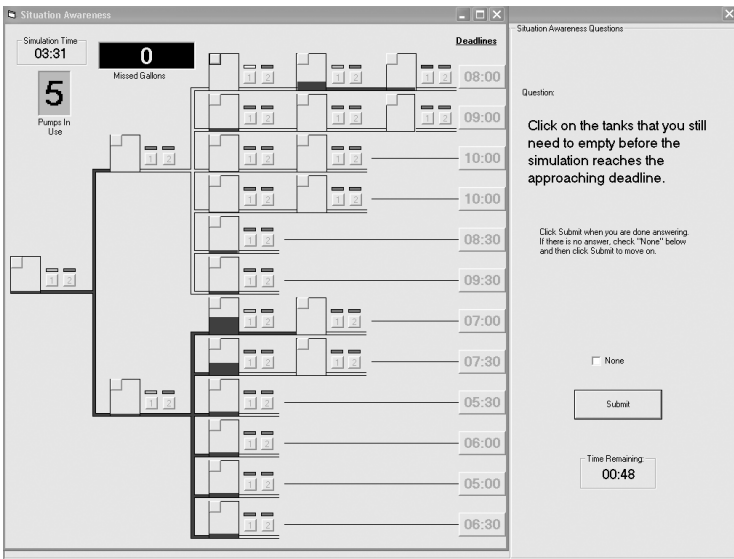
Figure 2a. Layout of the uncovered SAGAT condition. The time on the pictured screenshot is 3:31 p.m., and the operator has not yet missed any gallons of water. Thirteen tanks have water in them; five pumps are active. The SA queries appear on the right side of the screen. The operator has 48 s remaining to answer the series of queries.
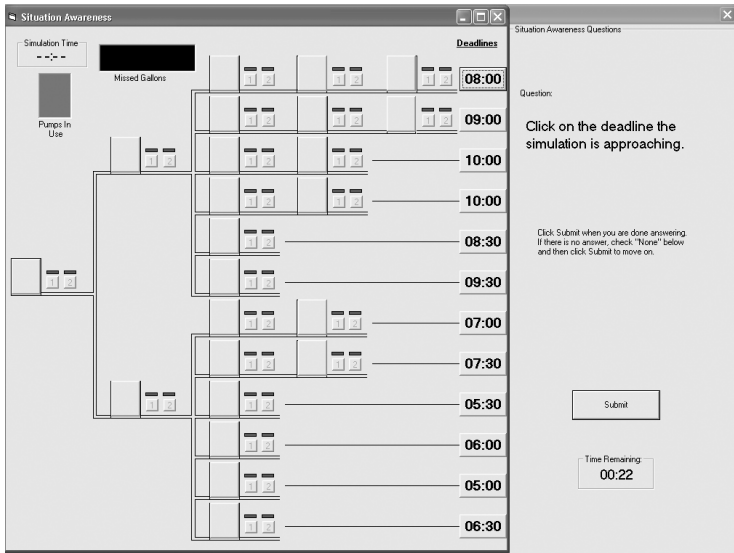

Figure 2b. Layout of the WPP-SAGAT simulation in the covered condition. All information relating to the task – time, status of the pumps, gallons of water missed, deadlines, etc. – is covered. The SA queries appear on the right-hand side of the screen. The operator has 22 s remaining to answer the series of queries.

of WPP, and we have found that VSPAN but not the Reading Span Test correlates significantly to performance in WPP (Gonzalez, Thomas et al., 2005). In addition, VSPAN reflects the processes of perception, encoding, and mental manipulation of spatial forms (Shah & Miyake, 1996) that are required for WPP. Thus, all participants in this experiment completed the VSPAN test.

VSPAN presents one letter at a time on the screen (a total of 70 letters); the letter is rotated clockwise and reversed. Participants are asked to determine if the letter is reversed or not while simultaneously remembering the direction in which the top of the letter points. At the end of the set of consecutive letters, individuals must recall the orientation of all letters in the set. The score is the total number of correctly recalled letters. Users required approximately 15 min to complete the VSPAN test (including instructions).

A Situation Awareness Measure

We used a query method similar to SAGAT to measure SA. WPP was paused at random times. The visual scene was blanked (for the covered condition only), and participants answered as many queries as possible from a randomly selected set of queries within a limited time. Queries addressed the participant's perception, comprehension, or projection given the current state of the simulation.

The usefulness of SAGAT depends on how well the queries reflect the most important aspects of awareness. Endsley (2000b) suggested that a task analysis would help determine the task goals, decisions, and information or resources needed to make those decisions. Both a cognitive task analysis (Gonzalez, Lerch, & Lebiere, 2003) and verbal reports (Gonzalez, 2003) helped to identify the key requirements for good performance in WPP. Queries were developed, pretested, and revised with the help of experts in this area and in accordance with data gathered from five participants in a pilot study (see Appendix for the complete list of queries developed for WPP). The queries were an integral part of the WPP task. That is, the answers to the queries, as well as the state of the environment at the time when queries were posed, were saved in a database.

Procedure

On the first day, participants completed the VSPAN and then a WPP training session. Individuals interacted with WPP at a very slow pace, and the experimenter answered questions about the task. Following the WPP training session, participants received instructions on how to answer the SA queries. They received several training trials with the SA queries.

After the training and during the subsequent two days, individuals played 18 trials of WPP. They were instructed to attend to the task first and consider the queries secondary (Endsley, 2000a). We asked participants to answer as many queries as possible during two pauses of 2 min each in randomly selected intervals per trial. Queries were presented in random order. After the 2 min allotted to answer the queries had passed, the simulation resumed.

The WPP trial with queries had a total duration of 12 min. Participants did not

rest between the trials and completed six consecutive trials per day over three consecutive days.

Data Preparation and Analysis

The main dependent variable was the average percentage correct for each query (SA score). (We also collected the response time to correctly answered queries. Such a measure is particularly useful in uncovered screen conditions and in simple tasks, in which there may be a ceiling effect if participants can determine the correct answers directly from information displayed. As will be clear in the results, a ceiling effect was not an issue in this task; thus an accuracy measure similar to the one used here is appropriate for both covered and uncovered, experimental conditions.) We calculated an average percentage correct for each trial. To explore SA for each of the perception, comprehension, and projection levels, we also generated three separate averages of the percentage of correct responses. We averaged the scores of the queries corresponding to the classification of each level (see Appendix).

For each query, the simulation program stored the status of the environment at the time the query was presented. Participants were not informed of the scoring system and were not given feedback on their responses to the queries.

Two statistical models were used in the analysis of these data: a repeated-measures general linear model (GLM), which used the trial as the repeated variable and the experimental condition (covered or uncovered) as a between-subjects factor, and a regression model using VSPAN as a predictor of SA.

Results

To investigate the change in the correlation between working memory and SA scores over practice (Hypothesis 1), we calculated the correlation between each average SA score per trial and VSPAN. Figure 3 depicts these correlations per trial. Then we used the slope of the regression line to identify the linear trend of these correlations. This is a common analytical technique in skill acquisition research (Ackerman, 1992). The results of this analysis indicate that the correlation between SA scores and VSPAN decreased over time (adj. $R^2 = .35$, $t(17) = -2.92$, $p < .01$). These results indicate that initially, SA required working memory, but with task practice, the influence of working memory on SA diminished. Thus, Hypothesis 1 was not refuted.

Figure 4 depicts the SAGAT scores per condition and trial. On average, the SA scores were higher in the uncovered condition ($M = 57\%$, $SE = 3\%$) than in the covered condition ($M = 39\%$, $SE = 2\%$). To test Hypothesis 2, we ran a GLM repeated-measures analysis to investigate SA differences that were attributable to the condition and practice among participants. The results showed a significant main effect of condition, $F(1, 33) = 19.76$, $p < .001$; no effect of trial, $F(17, 561) = 1.5$, ns; and a significant interaction between condition and trial, $F(17, 561) = 1.99$, $p < .05$. This interaction suggests a differential improvement in the SA scores that is dependent on the condition, in part confirming the second hypothesis. Post hoc tests indicated
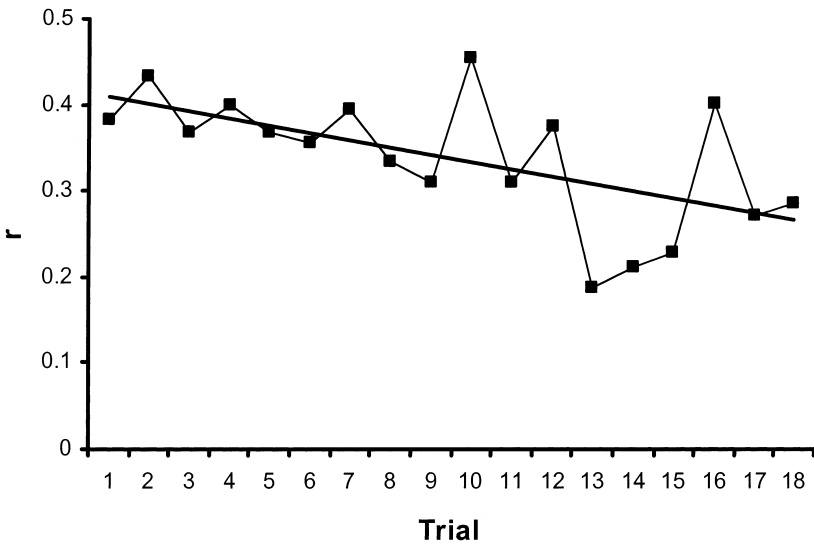
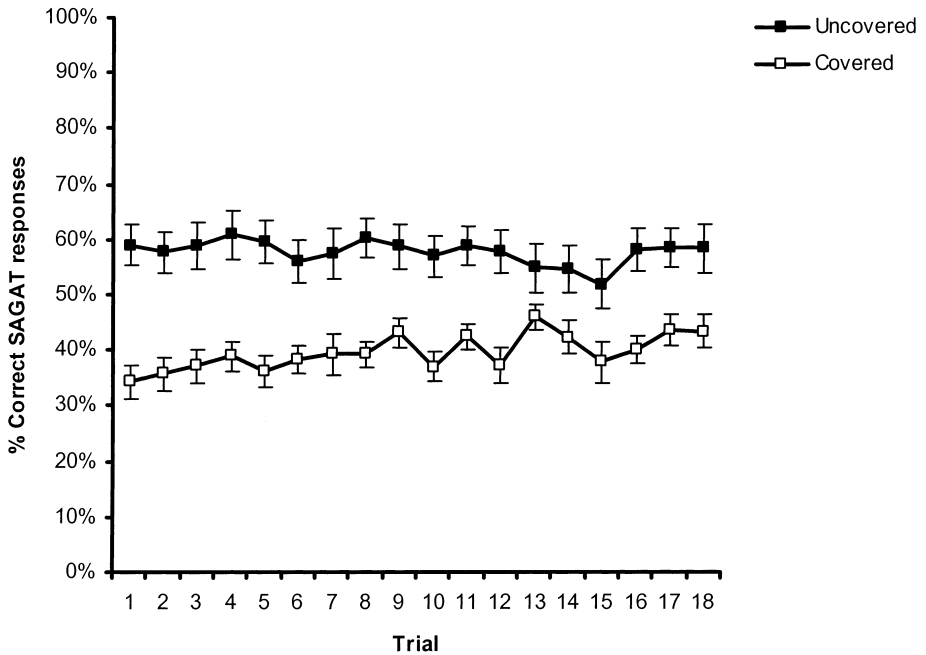*Figure 3. Correlation between VSPAN and SA scores per trial. There is a significant decreasing trend.*

Figure 4. Percentage of correct SAGAT responses per condition and for each of the 18 trials.

that the effect of practice was significant only in the covered condition, $F(17, 272) = 2.63$, $p < .001$, but not in the uncovered condition, $F(17, 289) = 1.01$, ns. This was counter to expectation, as greater improvements in SA were expected across trials with uncovered measurement.

To test Hypothesis 3 about whether working memory correlates with SA and how this relationship differs per condition, we used regression analyses. In the first regression, we used the average of SA scores as a function of condition, VSPAN, and the interaction between VSPAN and condition. The results indicated that only the condition was a predictor of the SA scores (adj $R^2 = .413$, std. $\beta_1 = .779$, $p < .05$). VSPAN was not a predictor of the SA scores (std. $\beta_1 = .620$, ns), and, contrary to our expectations, the effect of condition was not moderated by working memory (std. $\beta_1 = -.442$, ns).

However, a multicollinearity analysis also indicated a strong association between the interaction predictor and the other predictors (VIF = 19.132, Tolerance = .052). Therefore, we ran a second regression equation without the interaction factor. The results indicated that both condition and VSPAN were predictors of the SA scores (adj. $R^2 = .455$, std. $\beta_{1\ (Condition)} = .559$, $p < .001$, $\beta_{2\ (VSPAN)} = .288$, $p < .05$).

To further investigate if VSPAN predicted SA differently in the covered and uncovered conditions, we ran two separate regression equations, one for each condition, using VSPAN as the predictor of the SA scores. The results indicated that VSPAN predicted the SA scores only in the covered condition (adj. $R^2 = .455$, std. $\beta_1 = .656$, $p < .01$) and not in the uncovered condition (adj. $R^2 = -.010$, std. $\beta_1 = .222$, ns). These results supported Hypothesis 3.

Exploratory Analyses for SA Levels

We expected to obtain more information about the effects of the SA measurement and working memory through the exploration of SA scores for the perception, comprehension, and projection levels of SA. Figure 5 shows the average SA score per trial separated for each of the levels of SA.

SA scores for perception and comprehension queries were higher in the uncovered condition than in the covered condition (uncovered condition: perception: $M = 75\%$, $SE = 4\%$, comprehension: $M = 43\%$, $SE = 3\%$; covered condition: perception: $M = 42\%$, $SE = 2\%$, comprehension: $M = 33\%$, $SE = 3\%$). However, SA scores for projection queries were similar in both conditions (uncovered condition: $M = 46\%$, $SE = 3\%$; covered condition: $M = 44\%$, $SE = 2\%$).

We ran regression analyses for each level of SA using VSPAN as the predictor of SA scores. The results indicated that the relationship between VSPAN and the SA scores in the perception level was only marginally significant (adj. $R^2 = .078$, std. $\beta_1 = .324$, $p < .10$), and this relationship was not significant for the projection queries (adj. $R^2 = .009$, std. $\beta_1 = .196$, ns).

Discussion

The impact of practice on SA and the moderating effect of working memory

depend on how SA is measured. SA improves with practice when measured in a covered display condition but not when measured in the uncovered condition. Furthermore, the effect of experience appears only in the covered condition for perception and comprehension queries, not for projection queries. Several explanations for these results are possible.

Although individuals started with lower SA in the covered condition than did those in the uncovered condition, they improved their answers to SA queries over time, perhaps because as they acquired knowledge of the task, they used this knowledge to answer queries more accurately. It is possible that in the uncovered condition, participants felt no need to use their acquired task knowledge in order to answer the queries because the information was present on the screen. Uncovering information may have been a disincentive for using knowledge acquired in the task; individuals knew they could use the information available rather than making an extra effort to retrieve knowledge to answer the queries.

Interestingly, task knowledge acquired through practice did not make a difference in answers to projection queries, regardless of the condition. It seems that in order to improve the accuracy of their predictions, individuals need more than the presence of perceptual information on the screen. More important, to improve their predictions, individuals need more than the knowledge created through task practice. It is possible that with practice in the task, individuals simply exerted implicit learning (Berry & Broadbent, 1987). That is, they accumulated knowledge and perhaps were able to improve their pattern-matching abilities, but, at the same time, they were not able to compile this knowledge in the form of rules that could explain how the system works. Although we did not ask participants to describe or verbalize their knowledge of the task, there is evidence that in dynamic complex tasks such as WPP, individuals acquire implicit knowledge that it is hard to express in the form of rules (Berry & Broadbent, 1987, Dienes & Fahey, 1995).

Another main conclusion from this empirical study is that the moderating effect of working memory on SA changes with task practice and depends on the conditions in which SA is measured. Essentially, working memory predicted SA scores only in the covered condition. This result confirmed our expectations. SAGAT's requirement of blanking the screen while an operator answers queries elicits working memory.

We also found that the effect of working memory was more relevant for the perception than for the projection level of SA. Because of the nature of perception queries, correct answers fully depend on the visibility of the elements on the display. Conversely, accurate answers to projection queries depend on a more profound understanding of the task.

Finally, we found a significant decrease in the correlation between SA and VSPAN with practice in the task. This result suggests that as experience accumulates, there is a decreasing need for using working memory to maintain SA, although working memory may still be necessary for task performance. This result supports the contention that experienced decision makers develop ways to overcome their working memory limitations with practice (Ericsson & Kintsch, 1995; Ericsson & Staszewski, 1989). Perhaps in this case, practice in the task familiarized individuals with the
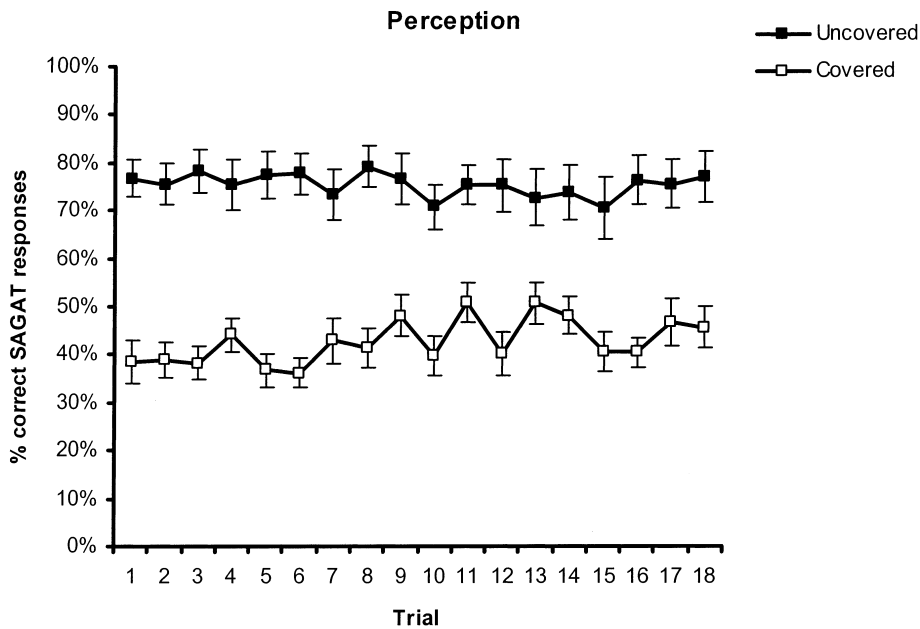
Perception

Figure 5a. The proportion of correct SAGAT responses per condition and trial for the SA level perception.
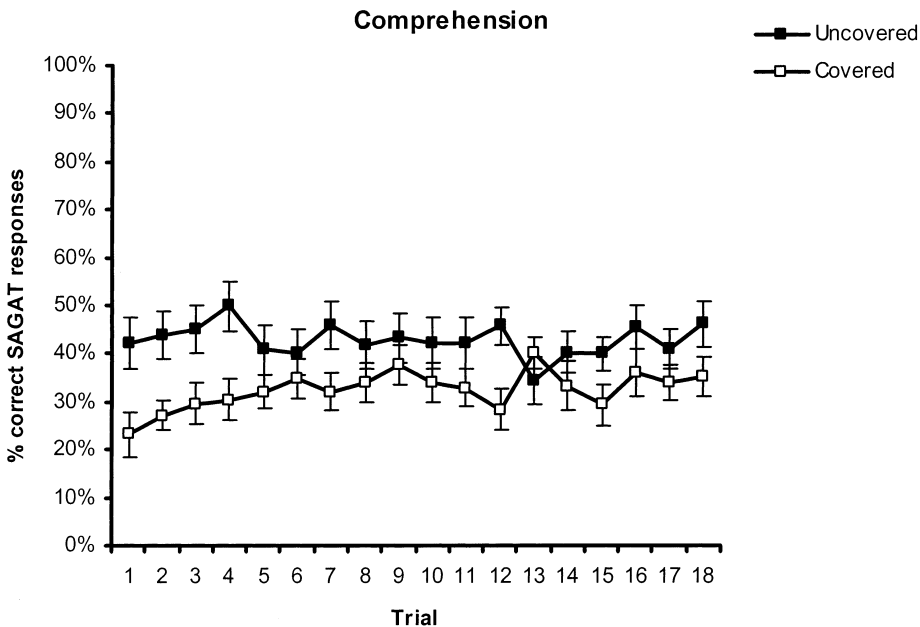
Comprehension

Figure 5b. The proportion of correct SAGAT responses per condition and trial for the SA level comprehension.
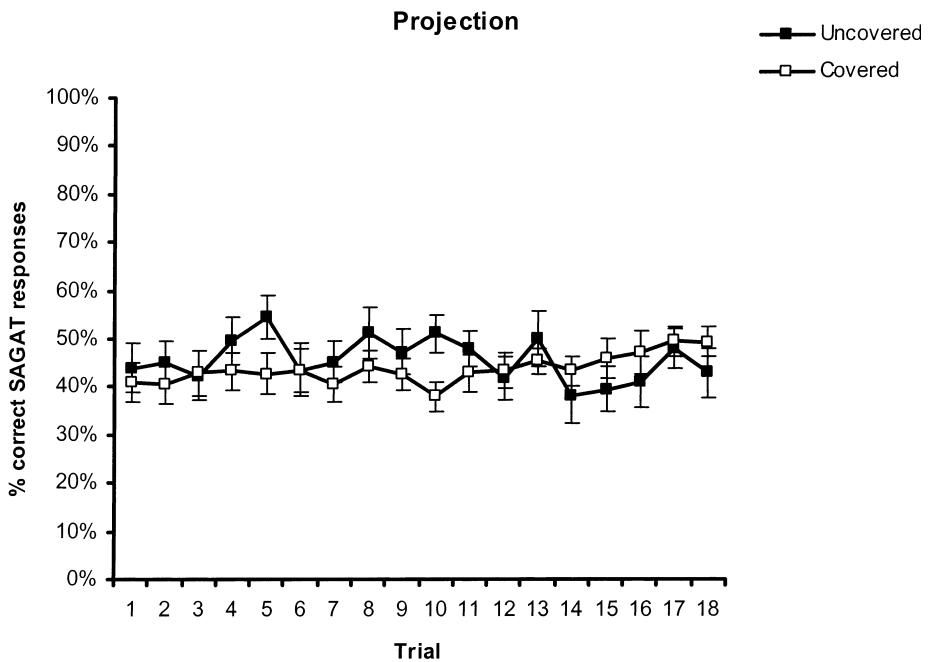
Figure 5c. The proportion of correct SAGAT responses per condition and trial for the SA level projection.

location of the items on the screen and generated knowledge that was easily retrievable, thereby circumventing any working memory limitations. In addition, queries may have drawn participants' attention to the elements they should be considering, inadvertently making them aware of important aspects of the task and reducing working memory demands with practice.

Conclusion

The results from this study have clear implications for systems design, for the design of learning aids, and for SA measurement. It is often assumed that SA is derived simply from people's senses or by direct observation from the world. If this were the case, systems designers would need only to make all the information available on the screen and SA would be best. Although data from this study indicate that availability of information at the interface resulted in better SA, the results also suggest that availability of information does not guarantee that SA will improve with practice.

Training efforts and learning experiments need to address the question of knowledge acquisition, particularly what is generated with practice in the task and what is generated by providing strategic information. For example, SA feedback interventions can help one understand how SA improves with practice and how this improvement could help in attaining a general understanding of the task. According

to our results, individuals need procedural knowledge or strategic information about how to address the task rather than simply task practice. More research is necessary to investigate how experience generates knowledge of the task – specifically, how task experience generates knowledge that can be used to improve our predictions. Clearly, a follow-up on this research should involve the generation of two forms of learning, implicit and explicit, to determine their differential effect on the improvement of SA.

Acknowledgments

APPENDIX: SA queries and their classification into three levels of SA

Query No.	Query Type	Query
1	Perception	Click on all the tanks with water in them.
2	Perception	Click on the active (green) pumps.
3	Perception	Click on the pumps that are in queue.
4	Perception	Click on the pumps that are cleaning.
5	Perception	Click on the deadline the simulation is approaching.
6	Perception	What is your current simulation time?
7	Perception	How many gallons of water have you missed up to now?
8	Perception	Click on the tanks that you still need to empty before the simulation reaches the approaching deadline.
9	Comprehension	How long is left before the simulation reaches the approaching deadline?
10	Comprehension	Click on the three tanks with the most water in them.
11	Comprehension	Find the chain with the most water and click on all the tanks in that chain.
12	Comprehension	Click on the tanks that should be emptied at every deadline.
13	Comprehension	Click on the tanks that are not important anymore.
14	Projection	Which would take longer: the water in tank 1 to reach 8:00 deadline or the water in tank 2 to reach 7:00?
15	Projection	Click on the tanks that you would expect to be filled with water from outside the system in the next hour.
16	Projection	By how much do you expect your missed gallons score to increase at the current deadline?
17	Projection	How many total gallons of water will you miss in this trial?
18	Projection	Do you predict that you will be able to achieve the simulation goal of zero gallons missed?

References

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

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinant. *Journal of Applied Psychology, 77*, 598–614.

Baddeley, A. D. (1992). Working memory: The interface between memory and cognition. *Journal of Cognitive Neuroscience, 4*, 281–288.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47–90). New York: Academic Press.

Berry, D. C., & Broadbent, D. E. (1987). The combination of explicit and implicit learning processes in task control. *Psychological Research, 49*, 7–15.

Bolstad, C. A., & Endsley, M. R. (1990). *Single versus dual scale range display investigation* (NOR-DOC 90-99). Hawthorne, CA: Northrop Corp.

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

Dienes, Z., & Fahey, R. (1995). Role of specific instances in controlling a dynamic system. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 848–862.

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

Durso, F. T., Hackworth, C. A., Truitt, T. R., Crutchfield, J. M., Nikolic, D., & Manning, C. A. (1997). Situation awareness as a predictor of performance of en route air traffic controllers. *Air Traffic Control Quarterly, 6*(1), 1–20.

Durso, F. T., Truitt, T. R., Hackworth, C. A., Crutchfield, J. M., Nikolic, D., Moertl, P. M., Ohrt, D. D., & Manning, C. A. (1995). Expertise and chess: A pilot study comparing situation awareness methodologies. In D. J. Garland & M. R. Endsley (Eds.), *Experimental analysis and measurement of situation awareness* (pp. 295–303). Daytona Beach, FL: Embry-Riddle Aeronautical University Press.

Endsley, M. R. (1987). *SAGAT: A methodology for the measurement of situation awareness* (No. NOR DOC 87-83). Hawthorne, CA: Northrop Corp.

Endsley, M. R. (1988). *Design and evaluation for situation awareness enhancement.* Paper presented at the Human Factors and Ergonomics Society 32nd Annual Meeting, Anaheim, CA.

Endsley, M. R. (1990). A methodology for the objective measurement of situation awareness. In *Situation awareness in aerospace operations* (AGARD-CP-478) (pp. 1/1-1/9). Neuilly Sur Siene, France; NATO-AGARD.

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

Endsley, M. R. (2000a). 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: Erlbaum.

Endsley, M. R. (2000b). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 3–32). Mahwah, NJ: Erlbaum.

Endsley, M. R. (2001). *Designing for situation awareness in complex systems.* Paper presented at the Second International Workshop on Symbiosis of Humans, Artifacts and Environment, Kyoto, Japan.

Endsley, M. R., Bolte, B., & Jones, D. G. (2003). *Designing for situation awareness: An approach to user-centered design.* London: Taylor & Francis.

Endsley, M. R., & Robertson, M. M. (2000). Training for situation awareness in individuals and teams. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness and measurement.* (pp. 349–366), Mahwah, NJ: Erlbaum.

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

Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert A. Simon* (pp. 235–267). Mahwah, NJ: Erlbaum.

Gonzalez, C. (2003). Verbal protocols in real-time dynamic decision making. *Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting* (pp. 293–296). Santa Monica, CA: Human Factors and Ergonomics Society.

Gonzalez, C. (2004). Learning to make decisions in dynamic environments: Effects of time constraints and cognitive abilities. *Human Factors, 46,* 449–460.

Gonzalez, C. (2005). The relationship between task workload and cognitive abilities in dynamic decision making. *Human Factors, 47,* 92–101.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science, 27,* 591–635.

Gonzalez, C., Thomas, R. P., & Vanyukov, P. (2005). Impact of individual differences and cognitive abilities on dynamic decision making. *Intelligence, 33,* 169–186.

Gonzalez, C., Vanyukov, P., & Martin, M. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior, 21,* 273–286.

Gugerty, L. J. (1997). Situation awareness during driving: Explicity and implicit knowledge in dynamic spatial memory. *Journal of Experimental Psychology: Applied, 3*(1), 42–66.

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

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: Erlbaum.

Jones, D. G., & Endsley, M. R. (1996). Sources of situation awareness errors in aviation. *Aviation, Space and Environmental Medicine, 67,* 507–512.

Kyllonen, P. C. (1985). *Dimensions of information speed* (No. AFHRL-TP-8-56). Brooks Air Force Base, TX: Air Force Systems Command.

Sarter, N. B., & Woods, D. D. (1992). Pilot interaction with cockpit automation: Operational experiences with the flight management system. *International Journal of Aviation Psychology, 2,* 303–321.

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,* 4–27.

Shebilske, W. L., Goettl, B. P., & Garland, D. J. (2000). Situation awareness, automaticity, and training. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 303–323). Mahwah, NJ: Erlbaum.

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,* 461–475.

Wickens, C. D. (2000). The trade-off of design for routine and unexpected performance: Implications of situation awareness. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 211–225). Mahwah, NJ: Erlbaum.

Cleotilde Gonzalez is an assistant professor in the Social and Decision Sciences Department at Carnegie Mellon University, 5000 Forbes Ave., Porter Hall 223C, Pittsburgh, PA 15213, conzalez@andrew.cmu.edu. She is also director of the Dynamic Decision Making Laboratory (*www.cmu.edu/ddmlab*). She was a Postdoctoral Fellow, focusing on cognitive workload in real-time dynamic decision making, and an associate professor of information systems at Universidad de las Américas Puebla, México in the Department of Computer Engineering.

Jacob Wimisberg received his B.A. (Magna Cum Laude) in behavioral sciences from the Academic College of Tel-Aviv and continued to study social psychology at Tel Aviv University, Israel. He received his master's degree in 2003 in the Social and Decision Sciences Department at Carnegie Mellon University. He is working in decision-making planning for the Israeli Defense Forces and planning to continue his Ph.D. degree.