Measuring and Predicting Shared Situation Awareness in Teams

Lelyn D. Saner
University of Maryland
Cheryl A. Bolstad
SA Technologies
Cleotilde Gonzalez
Carnegie Mellon University
Haydee M. Cuevas
SA Technologies

Abstract: In order to improve our understanding of situation awareness (SA) in teams performing in technologically advanced command, control, and communications (C3) operations, researchers need to develop valid approaches to assess both individual and shared SA. We investigated SA in an interdisciplinary military rescue operation training exercise. For this study, we developed procedures to measure the degree of shared SA between two team members and to improve the accuracy of their shared SA scores. We suggest that SA scores that are calculated using many existing methods may be *inflated* because they often fail to account for error in terms of both the amount of information that is thought to be relevant and in the accuracy of a person's knowledge of it. We calculated *true* SA scores that account for both of these types of error. The measures were then used to evaluate five potential predictors of shared SA. Our analysis suggested that failure to compensate for error in SA may lead to overestimation of performance in a situation. The results also revealed a significant relationship between shared SA and participants' distance from a central, joint service team, which acted as the organizational hub within the C3 structure. Shared SA was better the further away from the hub people were, which suggests that a person's role and position within an organization affects the level of shared SA that can be achieved with other individuals.

Introduction

Complex operational environments such as those found in the military require high levels of individual situation awareness (SA) but also a higher level of coordinated awareness among the team members (Gorman, Cooke, & Winner, 2006). In other words, individuals performing as teams in these contexts need to develop an accurate common understanding of the situation. Although much debate still exists regarding how best to define SA (Endsley, 2000a; Rousseau, Tremblay, & Breton, 2004), one commonly accepted definition is "the perception of the elements in the

---

ADDRESS CORRESPONDENCE TO: Lelyn D. Saner, University of Maryland, Center for Advanced Study of Language, 7005 52nd Ave., College Park, MD 20742; ldsaner@gmail.com.

Journal of Cognitive Engineering and Decision Making, Volume 3, Number 3, Fall 2009, pp. 280–308.
DOI 10.1518/155534309X474497

environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley, 1995a, p. 36). Based on this definition, SA can be understood as activated knowledge about a situation in which one is currently involved. A contrasting definition is that SA is a "continuous perception-action process in which ongoing activity plays an integral role in what there is to be perceived" (Gorman et al., 2006, p. 1314). By this definition, SA is conceptualized as a series of responses to constantly changing environmental conditions. Regardless of which definition is adopted, as both a state of knowledge and an appropriate action, SA is especially crucial in domains in which the effects of ever-increasing technological and situational complexity on the human decision maker are a concern.

Many complex tasks rely primarily on teams of individuals to achieve mission objectives, such that the concepts of team SA and shared SA are equally important (Gorman et al., 2006; Salas, Prince, Baker, & Shrestha, 1995). *Team SA* can be defined as "coordinated perceptions and coordinated actions" (Gorman et al., 2006, p. 1314) or as "the degree to which every team member possesses the SA required for his or her responsibilities" (Endsley, 1995a, p. 39). Thus, the success or failure of a team depends on the success or failure of each of its team members. In contrast, *shared SA* is defined as "the degree to which team members possess the same SA on shared SA requirements" (Endsley & Jones, 2001, p. 48).

As implied by these definitions, the information requirements relevant to each individual team member and the overlap of these requirements among the team members are both essential elements of SA in a team (Salas et al., 1995). However, despite the considerable research done in team performance (Brannick, Prince, Prince & Salas, 1995; Cooke, Gorman, Duran, & Taylor, 2007; Salas et al., 1995), valid and reliable measures of team and shared SA are still lacking (Salas, Cooke, & Rosen, 2008).

In this paper, we examine issues surrounding the measurement and prediction of shared SA in teams. In particular, we argue that both the degree of *accuracy* of an individual's SA and the *similarity* of two individuals' SA are integral to the attribution of shared SA between two team members, and we consider the role that each component plays in the measurement process. We rely on a measurement approach developed for individual SA and extend it to measure shared SA in a complex military team consisting of heterogeneous players coordinating in a dynamic, simulated military rescue operation training exercise. Next, we expand on the accuracy and similarity issues identified earlier, which serve as the basis for the development of the measurement approaches used in this study. That is followed by a discussion of potential predictors of SA. We then present the detailed method and results for this study and conclude with implications for future research in this area.

Accuracy and Similarity of Situation Awareness

The *accuracy* of an individual's understanding of a situation is always limited (Adams, Tenney, & Pew, 1995). The actual SA that is measured or observed is always less than or equal to the achievable SA (Pew, 2000). In Endsley's (1995a) definition of SA, the components of perceiving elements (Level 1), comprehending their meaning (Level 2), and projecting their future values (Level 3) represent progressively higher levels of SA, each of which builds upon the one below. When an error occurs at one level of SA, it may contribute to errors at other levels (Endsley & Garland, 2000a; Endsley & Robertson, 2000; Jones & Endsley, 1996). For example, a Level 1 SA error, such as "failure to monitor or observe data" can lead to a "lack of or incomplete mental model," which is a Level 2 or Level 3 SA error (Jones & Endsley, 1996, p. 508). In this study, our analyses of individual and shared SA provide a concrete example of how a person's mental model may be either incomplete or inflated when information is not properly monitored.

Shared SA is dependent on the SA of the individuals involved (Salas et al., 1995), but it is more complicated, in that often not all team members in a given situation need to be aware of all the same information (Endsley, 1995a; Jones & Endsley, 2002). In many situations, individuals in a team possess specialized knowledge to help them perform particular tasks, and they rely on others to do their jobs properly. Although each team member needs to monitor and have good SA of the information that is relevant to his or her job, the overlap of the individual SA between two team members is important only when they have the same task requirements. Shared SA can be evaluated by directly comparing any two given team members with regard to the similarity of their understanding of the situation elements that are relevant to both of them. Jones and Endsley (2002) identified five possible shared SA outcomes when two team members' SA is compared:

- Both team members answer correctly – indicates accurate shared SA
- Team Member A answers correctly, but Team Member B answers incorrectly – indicates nonshared SA
- Team Member B answers correctly, but Team Member A answers incorrectly – indicates nonshared SA
- Neither team member answers correctly, and their wrong answers are different – indicates team members have different SA and neither is correct
- Neither team member answers correctly, and their wrong answers are essentially the same – indicates inaccurate but shared SA

To illustrate the relationships between accuracy and similarity in shared SA, consider Figure 1, in which each of three understandings of a situation is shown in a separate box. In the first box, X1 to X7 represent all the facts that need to be known by someone in order to have complete awareness of a situation. Because the absolute truth is seldom (if ever) known in the real world, an individual's SA is often evaluated practically with respect to what is called *ground truth*. Ground truth is determined, for example, by an instructor who is controlling a simulated exercise and who can measure the relative performance of those participating in the simulation.

When we compare the set of facts reported by Person 1 to ground truth, we see that X2 and X5 are missing from the story and that two additional facts (X8

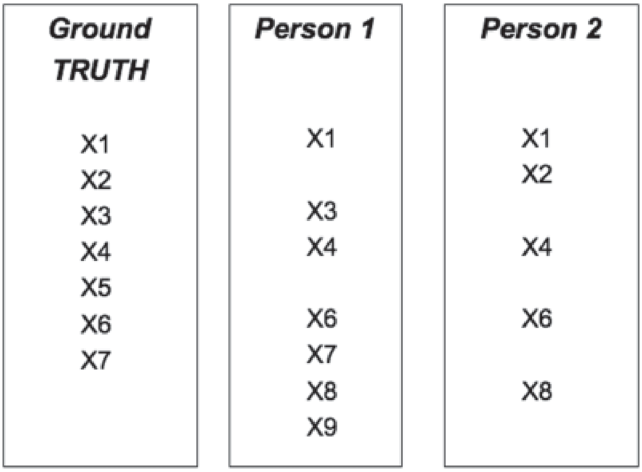

Figure 1. Accuracy and similarity of situation awareness.

and X9) were reported that are not accurate in this situation. So although it may appear that Person 1's mental model is the right size (i.e., seven facts reported), it is neither complete nor limited to what is true (i.e., it is inflated by X8 and X9). If any more facts had been reported, the story would clearly be inflated.

Person 2's story is even less complete. Although X2 was correctly reported, X3, X5, and X7 are still missing, and Person 2 also thinks that X8 is a relevant piece of information in this situation. With only five facts reported, the mental model is clearly incomplete, and a comparison to ground truth reveals some of the reported facts to be inaccurate.

This example also illustrates the distinction between accuracy and similarity of SA. As one looks across the boxes, it is apparent that X2, X3, and X7 were all true facts that were remembered by only one of the two people. Each fact was part of one person's accurate individual SA but was never shared with the other person, exemplifying the second and third shared SA outcomes in Jones and Endsley's (2002) list. Also, given that both people reported X8 as part of their account, that piece of information is an inaccurate but shared belief. That belief may increase both people's overall level of general agreement, but it certainly does not help them resolve the situation, thus illustrating the fifth outcome on the list. However, both people did have X1, X4, and X6 in their individual accounts, and all these facts were part of ground truth as well. As such, we can say that these two people had accurate shared SA (i.e., first outcome) with respect to these aspects of the situation.

Our study builds upon the shared SA outcomes identified by Jones and Endsley (2002), from which we propose a *similarity* measure that calculates the degree of SA shared between two team members. In Jones and Endsley's view, when two people provide the same incorrect response to individual SA measures, they have inaccurate but shared SA. We take a more conservative stance and maintain that there is no truly shared SA unless the understandings of individuals are shared and accurate. Furthermore, we propose that when there are multiple pieces of information to be aware of, or when the correct response to a SA measure is complex (e.g. Level 3), individuals can be correct to different degrees. By implication, neither individual nor shared SA is all-or-none, and not every error of SA is equally serious or far-reaching. Some errors may have immediate, fatal consequences, such as failing to detect a vehicle rapidly approaching on a collision course. Even in this example, however, a driver may have generally good awareness (e.g., watching signs and lights, tracking movements of other vehicles) but, because of a momentary distraction, may miss the one severe threat.

Other errors may be fatal only in accumulation or in unique combination (Chiles, 2002), as is the case when a system fault might be within safety parameters for a while and remain undetected by several operators before particular environmental conditions cause a broader system failure. It is rarely the case that a person will do everything wrong, and a single error, however serious or salient, is not necessarily an indicator that overall SA is poor. Therefore, in contrast to the traditional way of thinking about SA, we propose that evaluating SA with respect to its *degree of accuracy* would promote better accounting of the types and magnitudes of errors in SA measurement, and our similarity measure is designed to reflect this approach.

Measuring Situation Awareness

As suggested by Salas and colleagues (1995), two measures are critical in team situations: individual SA and team processes. We argue that both the accuracy and the similarity of individual SA levels will affect a team's shared SA. Although the SA construct has been widely researched, the multivariate nature of SA poses a considerable challenge in its quantification and measurement (for a detailed discussion on SA measurement, see Endsley & Garland, 2000b; Fracker, 1991a; 1991b). In general, techniques vary in terms of *direct* measurement of SA (e.g., objective real-time probes or subjective questionnaires assessing perceived SA) or *indirect* methods that infer SA based on operators' physiological state, behavior, or performance. Direct measures are typically considered to be *product-oriented,* in that these techniques assess an SA outcome; indirect measures are considered to be *process-oriented,* focusing on the underlying processes or mechanisms required to achieve SA (Graham & Matthews, 2000).

Our approach to assessing shared SA focuses primarily on using direct, objective measures of individual SA, as these have been extensively validated and shown to be reliable for a variety of domains. Essentially, objective measures assess SA by comparing an individual's perception of the situation or environment to some ground truth reality. Specifically, objective measures involve the collection of data from individuals by having them answer questions about the situation. Their responses can then be compared with what is actually happening to evaluate the accuracy of their SA at a given moment in time. As such, this type of assessment does not require operators or observers to make judgments about situational knowledge on the basis of incomplete information. Objective measures can be gathered in one of three ways: in real time as the task is completed, during an interruption in task activity, or in posttest following completion of the task.

One widely tested and validated approach to assessing SA is the Situation Awareness Global Assessment Technique (SAGAT; for a detailed description, see Endsley, 1995b). As a direct, objective measure of SA, the SAGAT methodology offers numerous advantages. First, SAGAT allows for the immediate assessment of SA by querying operators on their current perceptions of the situation. In addition, by comparing responses to the queries across team members, SAGAT can be used to provide insight into shared SA, and potentially team SA. SAGAT has been empirically validated with regard to its utility in providing valid and reliable assessment of SA across a variety of domains, including aviation (Endsley, 1990), air traffic control (Endsley, Sollenberger, & Stein, 1999), power plant operations (Hogg, Torralba, & Volden, 1993), teleoperations (Riley & Kaber, 2001), and military operations (Matthews, Pleban, Endsley, & Strater, 2000). The specific details of how SAGAT was used in our study to measure shared SA will be described in the Method section.

Prediction of Individual and Shared SA

Valid and reliable objective measures of SA, such as SAGAT, can be used to identify the critical variables underlying the formation of SA (and shared SA) as well as the possible relationships among these variables (Bolstad, Cuevas, Gonzalez, & Schneider, 2005). Our approach rests on the belief that shared SA is not a simple construct that can be attributed to a single predictor variable (e.g., the team's communications). Rather, shared SA entails a complex process in which multiple factors need to be considered. The complexity arises from having to account for both the factors that affect individual SA and those that contribute to any two team members' shared SA.

In the general model that this study was designed to test, there are three main components that affect SA formation: individual team member abilities, their interactions with other team members, and the environment in which they work (Bolstad et al., 2005). Within each of these components are multiple factors that affect shared SA formation and maintenance, such as geographical distribution, collaborative tool use, communication network proximity, and similarity of background knowledge and experiences. In the present study, we focused on the effects of individual and team factors on the SA of team members who were conducting military command, control, and communications (C3) operations from physically separated locations (i.e. distributed operations).

Individual Factors. One component of our analysis involved the similarity of the participants' cognitive workload, domain expertise, and task-relevant knowledge. Our first prediction was that (1) lower cognitive workload would be related to higher SA and shared SA. Fewer demands made on the participants' cognitive resources would potentially leave more capacity available for synthesizing the information needed to have good SA (Gonzalez & Wimisberg, 2007). We also hypothesized that higher SA and shared SA would be (2) directly related to higher domain expertise and (3) directly related to greater task-relevant knowledge, because these factors are known to be associated with better mental models of problems (Chi, Feltovich, & Glaser, 1981; Schunn, McGregor, & Saner, 2005).

Team Factors. Variables of social or organizational network distance are of particular relevance when distributed teams are relied upon to carry out mission objectives. A number of mathematical theories and modeling systems have been developed to characterize the degree of centrality and connectedness of people in networks (Graham, Gonzalez, & Schneider, 2007; Scott, 2000). However, we have deliberately chosen a less sophisticated approach. Most formal social network modeling techniques infer centrality and closeness from the behavior of the network, but this is a different direction of inference than we were attempting. In our study, a general network structure was dictated by the situation, and certain patterns of activity were expected based on that structure. The goal, therefore, was to determine whether observed levels of SA and shared SA would coincide with where participants were known to be located within the network.

In distributed C3 operations, a central location is often established as the organizational hub for team coordination. This gives rise to several possible association conditions, between any two given team members, that represent two types of distance at once: physical and organizational. Physical distance simply refers to whether the team members are colocated in the same physical location or are geographically distributed. As the term implies, *organizational hub distance* refers to the degree to which two team members are connected to the operation's central location (i.e., organizational hub). We hypothesized that (4) higher shared SA would be observed when participants were in close connection to the distributed team's organizational hub.

Furthermore, information exchange also serves as an input for building team member SA (Milham, Barnett, & Oser, 2000). Although communication is clearly important for information flow, it is also a social function in which team members can engage for reasons other than information exchange. Thus, another potential predictor of shared SA is the communication distance between any two team members. We hypothesized that (5) higher shared SA would result when team members were more closely linked with regard to communication distance.

Next, we describe a study that allowed us to test these hypotheses regarding the factors that may predict shared SA among distributed team members.

Method

Experimental Setting

Data were collected during a naturalistic study of military personnel engaged in a training exercise at the Joint Personnel Recovery Agency (JPRA), a subsidiary of the U.S. Joint Forces Command (JFCOM) of the U.S. military. JPRA is the primary coordinator of personnel recovery activities for the Department of Defense (DoD), which includes all efforts (e.g., military, civil, diplomatic) to return captured, missing, or isolated personnel from uncertain, potentially hostile environments. Such operations are formulated, planned, and executed under the supervision of JPRA, with the cooperation of multiple recovery centers abroad that are staffed by members of all military branches (Army, Navy, Marine Corps, and Air Force). The exercise was conducted at the Personnel Recovery Education and Training Center (PRETC), where the staff members for all the recovery centers are trained. Trainees complete a three-week course, composed of two weeks of learning the procedures followed by one week of simulated scenarios, all designed to mimic real recovery center operations.

Participants

Analyses were performed on data collected from a group of 17 DoD contractors and active service personnel (15 men and 2 women) from several branches of the military who were engaged in the simulation phase of their JPRA training. Participants were 27–40 years old (M = 34, SD = 4) and had 6–20 (M = 12, SD = 5) years of experience in their individual service branches. Seven participants were from the U.S. Army (41%), two from the U.S. Navy (12%), five with air force background (both U.S. and British Royal, 29%), and three from joint forces command (JFCOM, 18%). Three participants (18%) reported having prior experience with personnel recovery activities.

Simulation Exercise

Five personnel recovery simulation scenarios were conducted over the course of three days. Each scenario was unique and independent of the others, but all the scenarios were of the same general structure. We created the scenarios to become progressively more difficult by including increasingly more complex personnel recovery events. Each event was based on real-life recovery incidents (i.e., geographically isolated cases in which one or more people were found to be in need of recovery). For example, in Scenario 1, a fishing boat capsizes and the Navy recovery center is tasked with rescuing the missing persons. In Scenario 5, a U.S. plane is shot down over a hostile country and the surviving personnel are taken hostage by that country's military regime. In the latter scenario, the recovery center must not only coordinate the return of the personnel but must also negotiate with the foreign ambassador, the military-controlled government, and the Red Cross to ensure the safe release of personnel.

Each scenario took a half-day to complete; 2 hr of planning occurred before each scenario, followed by approximately 3-4 hr of execution. Participants were assigned to one of four team cells, which were distinct workstations for specialized activities. Three disciplinary branch team cells (Army, Navy, and Special Operations) conducted the operations in the simulation that would be conducted by the same military services in the real world, following the same protocols and with the same resources. For example, if an aircraft were shot down over the ocean, the Navy cell would be best equipped to deploy rescue and salvage ships, but for hostages on land, the Army and Special Operations teams would know better how to coordinate ground forces. An individual's real military branch affiliation and ranking were not considered when making assignments to these team cells, however. The fourth team cell was a joint service team that was expected to be responsible for overseeing and directing joint recovery efforts, but in order to do so, the members needed to be aware of all recovery activities at all times.

In this exercise, the introduction of new information to the network of team cells was controlled by a separate group of senior exercise administration staff who were designated the *white cell*. The white cell did not participate in the exercise but would inject new information into one of the four participating cells according to the procedure described later. Once the information was in the system, participants were responsible for sharing it with each other. In that regard, one of the intentions of the exercise was for participants in branch cells to focus on their specialized tasks rather than expend resources determining where to send new incoming information. It was expected that all branch team members would utilize the joint service cell as a central hub for information integration and distribution, but they were not told to do so. Instead, part of their training was to determine for themselves what would be the optimal strategy for information sharing.

Whether or not they discovered this strategy, the participants in all four team cells worked together simultaneously to complete each scenario. Each team was in a physically separated location from the others, but necessary communication between them was supported through e-mail, chat, a shared bulletin board, and fax. In addition, between scenarios, each participant was reassigned to a different team cell. This reassignment was also random, aside from the criterion that all participants serve at least once as a member in each of the four cells. Thus, all participants completed all the scenarios, even though the roles they played changed from one scenario to another.

Demographics Questionnaire

Questionnaires were administered to all participants to collect data on their age, gender, and current military service, as well as their branch and rank within that service. Expertise was based on participants' general military experience and was measured as the number of years served in the individual's service branch. Participants were also asked about their knowledge of procedures specific to personnel recovery. When individuals had prior experience with JPRA operations, they had additional background knowledge to apply to this task. Because only some of the participants had been involved in similar activities before, this factor was measured categorically, with a distinction between those who had such experience and those who did not.

Perceived Cognitive Workload

The National Aeronautics and Space Administration's Task Load Index (NASA-TLX) consists of a 6-item questionnaire that asks participants to rate their levels of perceived workload in terms of mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988). In this study, participants were asked to report their perceived cognitive workload by completing a modified version of the NASA-TLX. The same version was originally designed and used in other military exercises (Graham et al., 2007). Five items queried participants about the level of mental demand they had experienced in the intervening period, the amount of effort that they had to exert to complete tasks, their level of frustration, and so on. Responses were recorded on a 7-point rating scale, where 1 represented the least workload and 7 represented the highest level of workload. The cognitive workload measure was administered along with the SAGAT during each stop in the scenario.

Measurement of Individual and Shared SA (SAGAT)

Utilizing a concurrent memory probe technique, SAGAT involves, first, temporarily stopping the simulation at randomly selected times and removing task information sources (e.g., blacking out information displays); next, administering a set of queries that target each individual's dynamic SA information requirements (i.e., what they need to know at that point in time) with respect to the domain of interest; and then resuming the simulation (Endsley, 1995b, 2000b). In this study, each scenario was stopped at three random times to administer a set of 7 SAGAT queries (see Table 1). All together, SAGAT was administered 15 times: five scenarios, each stopped three times. To ensure that responses were independent, communication among participants was not allowed during the SAGAT stops.

Participants received training from the recovery staff and the experimenter on how to answer the SAGAT queries prior to beginning the exercise. Thus, for example, participants knew how to distinguish between low, medium, and high threats in the context of recovery incidents. At each stop, the white cell coordinator determined and officially recorded the correct answers to all 7 SAGAT queries for each current incident. The coordinator had assistance from the other members of the white cell (up to five other individuals) to help with the answer determination.

Responses to the SAGAT queries were scored in terms of proportion correct. Scoring of participants' responses to Query 1 was based on whether or not the participant identified the correct number of incidents occurring prior to that SAGAT stop. However, participants had to respond to queries 2 through 7 for each of the incidents that they identified in Query 1. Thus, scoring was based on the percentage correct for six queries times the number of incidents identified. Participants

TABLE 1. SAGAT Queries

1. How many isolated incidents are you aware of?
2. How many of these isolated incidents have been verified and validated as actual incidents?
3. Who is the SMC (SAR Mission Coordinator) for each incident?
4. Indicate the number and status of isolated personnel (IP) for each incident (OK, slightly injured, severely injured).
5. What is the current tactical situation around the IPs for each incident (high threat, medium threat, or low threat)?
6. What appropriate Joint Task Force and subordinate staff sections are aware of this incident?
7. What additional assets do you require to conduct a recovery?

were not told whether they had identified an incorrect number of incidents, but we asked them to provide additional details for as many as they indicated.

The number of incidents that participants had to resolve differed between scenarios, and their tracking of incidents was pivotal to this analysis. The SAGAT queries were connected as a set primarily by their focus on incident-level knowledge, and a given scenario could have a total of 3 to 12 incidents. Not all incidents were presented to participants at once. Scenarios would begin with the presentation of at least one incident, and during the time participants were working on resolving known incidents, others would be brought to their attention. As such, the workflow had to integrate new incidents as they were presented, but the scenarios were not complete until all incidents had been resolved.

The results that will be presented hinge on two unique features of the way SAGAT was administered in this study. The first notable distinction is that the SAGAT queries were asked in the same order at every stop. Most assessments of situation awareness that use SAGAT randomize the order of the items in an effort to control for confounds in the measures (Endsley, 1995b). Second, unlike in other studies, we did not have a large selection of SAGAT queries, because our ability to collect ground truth was severely limited. We had to rely on the white cell coordinator, who was also coordinating the information insertion process, to provide us with the correct answers at each stop. Thus, unlike in more traditional SAGAT implementations, we elected to use only seven queries and ask the queries in the same order each time.

In this study, the particular value of asking all the queries in the same order was that a participant's answer to the first query ("How many isolated incidents are you aware of?") was an immediate indicator of how accurate that participant would be on the overall knowledge measured by the remaining queries. All the other queries had to be answered for the number of incidents indicated by the participant in Query 1. If the participant's response was incorrect on Query 1, then the mental model characterized by the responses to the other six queries either would be incomplete or would include more information than was true, depending on whether the participant identified a lower- or higher-than-correct number of incidents, respectively. Furthermore, assuming that a participant correctly identified the number of incidents, and that all the incidents he or she had recalled were real, it was still possible for the individual to characterize the details of the incidents incorrectly when answering the other queries. For any given incident under consideration, participants could misidentify the coordinator or the number of personnel involved, or they could incorrectly describe the tactical situation, and so on. As such, there were two key ways in which participants could exhibit error in their SA.

To take an example, suppose that three incidents have actually been presented by the time a stop occurs, but a participant reports that he or she is aware of four incidents. This response incorrectly represents awareness of up to 33% more information than actually exists in the situation. Thus, the measure of SA will be inflated by one incident's worth of information that is not real (i.e., imagined; an incident that exists only in the mind of that person). In addition, the details of this imagined incident could interfere with the accuracy of the participant's SA for the real incidents, in that it might be a false amalgamation of information from those incidents. Alternatively, if the participant reports being aware of only two incidents, there is already a clear 33% lack of SA at that point in time, to which any other errors of SA on the other two incidents must be added.

This is not to say that reporting too much information is the same kind of mental model error as reporting too little information. Detailing the distinctions between these errors is beyond the scope of this study, and our ability to pinpoint particular types of errors was limited. In this investigation, the general magnitude of error could be evaluated relative to how much more or how much less information was reported than was required, and even this rough measure of general error was a good starting point to examine the dynamics of accounting for SA error. Within the context of this study, participants' SA was evaluated according to the extent to which they had the correct number of incidents and characterized most or all of them correctly.

Procedure

Before the simulation exercise began, participants completed a demographic questionnaire, which included the questions about their military and prior JPRA experience. Copies of booklets that contained the communication and SAGAT questionnaires were distributed, and participants were given a chance to ask any clarification questions about the questionnaires. The training instructors provided an overview of the exercise and the team cell tasks. These descriptions were brief, as part of the exercise included students learning what to do and how to do it.

Participants were given their cell assignments and position assignments on the morning they arrived at the training center to begin the exercise. Each center contained a large table, a whiteboard, full-wall map, two phones, and one computer. Additionally, each team cell was given an update briefing, which included a set of resources with information about their assets (e.g., supplies and supply lines, vehicles and equipment, manpower) and their location, status, and capabilities for the specific branch of service.

Participants began each scenario with a 2-hr planning session in which they took stock of their assets. Once the scenario began, the recovery incidents were injected into the exercise by the white cell. At semiscripted times within the exercise, a member of the white cell would e-mail, make a call, or send a fax to a predetermined cell to communicate the incident. Over the course of the exercise, all of the possible communication methods were used to inject information into all of the cells at some point. The scenarios were completed when all the incidents were resolved.

Preparation of SA and Shared SA Measures

Because all SAGAT queries are assumed to be of the same importance in most studies that use this method of SA measurement, the baseline measure of individual SA at each stop was calculated as the overall mean accuracy; the sum of the number of correct answers from all items was divided by the total number of answers possible at that stop. We will refer to this measure as *inflated SA,* for two reasons. First, the aggregated score included Query 1, which was redundant with the other items and artificially increased the average. Second, this measure did not correct for any SA error and therefore false memories of information were not adequately penalized.

The first step in getting a true measure of SA, then, was to calculate the overall mean accuracy of the SAGAT scores for only queries 2 through 6 (i.e., not including Query 1), which captured the substantive information of a person's mental model in this situation. This mean was then scaled by the accuracy in Query 1 (SAQ1), to result in the *true SA*. Accuracy in Query 1 was calculated as 1 minus the difference between the number of incidents that a participant reported and the actual number of incidents, divided by the actual number of incidents.

To illustrate this procedure using the earlier example, if four incidents are reported when only three actually occurred, the participant's SA score is off by 0.33, which implies that the mean SA of queries 2–7 would be scaled by 0.67 (i.e., 1 – 0.33). The derivation of these measures of individual SA is illustrated in Figure 2.

Shared SA measures corresponding to individual SA measures were calculated using the same general set of steps that were used for individual SA, but with some modifications. An additional step was to calculate the degree to which SA was shared between two participants. First, each participant's data were paired with every other participant's data, such that all possible unique participant pairings were generated (a total of 136 pairings). Once the data for each pair were aligned, the following standard similarity formula was used to assess the similarity of the levels of SA exhibited by each individual participant in the pair, where p_1 and p_2 are the SA measures of each participant and $(p_1 - p_2) / (p_1 + p_2)$ represents the proportion of deviation between their scores. This formula was used to generate shared SA measures as well as several predictors of shared SA: workload similarity, expertise similarity, and knowledge similarity.

$$\text{Similarity} = 1 - \text{absolute value of } [(p_1 - p_2) / (p_1 + p_2)] \quad (1)$$

To determine shared SA, the similarity formula was applied to the SAGAT query scores of each unique pair and then the mean of the similarity scores was

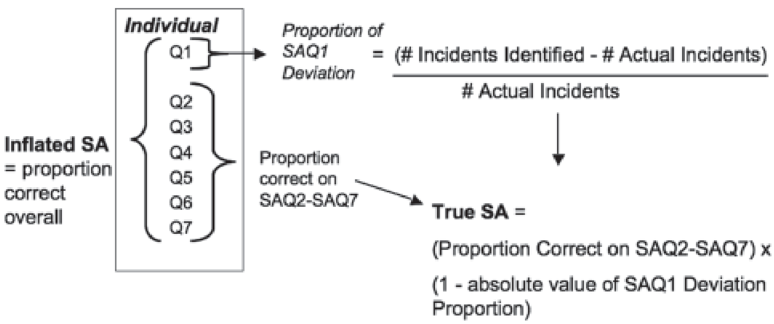

Figure 2. Derivation of individual situation awareness measures.

calculated. An important feature of the formula is how conservative it is with respect to assessing similarity. When both participants in a pair have an SA score equal to zero, their final similarity score will be zero by default. When only one individual has an SA score equal to zero, then the resulting proportion will be the other participant's SA score divided by itself, which is 1, and ultimately this will also yield a similarity score of zero. Thus, both participants must have an SA score greater than zero to generate a similarity score greater than zero. When both participants have a perfect SA score, the result of the division is zero, such that the final similarity score will be 1 (i.e., perfect similarity).

In sum, applying this formula to two participants' SA on a given SAGAT query provides a measure of their shared SA specifically based on the similarity of proportional accuracy, and it accounts for the shared accuracy of the two individuals at the item level. At this point, the average of the similarity scores (queries 1 through 7) was calculated to get an overall *inflated shared SA* score for a participant pair.

We again adjusted the inflated shared SA by calculating it without the similarity score from Query 1. Another modification had to be made with respect to scaling for Query 1 SA error, however, because the deviation proportions of the two participants' scores on Query 1 could have opposite signs. One participant might overestimate the number of incidents and the other might underestimate the number, which would make their SA less similar than if they had both overestimated to different degrees.

This is an example of the fourth outcome suggested by Jones and Endsley (2002); both participants are wrong, and they have different answers to the query. As such, the deviation proportion of each participant was squared, the products were added together, and the square root of the sum was calculated. Essentially, this procedure is identical to calculating the standard deviation of a data set, except that only two data points are involved and the deviations are from ground truth rather than from a mean. The value that resulted from this procedure was a measure of shared error, and the inverse of this value was multiplied by the adjusted mean to yield the *true shared SA*. The process for generating this measure is illustrated in Figure 3. In sum, analyses of the SAGAT data collected in this study will be based on two measures of individual SA (inflated SA and true SA) and two measures of shared SA (inflated shared SA and true shared SA).

Predictors of Individual and Shared SA

The measures of individual and shared SA described earlier served as the dependent measures of SA. Four factors (expertise, knowledge, cognitive workload, and team cell membership) were measured at the individual level and used first as predictors of individual SA. Expertise and knowledge (of procedures specific to personnel recovery) were derived directly from the demographic questionnaire. Cognitive workload was determined from responses to the NASA-TLX. Team cell membership was recorded as part of the research protocol.

For predicting shared SA across pairs of participants, we calculated the values of shared cognitive workload, shared expertise, and shared knowledge for each of

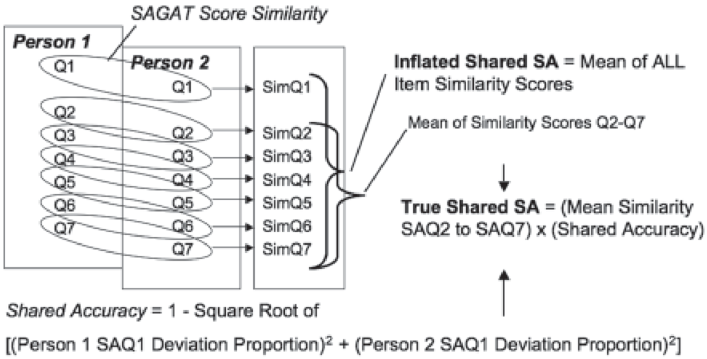

Figure 3. Derivation of shared situation awareness measures.

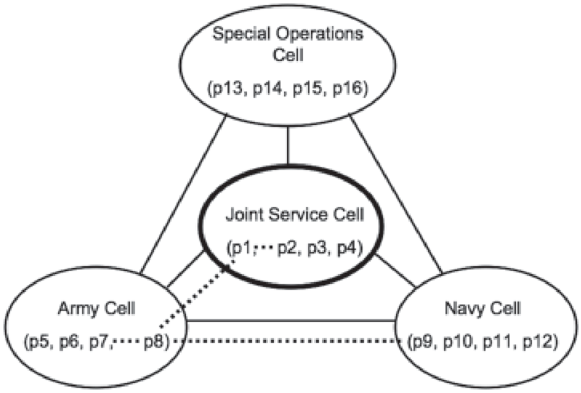

Figure 4. Diagram of organizational hub distances between participants.

the unique pairs of participants, using the similarity formula described earlier. In addition, membership in the joint service cell, which served as the organizational hub, was used as a predictor of shared SA; however, the relationship to the joint cell for each pair had to be expanded. Because two participants had to be considered with each case and because each cell was in a different physical location, distinctions also had to account for physical distance between participants in the pairs (Bolstad et al., 2005). The diagram in Figure 4 illustrates these relationships.

In this study, the variable *organizational hub distance* was used to capture the degree to which a pair of participants was connected to the centralized joint service cell and the degree to which the individuals in a pair were connected to each other by location. We established levels of this variable by assigning numbers to each condition, with higher numbers corresponding to greater distance from the joint cell (both in joint cell = 1, one in joint cell = 2, both in same branch cell = 3, each in separate branch cell = 4).

An additional variable that was measured with respect to paired participants was the *communication distance* between the members. At each scenario stop, participants were asked to consider all the other participants in the whole group (i.e., members of all cells) and to rank them according to frequency of communication with them. This measure has also been used in other research with military teams (Graham et al., 2007). When one participant did not communicate with another, he or she put zero in the communication frequency rank space or left it blank. Frequency ranks ranged from 1 to 9, with the rank of 1 given to the person with whom he or she communicated most.

The communication distance measure used in our analysis was a function of each member's rank of the other within each pair of participants. If both members of a pair ranked the other as number one, they were as closely linked as they could be in terms of communication. If both members of the pair had zero or blanks, indicating no communication between them, the participants were at maximum communication distance from each other (i.e., neither was even in the other's communication network). The frequency ratings of each pair, at each time point, were reverse-coded and scaled to yield communication distance scores that ranged from 0.05 to 1, with 0.05 representing the lowest possible distance and 1 representing no communication between them.

Results

To test which factors were generally predictive of SA and shared SA, we averaged the values on every measure (both SA and predictors) across the 15 time points at which they were measured. This was done for each participant and then for each unique participant pair. Linear regression was used to generate a predictive model. As described earlier, we believe that the level of SA in a team situation is influenced by cognitive, social, and environmental factors all at once. For this reason, all predictors were entered into the models simultaneously. However, models were of interest only when the F statistic for the model was significant at the $p < .05$ level and when at least one predictor significantly predicted the measure of SA or shared SA at the $p < .05$ level. It was also expected, however, that the effects of predictors would change as the participants progressed through successive scenarios and gained more practice. As such, in addition to the overall models, the data were also split by scenario to determine whether the impact of predictors would change in relation to increased experience with the simulations.

Prediction of Individual SA

With 17 participants measured at 15 time points, there were $N = 255$ cases to consider for assessing the frequency of individual SA outcomes. Perfect SA was assessed as 100% correct on the SAGAT queries, and 0% correct was defined as a complete lack of SA. Any value between these represented partial SA. Nearly all the participants had partial SA at each of the 15 time points ($N = 247$, 97%). In no case did participants have perfect SA across the queries, with or without counting

Query 1. Only in three cases (1%) was there SA for the first query, one of which was perfectly correct and two of which were partially correct. There were five participants for whom no value of SA was recorded at the last time point, and those cases (2%) were excluded from analysis.

As expected, a significant difference was found between the individual inflated SA and true SA measures, $t(16) = 10.49$, $p < .0001$. When corrected for accuracy, participants' true SA scores ($M = 0.21$, $SD = 0.07$) were significantly lower than their inflated SA scores ($M = 0.27$, $SD = 0.07$). We chose to use the more conservative true SA scores to determine which factors predicted individual SA, as this measure would provide the truest indication of participants' SA.

The predictors cognitive workload, expertise, knowledge (of JPRA procedures), and joint-cell membership were entered simultaneously and regressed onto true SA. The linear regression analyses showed that these variables did not significantly predict individual SA in this situation. The model was not significant ($F < 1$), and none of the predictors was found to be significant. There was also no significant prediction of individual SA when the data were broken down by scenario. As such, none of the hypotheses about individual SA (Hypotheses 1–3) was supported.

Prediction of Shared SA

The general outcome pattern for shared SA was comparable to that of individual SA; however, outcomes for shared SA were measured with respect to similarity of individual SA scores. Perfect shared SA, therefore, exists when both individuals were completely correct on all the SAGAT queries, thus implying perfect agreement. A complete lack of shared SA would require that both individuals have a complete lack of individual SA.

A similarity score between 0 and 1 represents partial shared SA. There were 136 unique pairings of participants, which, when multiplied by 15 time points, yielded $N = 2{,}040$ cases. When those who had no observed values of individual SA at the one time point (as discussed earlier) were paired, there were 70 cases (3%) with no shared SA measured. These were excluded from the analysis of shared SA.

There were no cases of perfect shared SA. Most of the cases ($n = 1{,}278$, 63%) involved partial shared SA on all the queries (1 through 7). In 471 cases (23%), there was perfect shared SA on Query 1 and partial shared SA on the remaining queries (1 through 7). In addition, in 60 cases (3%) there was no shared SA on Query 1 and partial SA on the remaining queries.

Once again, as expected, true shared SA based on participants' SAGAT similarity scores ($M = 0.40$, $SD = 0.05$) was significantly lower than their inflated shared SA ($M = 0.60$, $SD = 0.06$), $t(135) = 31.71$, $p < .001$. As before, the more conservative true shared SA scores were used as the dependent measure for the linear regression analysis. Workload similarity, expertise similarity, knowledge similarity, organizational hub distance, and communication distance were regressed as predictors onto true shared SA, both overall and by scenario. The models are shown in Table 2.

TABLE 2. Regression Models of True Shared SA

True Shared SA	F	Adj. R^2	Constant	Experience Similarity	Shared Knowledge	Workload Similarity	Organizational Hub Distance	Communication Distance
Overall	5.11**	.21	–.03	.09	.26*	.08	.50**	–.18
Scenario 1	2.56*	.09	.02	–.07	–.02	.18	–.26*	–.08
Scenario 2	1.55	.03	.39	.04	–.05	–.19	–.26*	.00
Scenario 3	1.66	.05	.10	.19	.09	.04	–.26*	–.02
Scenario 4	2.62*	.11	.26	.08	.31*	–.03	–.17	–.06
Scenario 5	5.79*	.24	.42	.11	–.19	–.16	.45*	–.24*

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

With regard to the overall model, two predictors—shared knowledge and organizational hub distance—exhibited a significant relationship with true shared SA scores. Shared knowledge was positively related to shared SA, such that when two participants had prior experience with JPRA operations, their shared SA was higher and Hypothesis 3 was supported overall. Organizational hub distance was also positively related to shared SA. The more disconnected a pair of individuals was from the joint service cell, the greater was their shared SA, but this is the opposite of what we expected, and Hypothesis 4 was not supported in the overall model. Also, because there were no significant effects of shared workload, shared expertise, or communication distance, we found was no support for Hypotheses 1, 2, and 5, respectively, in the overall model. To investigate these relationships further, we examined the corresponding models for each scenario.

The overall model for Scenario 1 was significant. However, the only significant predictor was organizational hub distance, and in this scenario, it was negatively related to shared SA. This meant that if a participant pair was more separated from the joint service cell, their shared SA was lower.

Organizational hub distance was also a significant predictor in Scenarios 2 and 3, with the same relationship to shared SA, but the overall models were not significant. In Scenario 4, the model again predicted shared SA. In this case, the only significant predictor was shared prior knowledge of JPRA operations. Finally, in Scenario 5, the overall model was significant, with organizational hub distance again as a significant predictor. Particularly in this scenario, when a participant pair was more disconnected from the joint service cell (i.e., at a greater distance from it), they were likely to have higher shared SA. At the same time, communication distance was also a significant predictor in Scenario 5. As the communication distance between two participants increased, their shared SA decreased. In summary, there was a strong relationship between organizational hub distance and shared SA, as expected. However, this relationship was in the opposite direction of what we hypothesized overall.

Additional Considerations

Before we proceed to the discussion of our results, several validity checks must be reported. First, the correlations between the predictors were calculated to ensure no confounds with interpretation of the results. The matrix of correlation coefficients is shown in Table 3.

For the most part, no collinearity was found between the predictors. The positive correlation between communication distance and organizational hub distance was significant but small, indicating that, to some degree, when people had a weaker link to the joint service cell, they were less likely to communicate with each other as well. However, communication distance was not a significant predictor overall. The only time at which it was significant was in Scenario 5, and then its relationship to shared SA was opposite that of organizational hub distance.

Shared knowledge also had a small, significant, but negative correlation with organizational hub distance. As the latter increased, the likelihood that both

TABLE 3. Correlations Among Predictors of True Shared SA

Predictor	1	2	3	4	5
1 = Communication distance	1.00	−.02	.03	.00	.199*
2 = Workload similarity		1.00	−.01	−.13	−.13
3 = Experience similarity			1.00	.09	.03
4 = Shared knowledge				1.00	−.207*
5 = Organizational hub distance					1.00

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

participants had prior background knowledge of the task-relevant procedures decreased. This relationship is most likely attributable to the fact that the likelihood of being in a branch cell was roughly three times the likelihood of being in the joint service cell at any point in time.

Another issue to consider is that although we generally expect peoples' SA to improve with general task experience, the procedure by which SA is measured should not cause better SA. The reason SAGAT questions are usually randomized is to prevent participants from being aware of what they will be asked. In this paper, we emphasized that asking the queries in the same order gave us a unique opportunity, but to determine whether there was any confound associated with this approach, we ran a 3×5 repeated-measures ANOVA with scenario and stop as predictors of SA.

No effect of stop on participants' SAGAT scores was exhibited within each scenario, indicating that there was no general improvement from the beginning of a scenario to the end. There was a significant effect of scenario, $F(4, 64) = 15.42$, $p < .01$, and an interaction between scenario and stop, $F(8, 128) = 9.82$, $p < .05$. However, the overall differences to which these effects refer were not associated with a steady increase in SAGAT performance, as can be seen in Figure 5. The same level of SAGAT performance was observed in Scenarios 2 and 3, and the level of SAGAT performance observed in Scenario 4 was much higher than that observed in Scenario 5. In general, then, because SA did not simply increase with each additional scenario that was completed, the pattern of SAGAT performance does not support a learning effect interpretation of differences across scenarios. It is more likely that the predictors measured in this study, alone or in combination with some other factor or factors, were responsible for the SA observed in each scenario.

Finally, because the organizational hub distance variable was inclusive of both physical distance between individuals and joint cell association, it was important to determine, in a manner that would not confound them, whether either of these component factors by itself was a significant predictor. As such, all the models were rerun twice, with all the same predictors, but once with physical distance substituted for organizational hub distance and once with joint cell distance substituted. When the analyses were run using physical distance, none of the models was significantly predictive of shared SA overall, and physical distance was a sig-

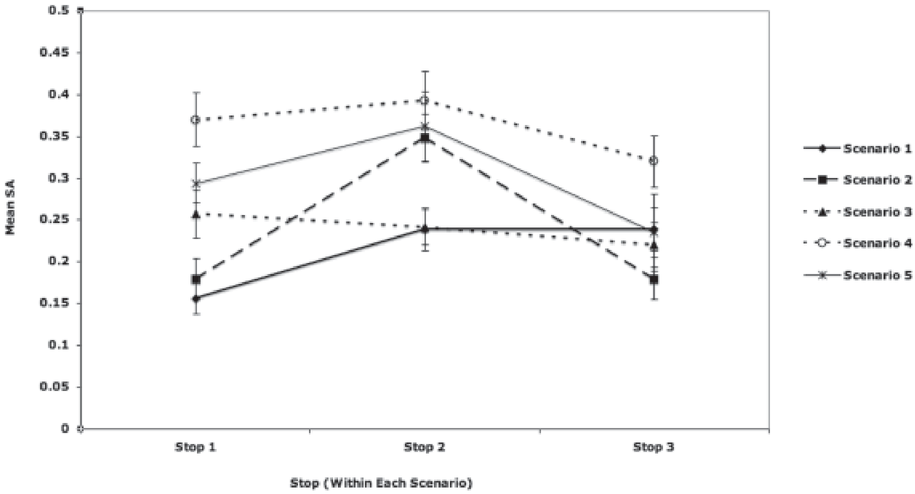

Figure 5. Mean SAGAT scores over time.

nificant predictor only in Scenario 3. As such, physical distance alone did not account for the patterns of prediction.

In contrast, when joint cell distance alone was used as a predictor, the pattern of prediction across the five scenarios was almost identical to that observed with organizational hub distance. Aside from small differences in the beta weights, the only difference in prediction pattern was in Scenario 5. When organizational hub distance was used, it and communication distance both predicted shared SA. When joint cell distance replaced organizational hub distance as the predictor, it was the only predictor of shared SA in that model, and its effect was at a higher level of significance. This result suggests that the particular cell memberships of paired participants are more strongly related to shared SA than is the physical distance between the individuals.

Discussion

This research contributed in two key ways to the current literature on the measurement of shared SA in teams. First, we proposed a similarity metric for use with individual awareness measures to calculate shared SA. Cooke, Stout, and Salas (2001) asserted that a team mental model is more than a summation of individual models, and that a holistic shared model requires that knowledge be integrated based on its similarity and distribution across individuals. Our proposed similarity metric extends the theory of Cooke et al. that a shared model emerges through coordination, in that we propose a concrete and simple way in which individual SA may contribute to generating a team model. Our simple and intuitive measure is not meant to capture all the complexity in a team interaction process but, rather, to be a potentially useful first step in assessing shared SA between paired team members. Second, we elucidate the process by which integrated team SA can be based on individual comprehension and other predictive factors. Although several of our results do support an action process account of team SA (Gorman et al., 2006; Salas et al., 1995), the fact that knowledge-based measures of shared SA were predicted by coordination-related factors suggests that knowledge elicitation is also an important component in measuring shared and team SA (Cooke et al., 2001). These two main contributions are discussed next.

Measuring Shared SA

The first issue examined in this paper was a way in which inaccurate individual SA can limit overall SA formation and potentially compound error in shared SA. We observed that scaling SAGAT queries 2 through 6 by the accuracy of SAGAT Query 1 led to a significant reduction in the observed magnitude between an inflated assessment of SA and a truer assessment of SA for both individual and shared SA. This finding illustrates that failure to account for all sources of error may result in exaggerated values of SA and overestimation of human performance. Furthermore, with respect to shared SA, inasmuch as two people working in a team may sometimes correct each other's representations of the situation (Cannon-Bowers & Salas, 1998), the errors of individuals can amplify the error in shared SA when different understandings become more divergent.

The error correction procedures that we have proposed here allow for more accurate measurement of the degree of shared SA under each of the outcome conditions that Jones and Endsley (2002) identified, albeit on a proportion basis. When both individuals have good SA, the adjustment to shared SA will be minimal, if it is needed at all. When one individual has good SA and the other has poor SA, the shared SA will be measured as lower in proportion to the second individual's level of error. When both individuals are incorrect and have different responses, then our procedure can factor in the degree of divergence between their responses, penalizing their measure of shared SA more than would be the case if both had the same incorrect answer.

The shared SA measurement procedure that we proposed here was not without limitations. One limiting factor in this study is that because the items were scored by the white cell coordinator, we were unable to measure the exact nature and direction of the specific errors on items 2 through 6. Members of the observing white cell were also trying to record what was correct in real time at each stop, which made it difficult for them to be detailed in their evaluations. In addition, some questions could have a number of valid responses, and the correct answers for those questions required more subjective judgments, which is why items were scored on a proportion basis. In any case, people were almost never categorically incorrect on their SA, and it is only when the extreme outcomes occur (i.e., completely correct or completely incorrect) that the discrete outcomes identified by Jones and Endsley (2002) can be counted and compared directly. Neither of these outcomes was observed in our data. The majority of cases of both individual SA (97%) and shared SA (96%) exhibited partial SA. As such, the intermediate outcomes (i.e., cases in which Person A was correct and Person B was incorrect, or vice versa) were observable only in relative terms.

The features of our similarity formula were also such that the formula did not distinguish which person was more correct or less correct, and the agreement between individuals was inseparable from their shared agreement with the truth. The formula was specifically generated to provide an assessment of how close the individuals were to each other in their general level of SA when there was some proportional deviation from ground truth between them.

Predicting SA and Shared SA

In this study, the individual factors measured and used as predictors were not significant predictors of individual SA, whereas factors related to shared SA were significant. We believe that the reason individual factors were unreliable predictors of individual SA in this context is that the personnel recovery mission is inherently a team task. Personnel recovery missions were planned and executed by multiple military branches working in cooperation, so an individual's SA depended on sharing information with other participants. This finding appears to support the theory of Gorman, Cooke, and Winner (2006) that team SA is determined more by coordination among team members than by the knowledge of individual team members. Furthermore, inasmuch as knowledge is coordinated in a team setting, it may be that the effects of individual factors that would generally be expected to predict individual SA, such as experience and cognitive workload, were attenuated through the coordination process as well.

If being part of a team transforms SA itself, then it stands to reason that the predictors of individual SA are also transformed in such a setting. Perhaps cognitive workload fluctuates differently when one works with others than when one works alone. Similarly, one's own experience in an area of specialization may be a strong predictor of individual SA, but when one is put into a situation in which the task role is independent of one's career training, experience in one's own field will be of minimal advantage. In any case, these null results do not eliminate the possibility that shared SA derives from individual SA, but they do suggest that the relationship between them may be more complicated and less pronounced than we originally hypothesized.

Two factors, organizational hub distance and shared knowledge, were significant predictors of shared SA. Organizational hub distance was shared by definition. However, the observed relationships between organizational hub distance and shared SA were the opposite of what we hypothesized, both overall and with respect to changes in team functioning across the scenarios. The general expectation (i.e., Hypothesis 4) was that a weaker link to the joint service cell would be associated with lower shared SA. Instead, we found that shared SA was lower as participants were more closely linked with the joint service cell. One possible explanation for this is that the information flow (i.e., the pattern of information exchange among participants, both within and between cells) did not develop in the way we expected. The theory behind the joint service cell was that information flow would be most efficient if members of branch cells would funnel new information into that cell, and that members of the joint service cell would then disseminate the information to all the branch cells that needed it. Because there were always about three times as many branch cell members as there were joint service cell members, however, it is possible that the most relevant information simply floated among the branch cells without actually going through the joint cell. Consequently, discussion of information within the branch cells may have led to higher shared SA among the members of those cells, thus reducing the need for the joint service cell.

This possibility is also consistent with the SA through coordination hypothesis (Gorman et al., 2006) and is a plausible outcome of team self-correction (i.e., the knowledge of other team members corrects one individual's error in understanding; Cannon-Bowers & Salas, 1998). Further support is found in the observed shift in the influence of joint service cell membership from the first few scenarios to the later scenarios. The next step in this research will be to examine more closely how participants respond to new information, accounting for where the information is injected into the task system and observing more closely the strategies used by operators to interpret and share information.

Shared knowledge emerged as the only significant predictor in Scenario 4. This break in the influence of organizational hub distance was striking and marked the shift in direction of its relationship to shared SA, from negative to positive, as described earlier. In this study, shared knowledge specifically of JPRA-related procedures was a separate variable from shared level of general military expertise because the latter was not a determining factor in what roles participants had in the exercise.

It was important to test two opposing possibilities: that general experience would help participants compensate for not having had prior specific experience on the task at hand, or that specific prior experience with the task would help participants navigate the gap between their usual job responsibilities and the tasks to which they were assigned in this exercise. It is therefore notable that having specific shared knowledge did have a significant effect on shared SA in this situation and that having similar levels of general expertise did not, which is consistent with the second possibility. Given that only a small minority of participants had prior knowledge of JPRA-related procedures ($n = 3$, 18%), however, it is not unreasonable for those individuals to have taken three scenarios to connect with each other and coordinate their knowledge. If those who had expertise in the domain were located primarily in branch cells at that point, they may have emerged as single-person hubs for interpreting and applying incoming information.

The influence of shared knowledge also suggests an alternative explanation for the change in impact of the organizational hub distance: namely, that new, overall shared knowledge emerged over the course of the exercise. Participants were reassigned to new cells in each scenario and played different roles from one stage of training to another, a procedure referred to as *cross-training* (Cannon-Bowers & Salas, 1998). Because of this, most of them were likely to have been in the joint service cell at least once by the fifth scenario, and it is reasonable to expect that those participants who had no background knowledge of JPRA logistics before the exercise began would have gained a sufficient amount by the fourth scenario for this knowledge to make a difference in their SA. The follow-up analysis revealed that the effect of organizational hub distance was indeed carried by the cell membership factor (both paired participants in the joint service cell, only one in joint service cell, or neither in joint service cell), such that it made a difference where paired participants were in relation to the joint service cell. Consequently, the joint service cell experience that participants gained just from completing the first few scenarios may have served as a catalyst for the shift in relationship between organizational hub distance and shared SA.

Communication distance also emerged as a significant predictor in the last scenario, and it had a negative relationship to shared SA, as expected. The more communication distance there was between people, the lower their shared SA. It is also notable that communication distance first emerged as a limiting factor in the sharing of knowledge among participants in Scenario 5, just at the point that joint cell experience was showing itself to be beneficial to people in branch cells. It is possible that communication distance was a factor throughout all the scenarios but that its effect was overshadowed by the gradual development of coordination practices. Once the participants knew what information to share with each other, the only limiting factor was their ability or opportunity to communicate with particular people across distances. There also may have been other pragmatic factors at work in the situation, such as peoples' willingness to share information or selecting the best mode of communication. In this study, we were unable to gather detailed data on these factors, but this too is a goal for future research.

Generalizing Shared SA Similarity Metric

We have presented a procedure for assessing shared SA and its relative accuracy with a direct measure of individual SA, and we believe this general procedure can be applied to other individual SA measures as well. A core thesis of this study is that there is no SA when the information that a person or team believes is not actually true, and accuracy has been recognized as important in most of the major SA measurement methods as well. For example, as described by Jones (2000, p.118), "information quality" is one scale in the Situation Awareness Rating Technique (SART) on which an operator's "understanding" of the situation is measured. This scale measures the operator's assessment of how reliable and valuable the information is in resolving the situation, but in this regard the accuracy of information is itself a matter of subjective judgment, and there is no ground truth with which to compare SA. In addition, when SA is being measured by an observer's rating, it is only when the questions are specific and refer explicitly to a particular piece of information that an observer's rating of quality can be weighted and interpreted.

Finally, in the SA as performance approach, air traffic control operational error reports also have been used to link the level of SA of those involved to the specific effects of the errors (Durso, Truitt, Hackworth, Crutchfield, & Manning, 1998;

Rodgers, Mogford, & Strauch, 2000), but this performance can be evaluated only when there is a predetermined set of correct actions, safety parameters, or error categories as well. In sum, the procedure that we have presented for assessing the similarity of individual SA scores can be applied to any measurements taken on a ratio scale, but the degree to which the shared scores reflect accuracy will always depend on how well accuracy is integrated into the individual scores first.

Conclusion

In conclusion, this study suggests that sources of error in SA and methods for correcting it should be a standard topic in discussions of individual and shared SA measurement. Although our method will be more easily tailored for use in some situations than others, we hope that these analyses will raise further awareness and lead to more dialogue about how to measure SA accurately. Our results do indicate that good shared SA is a matter of both knowledge and coordination, and that neither aspect can be focused on to the exclusion of the other. Generally speaking, it would be ideal if all the knowledge elements that SA requirements comprises in a given situation could be broken down into basic, objective items that could be assessed by checking off their correctness or incorrectness in each operator's response. In those conditions, our similarity measure would not be needed, and correlations of outcome patterns would be much stronger assessments of similarity. As it stands, it is often possible to judge the correctness of SA only with respect to the relative or proportional quality of a response, regardless of whether SA is assessed in terms of action or knowledge. The degree to which an assessment of SA is objective is limited by the degree to which there is ground truth or a priori prioritization of operational responses to situations. It is precisely because such rubrics are not available in many situations that a calculation such as the one we propose is a necessity, but more research is needed to develop and validate this metric further.

In future work on the measurement of shared SA, we will attempt to address more of these dynamic factors involved in team work (information flow, communication flow, physical location, etc.), particularly the specific roles adopted by individuals and the distribution of task experts within the larger team. If a team is ultimately successful in completing its task, there must be a way to describe why particular knowledge, actions, or coordination patterns led to the success. Without a measure of what particular beliefs the participants were operating on, and whether participants were operating on similar or different beliefs when they took similar actions, there is no sure way to reproduce success in later, similar situations. The effect of shared SA on performance is a question that is still unanswered in team research, and we plan to pursue this in future work.

Acknowledgments

This research was partially supported by the Army Research Laboratory (DAAD19-01-2-0009) awards to Cheryl Bolstad and Cleotilde Gonzalez. We thank

Diane Eberly of the U.S. Army Evaluation Command and the staff at the Joint Personnel Recovery Agency, Personnel Recovery Education and Training Center, for their support of this research effort.

References

Adams, M. J., Tenney, Y. J., & Pew, R. W. (1995). Situation awareness and the cognitive management of complex systems. *Human Factors, 37*, 85–104.

Bolstad, C. A., Cuevas H. M., Gonzalez, C., & Schneider, M. (2005, May). *Modeling shared situation awareness.* Paper presented at the 14th Conference on Behavior Representation in Modeling and Simulation (BRIMS), Los Angeles.

Brannick, M. T., Prince, A., Prince, C., & Salas, E. (1995). The measurement of team process. *Human Factors, 37*, 641–651.

Cannon-Bowers, J. A., & Salas, E. (1998). Team performance and training in complex environments: Recent findings from applied research. *Current Directions in Psychological Science, 7(3)*, 83–87.

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science,* 5, 121–152.

Chiles, J. R. (2002). *Inviting disaster: Lessons from the edge of technology*. New York: Harper Collins.

Cooke, N. J., Gorman, J. C., Duran, J. L., & Taylor, A. R. (2007). Team cognition in experienced command-and-control teams. *Journal of Experimental Psychology: Applied, 13(3)*, 146–157.

Cooke, N. J., Stout, R., & Salas, E. (2001). A knowledge elicitation approach to the measurement of team situation awareness. In M. McNeese, E. Salas, & M. R. Endsley (Eds.), *New trends in cooperative activities: Understanding system dynamics in complex environments* (pp. 114–139). Santa Monica, CA: Human Factors and Ergonomics Society.

Durso, F. T., Truitt, T. R., Hackworth, C. A., Crutchfield, J. M., & Manning, C. A. (1998). En route operational errors and situation awareness. *International Journal of Aviation Psychology, 8(2)*, 177–194.

Endsley, M. R. (1990). Predictive utility of an objective measure of situation awareness. In *Proceedings of the Human Factors Society 34th Annual Meeting* (pp. 41–45). Santa Monica, CA: Human Factors and Ergonomics Society.

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

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

Endsley, M. R. (2000a). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 147–173). Mahwah, NJ: Erlbaum.

Endsley, M. R. (2000b). Direct measurement of situation awareness: Validity and use of SAGAT. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 3–32). Mahwah, NJ: Erlbaum.

Endsley, M. R., & Garland, D. J .(2000a). Pilot situation awareness training in general aviation. In *Proceedings of the 14th Triennial Congress of the International Ergonomics Association and the 44th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 357–360). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R., & Garland, D. J. (Eds.). (2000b). *Situation awareness analysis and measurement*. Mahwah, NJ: Erlbaum.

Endsley, M. R., & Jones, W. M. (2001). A model of inter- and intrateam situation awareness: Implications for design, training, and measurement. In M. McNeese, E. Salas, & M. Endsley

(Eds.), *New trends in cooperative activities: Understanding system dynamics in complex environments* (pp. 46–67). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R., & Robertson, M. M. (2000). Training for situation awareness. In M. R. Endsley & D. J. Garland (Eds.) *Situation awareness analysis and measurement* (pp. 349–365). Mahwah, NJ: Erlbaum.

Endsley, M. R., Sollenberger, R., & Stein, E. (1999). The use of predictive displays for aiding controller situation awareness. In *Proceedings of the 43rd Annual Meeting of the Human Factors and Ergonomics Society* (pp. 51–55). Santa Monica, CA: Human Factors and Ergonomics Society.

Fracker, M. L. (1991a). *Measures of situation awareness: An experimental evaluation* (Report No. AL-TR-1991-0127). Wright-Patterson Air Force Base, OH: Armstrong Laboratories.

Fracker, M. L. (1991b). *Measures of situation awareness: Review and future directions* (Report No. AL-TR-1991-0128). Wright-Patterson Air Force Base, OH: Armstrong Laboratories.

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

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics, 49,* 1312–1325.

Graham, J., Gonzalez, C., & Schneider, M. (2007). A dynamic network analysis of an organization with expertise out of context. In R. Hoffman (Ed.), *Expertise out of context* (pp. 385–402). Mahwah, NJ: Erlbaum.

Graham, S. E., & Matthews, M. D. (2000). Modeling and measuring situation awareness. In J. H. Hiller & R. L. Wampler (Eds.), *Workshop on assessing and measuring training performance effectiveness* (Tech. Rep. 1116, pp. 14–24). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). New York: North-Holland.

Hogg, D. N., Torralba, B., & Volden, F. S. (1993). *A situation awareness methodology for the evaluation of process control systems: Studies of feasibility and the implication of use* (No. 1993-03-05). Storefjell, Norway: OECD Halden Reactor Project.

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 in aviation. *Aviation, Space and Environmental Medicine, 67*, 507–512.

Jones, D. G., & Endsley, M. R. (2002). *Measurement of shared SA in teams: Initial investigation* (No. SATech-02-05). Marrietta, GA: SA Technologies.

Matthews, M. D., Pleban, R. J., Endsley, M. R., & Strater, L. G. (2000). Measures of infantry situation awareness for a virtual MOUT environment. In *Proceedings of the Human Performance, Situation Awareness and Automation: User-Centered Design for the New Millennium Conference* (pp. 262–267). Savannah, GA: SA Technologies.

Milham, L. M., Barnett, J. S., & Oser, R. L. (2000). Application of an event-based situation awareness methodology: Measuring situation awareness in an operational context. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (2.423–2.426). Santa Monica, CA: Human Factors and Ergonomics Society.

Pew, R. W. (2000). The state of situation awareness measurement: Heading toward the next century. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement*. Mahwah, NJ: Erlbaum.

Riley, J. M., & Kaber, D. B. (2001, July 28–August 2). Utility of situation awareness and attention for describing telepresence experiences in a virtual teleoperation task. In *Proceedings of the International Conference on Computer-Aided Ergonomics and Safety* (CD-ROM). Maui, Hawaii.

Rodgers, M. D., Mogford, R. H., & Strauch, B. (2000). Post hoc assessment of situation awareness in air traffic control incidents and major aircraft accidents. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 73–112). Mahwah, NJ: Erlbaum.

Rousseau, R., Tremblay, S., & Breton, R. (2004). Defining and modeling situation awareness: A critical review. In S. Banbury & S. Tremblay (Eds.), *A cognitive approach to situation awareness: Theory and application* (pp. 3–21). Burlington, VT: Ashgate.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors, 50*, 540–547.

Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors, 37*, 123–136.

Schunn, C. D., McGregor, M. U., & Saner, L. D. (2005). Expertise in ill-defined problem solving domains as effective strategy use. *Memory and Cognition, 33*. 1377–1387.

Scott, J. P. (2000). *Social network analysis: A handbook* (2nd ed.). London: Sage.

Lelyn D. Saner is an assistant research scientist at the University of Maryland Center for Advanced Study of Language. Saner's research focuses on the psychology of how representations are used for decision making and reasoning, with recent research looking at how people naturally frame decisions.

Cheryl A. Bolstad is a senior research associate with SA Technologies with over 15 years of experience as a human factors engineer. Her recent research has focused on developing methods for supporting team situation awareness in distributed systems and developing training systems for supporting situation awareness.

Cleotilde Gonzalez is an associate research professor and director of the Dynamic Decision Making Laboratory at Carnegie-Mellon University. Her research has focused on the psychology of decision making in complex, dynamic situations. Past projects include the development of ACT-R cognitive models of situation awareness and learning in command and control during the execution of a battle.

Haydee M. Cuevas is a research associate with SA Technologies. Her recent research has focused on training and system design to support team cognitive and metacognitive processes related to both human-human and human-agent group dynamics across pre-, in-, and postprocess interactions in distributed military teams.