CRM D0002722.A2/Final
November 2000

Gaming and Shared Situation Awareness

Center for Naval Analyses:
Peter P. Perla • Michael Markowitz
Albert Nofi • Christopher Weuve

ThoughtLink Incorporated:
Julia Loughran • Marcy Stahl

Center for Naval Analyses
4825 Mark Center Drive • Alexandria, Virginia 22311-1850

Approved for distribution: November 2000

Peter P. Perla
Director for Interactive Research Products
Office of the Senior Vice President for Research

This document represents the best opinion of CNA at the time of issue.
It does not necessarily represent the opinion of the Department of the Navy.

Distribution unlimited. This document is approved for public release.
For copies of this document call :CNA Document Control and Distribution Section a t703-824-2943.

Copyright © 2001 The CNA Corporation

Contents

Summary 1

Background 5

Defining and measuring shared situational awareness. 9
- What is situational awareness? 9
  - The process of situational awareness. 9
  - Situational awareness as a mental model 13
- What is shared situational awareness? 17
- Measuring situational awareness and shared situational awareness 20

The game: SCUDHunt 27

The SCUDHunt experiment 31
- Data and measures 31
- Underlying experimental model 32
- Statistical analysis 34

Ancillary data and observations. 37
- Ancillary data 37
- Observations 38

Discussion and conclusions 41

Appendix A: SCUDHunt 45
- Objective 45
- Components 45
- Background. 46
- Sequence of play 47
- Player assets. 48
- Player roles 49
- Search results 50
- Web implementation. 50

Appendix B: The experimental model 57

Appendix C: Questionnaires 61
Pre-game questionnaire 61
Post-game questionnaire 64

Summary

The Defense Advanced Research Projects Agency (DARPA) is interested in exploring key factors that affect how teams, particularly distributed teams, develop what is called *shared situational awareness* in an operational environment. The DARPA Program Manager for the *Wargaming the Asymmetric Environment* program asked the Center for Naval Analyses to address these issues, with subcontracting support from ThoughtLink Incorporated. The focus of the project was to demonstrate how wargaming could be used as a testbed for conducting experiments to explore these key factors in team shared situational awareness.

Our survey of the state of the art of research about situational awareness and shared situational awareness indicated some fragmentation of thinking; however, that thinking appeared to be beginning to coalesce into some weak agreement on fundamental concepts. Key among these is the notion that situational awareness is best defined as a dynamic mental model of our operating environment and our place in it. We build this model through a process we call situation assessment, which consists of four interwoven subprocesses: perception, comprehension, projection, and prediction. Similarly, for our purposes, we define *shared situational awareness* of a team as the overlap in the situational awareness of team members.

But how do we measure situational awareness, not to say shared situational awareness? We found no evidence of a clearly understandable, generally accepted, objective way to measure situational awareness and how it develops, whether for individuals or for groups. As a result, we chose to develop a fresh approach. Our goal was to devise an experimental regime that would allow us to define both an individual's situational awareness and that of a team in clear and understandable terms. In addition, we wanted to minimize the dangers that our attempts to measure situational awareness would affect the behavior of the test population.

Our approach was to develop an experiment in which multiple teams played an online game. We designed the game so that:

- Team members had to share information to do well
- Their decisions could be directly and easily measured and recorded
- The measure of their decisions would describe the degree of shared situational awareness of the team.

We focused on developing a sample testbed and using that testbed in a limited-objective experiment to demonstrate the applicability and feasibility of this technique. The game we developed was SCUDHunt. The experiment we designed used SCUDHunt to explore the effects of communications and shared visualization tools on the ability of a distributed team of players to develop a shared mental model or operational picture of a simple operating environment. This operational environment consisted of a 5 x 5 grid of featureless squares on which 3 SCUD launchers were concealed.

Each player controlled one or two reconnaissance assets which could provide information about whether a particular square might contain a launcher. Players shared information about the search capabilities of their assets, the results of their searches, and their future search plans and attempted to identify and recommend a number of target squares they believed were their best estimates for containing launchers. We measured the efforts of the players to build their shared picture based on these recommendations and how well the recommendations overlapped among all the players on the team.

The overall experiment was designed in the form of a 6 x 6 Latin Square, in which each team played a well-defined series of 6 games, defined by combinations of 3 different types of communications capabilities crossed with two types of shared visualization tools. There were six four-person teams. Each team played the games in a different order so that we could control for possible learning effects.

The results of the experiment indicated that both communications and shared visualization played statistically significant roles in the ability of our teams to develop shared situational awareness. Perhaps

surprisingly, however, there appeared to be no significant difference between the effects of voice communications and that of real-time text chat.

The broader results of this effort show that the use of simple games, designed to target specific experimental goals, is a promising technique for conducting research in this field. We collected significant amounts of data that we had neither the time nor the resources to analyze during this effort. Follow-on research to analyze this data may prove useful. In addition, we could apply the SCUDHunt game and experimental technique in support of the research directions taken by several leading practitioners in the field of situational awareness. Finally, a new research approach that combines the game-based environment of SCUDHunt with the burgeoning popularity and feasibility of agent-based computer simulation holds out the promise of exciting new techniques to supplement and help direct human-based testing in this field.

Background

At the start of the new century, the U.S. military increasingly finds itself operating in complex contingencies and as part of heterogeneous teams involving military, interagency, non-governmental organizations (NGOs), international organizations (IOs), and other coalition team members. At the same time, military doctrine and organization for command and control is becoming increasingly "network-centric"—meaning that resources such as information, people, and communication methods and tools are highly distributed—and procedures are moving more and more toward self-organization (flattened hierarchies) and self-synchronization (act, not ask).

One important element of this changing environment is the proliferation of *virtual teams*, that is, teams composed of individuals from distinct organizations and physically separated from each other. These teams, also called *distributed teams*, work together by using computer networks and other communications and information-sharing technologies. We see such distributed teams everywhere—in training, operations, and research environments.

As the United States faces new threats and new situations, understanding how our own distributed teams can improve their shared situational awareness (SSA) will help us create better tools for operating in the new asymmetric environment. The same research can also shed light on how we can degrade the SSA of our adversaries and so magnify our own effectiveness.

There is an extensive body of research in the field of situational awareness (SA) in general, as well as the subset of issues associated with SSA of teams. We summarize much of this research in the next section and deal with it more extensively in a separate paper.[1]

---

1. Albert A. Nofi, *Defining and Measuring Shared Situational Awareness*, CNA Research Memorandum D0002895.A1 (November 2000).

A great deal of the research in this field centers around observational studies of limited scope and scale, frequently conducted in real-world operational environments.[2] Such studies are difficult or impossible to repeat under controlled experimental conditions.

Some experimenters have conducted studies that make use of gaming environments as a substitute for real ones. Many of these studies have used complex games that attempt to simulate real-world environments with some degree of fidelity.[3] Their emphasis on fidelity can make experiments based on such games difficult to control or to focus on specific research topics.

Other researchers have employed simplified games to explore specific, tightly defined issues. In many cases, however, even those researchers have defined their measures in terms of how well a team performs an operational mission requiring SSA, rather than on the process and results of building that SSA, or a shared picture, itself.[4]

Our approach centers on the use of a simplified, though not quite abstract, game that allows us to tailor its design and mode of play to focus on the specific research items of interest. In the case of SSA, we designed the game so that the bulk of the operational task faced by the players lies precisely in building a shared picture—their shared SA, if you will—of their operating area. This approach removes much of the potential confounding between SSA and game-playing skill, a problem that can be associated with measuring a team's performance in a game primarily by measuring its success in performing a specific operational game task (such as "winning" the game).

---

2. See, for example, Gary Klein, J. Orasanu, R. Claderwood, and C. Zsambok, *Decision-Making in Action: Models and Methods.* Alex Publishing Corporation, Norwood, NJ, 1993.

3. See, for example, Diane L. Rulke and Joseph Galaskiewicz, "Distribution of Knowledge, Group Network Structure, and Group Performance," *Management Science*, Vol. 46, No. 5, May 2000, pp. 612-625.

4. A good example of this is Cheryl A. Bolstad and Mica R. Endsley, "Shared Mental Models and Shared Displays: An Empirical Evaluation of Team Performance," *Proceedings of the 43rd Meeting of the Human Factors and Ergonomics Society,* 1999.

By using this approach, we were able to develop a game and an experimental regime based on a statistical experimental design. This design allowed us to test rigorously the sorts of hypotheses of interest. We collected and analyzed our data using standard statistical techniques.

In the remainder of the paper, we summarize our survey of SA and SSA research, describe the game we used as our testbed, and outline our experiment and its results. We conclude by discussing what we learned and speculating on where our research could lead in the future.

Defining and measuring shared situational awareness[5]

What is situational awareness?

Situational awareness (sometimes also called *situation awareness*) is one of those concepts that everyone (supposedly) understands but that has successfully resisted a precise definition we all share. In common usage, we frequently understand the concept as a sense of knowing what's going on in our current environment, what could happen next, what options we have for action, and what the possible outcomes of those actions might be. In military usage, the concept apparently originated in the aviation community, in which a pilot's situational awareness reflects his ability to know and understand what's going on around him, whether in the cockpit, on the ground, or in the air. This concept has been a major driver in the development of systems and procedures to maintain and increase the safely of flight.

To answer the question above, we must first address a basic issue frequently lost in the shuffle of intellectual debate: Is situational awareness a process or a state?

The process of situational awareness

Most of the '"definitions" of situational awareness we found in our research focus on what amounts to a process. For our purposes, the best description of the processes involved in situational awareness that we have seen is the one proposed by Mica Endsley in a 1995

5. This section is largely derived from Albert A. Nofi, *Defining and Measuring Shared Situational Awareness*, CNA Research Memorandum D0002895.A1 (November 2000).

paper published in the journal *Human Factors*.[6] Endsley defines situation awareness as

> the perception of the elements in the environment within a volume of space and time, the comprehension of their meaning, the projection of their status into the near future, and the prediction of how various actions will affect the fulfillment of one's goals.

A "competing" definition is espoused by Barry McGuinness, a psychologist at the Sowerby Research Centre of British Aerospace in an internet discussion board dedicated to the topic of situational awareness.[7] McGuinness began his description of situational awareness with a quote from a pilot: situational awareness is "knowing what's going on so you can figure out what to do." McGuinness elaborated as follows:

> This says it all. If you have a function to perform in a situation that is fairly complex and dynamic, such that you have to make decisions, then you have to be aware of what is going on—at different levels—if you are to make the right decisions to achieve your goals.
>
> We can unpack this further: To be aware of what's going on you have to
>
> (a) take in the available facts, and
>
> (b) understand them in relation to your expert knowledge of such situations. Furthermore, to make the best decisions you have to
>
> (c) anticipate/predict how the situation is likely to develop in future and
>
> (d) understand your options and courses of action relative to your goals.

---

6. Mica Endsley, "Toward a Theory of Situation Awareness in Dynamic Systems," *Human Factors*, 37 (1995), pp. 32–64.

7. Barry McGuinness, SABRE, December 8, 1999, http://users.ox.ac.uk/~pemb0595/wwwboard/44-sa.html. The extract has been edited slightly for clarity, but with emphasis as in the original.

(e) Finally, you also need to take into account how accurate / complete / reliable your SA is likely to be.

So you can identify these different aspects to SA. Endsley calls the first three PERCEPTION, COMPREHENSION and PROJECTION. I would say you have to include the last two as well, and call them INTENTION & METACOGNITION respectively.

Each of these factors is associated with specific cognitive processes and with particular contents (mental representations). For instance, at the "level" of perception you're monitoring, attending, detecting and identifying (processes), which gives you factual awareness of current objects, events, states and so on (contents).

It is not clear why McGuinness chose to replace Endsley's fourth element, prediction, with the concepts of intention and metacognition. Consider the differences. Intention seems less an aspect of the process by which situational awareness is developed than a necessary precondition for developing it. Situational awareness is essentially goal oriented; our reason for being in a particular "situation" explains why we need to develop the "awareness" necessary to cope with it. Likewise, metacognition would seem to be more appropriately a combination of projection and prediction.

In the end, we chose to adopt Endsley's definition for our purposes, and so concluded that the critical factors in the development of situational awareness are:

1. Perception—acquiring the available facts
2. Comprehension—understanding the facts in relation to our own knowledge of such situations
3. Projection—envisioning how the situation is likely to develop in the future, provided it is not acted upon by any outside force
4. Prediction—evaluating how outside forces may act upon the situation to affect our projections.

McGuinness makes an important point, however, when he observes that we should not envision the elements of situational awareness

> as lying in a chain or sequence, but as interlocking cycles. None of them comes first. E.g., we don't suddenly enter a situation and gradually pick up raw information; we always have an ongoing action schema.[8]

This is an important point. In developing situational awareness, we don't necessarily follow the neat flow from perception, through comprehension, then projection, and finally prediction. These stages occur virtually simultaneously, given the speed with which our minds work. As we perceive the information, we are already processing it for comprehension and its implications for our purposes. In short, these stages of awareness form a dynamic tapestry of interwoven threads rather than a static sequence followed like a flow chart. Moreover, this process goes on continuously, so that our situational awareness evolves continuously as well.

Even if we accept the value of Endsley's (or McGuinness's) "definitions," can we really accept their processes as a definition of situational awareness itself? It seems more reasonable to think of situational awareness as a state, a state that is the product of a process that involves the matrix of perception, comprehension, projection, and prediction. Indeed, Mica Endsley suggests that the process by which situation awareness develops might best be termed "situation assessment."[9] It is this process of situation assessment that Endsley and McGuinness have defined, not its product, situational awareness.

If that is the case, then how can we characterize situational awareness itself?

---

8. Barry McGuinness post at http://users.ox.ac.uk/~pemb0595/www-board/64-sa.html on February 24, 2000 at 16:49:32. McGuinness, of course, is actually speaking of the five factors that he has identified, rather than Endlsey's four factors, but the principle would remain unchanged in either case.

9. Endsley, "Toward a Theory . . . ," p. 36.

Situational awareness as a mental model

Clearly, what constitutes situational awareness is a function of your "situation" and how it develops. In a wartime environment, for example, the things that are important to a soldier dodging bullets in the front lines are different from those that are important for a soldier shuffling papers in the rear. We could develop contrasting models of what matters most in the situational awareness of each of the two soldiers, given their differing environments. Indeed, the concept of *mental models* would seem a natural way to think about the nature of situational awareness.

A mental model is a "psychological representation of the environment and its expected behavior." The purpose of a mental model is "to provide conceptual framework for describing, explaining, and predicting future system states."[10]

Barry McGuinness argues that situational awareness postulates the construction of precisely such a mental model:

> In the context of SA, it is usually assumed that the human operator working in a complex, dynamic task environment must construct and maintain a mental model of 'the situation'. So this is in effect an organized set of 'working hypotheses', rather than a stored-away file.
>
> The two types interact, however: When I 'understand' the present situation, it is because all the details fall into a familiar pattern that corresponds to a generalized model I have learned; but when something new or odd occurs I have to generate hypotheses about it which I can hopefully test out with further information.
>
> So the contents of a person's SA at any one moment can be thought of as a set of
>
> 1. references to confirmed schemas (e.g., "landing phase"),

10. Richard Klimoski and Susan Mohammed, "Team Mental Model: Construct or Metaphor?," *Journal of Management*, Vol. 20, No. 2 (1994), p. 405.

2. yet-to-be-tested hypotheses ("I expect to see runway once below cloud"), and

3. specific details of significance, like so-and-so's call sign.[11]

The characterization of situational awareness as a mental model seems to underlie most military usage of the term. Thus, we find, with reference to the functioning of an AEGIS Command Direction Center,

> Each officer has a mental model of the position he is manning, its responsibilities and requirements, and he also maintains a highly dynamic mental model of the current tactical situation. The tactical situation mental model may include a model for the tactical situation of his own ship, as well as a separate, but related model for the entire ship group.[12]

Used in this context, situational awareness is readily seen as the mental model developed by the operator, in some ways embodying some of the characteristics of a "working hypothesis." McGuinness goes so far as to say that "there is no 'situation' in the real environment; rather, we experience being in a situation whenever our goals and actions are being affected, and the impact of the environment becomes a part of it."[13] So our situational awareness changes as our environment changes. It's a subjective condition, a "dynamic mental model," as it were.[14]

Individual situational awareness is very much a personal attribute. We see the world around us in individual terms, based on our cultural background, education, and experiences, not to mention the strengths and limitations of our senses. The mental model we

---

11. Barry McGuinness, SABRE, January 17, 2000, http://users.ox.ac.uk/~pemb0595/wwwboard/56-sa.html.

12. Stottler Henke Associates, "Intercepting Missile Control Adaptive Training System," http://www.shai.com/projects/missile_control.htm.

13. Barry McGuinness, SABRE, February 24, 2000, http://users.ox.ac.uk/~pemb0595/wwwboard/64-sa.html.

14. Barry McGuinness, SABRE, November 10, 1998, http://users.ox.ac.uk/~pemb0595/wwwboard/29-sa.html.

develop as a result of these inputs is essentially self-centered, as it necessarily must be, given that the self is the prime referent.

Factors affecting an individual's situational awareness would seem to include training, experience, personality, interests, and skill level. Each of these factors plays its role within the framework of the mission that one is performing and the environment in which we are operating.

Individual talents and personality traits are also important elements in the development of situational awareness. For example, some people are better observers than others, a trait that would certainly influence their situational awareness. In this regard, consider people who fall into the Myers-Briggs personality classes of *Sensors* and *Intuitives*. *Sensors* like to operate in an exact and systematic fashion, preferring to focus on facts and details. *Intuitives* tend to look for possibilities, meanings, and relationships, and try to take a holistic look at problems.[15]

In addition to such long-term, or *structural*, factors as personality, culture, training and education, experience, and others, the process of building situational awareness can be influenced strongly by transitory, or *situational*, factors. The latter includes things such as mood, fatigue, stress, time pressure, and the complexity and ambiguity of the situation.

It seems reasonable to conclude that different people will develop different situational awareness—in terms of their personal, dynamic, mental models—even when in similar situations. The situational awareness of a veteran soldier under fire will probably be much different from that of a raw recruit in the same circumstances.

In research focused on identifying characteristics important for systems designed to help produce what was called a "coherent tactical

15. Otto Kroeger and Janet M. Thoresen, *Type Talk: The 16 Personality Types That Determine How We Live, Love, and Work* (New York: Dell, 1988), pp. 24-25.

picture," CNA analysts Allen T. Hjelmfelt and Marvin A. Pokrant conclude that situational awareness,

> 1. . . . is a volatile mental state. In most cases SA is built up over time. That is, it requires a knowledge of the current state of the environment and at least some past history. Because SA is strongly time-dependent, without continual refreshing it decays as the environment changes.
>
> 2. . . . refers to one's perception of the dynamic state of the environment. It does not directly refer to static factors such as the knowledge of established procedures and doctrine. Nor does it refer to an individual's skills. It is assumed that the individual is competent in all of these.
>
> 3. . . . does not encompass awareness of all available information. The basic information elements required for SA are bounded by time, space, and the individual's goals.
>
> 4. . . . requires the ability to predict how the situation will change due to one's action or lack of actions.
>
> 5. . . . is goal-oriented and not task-oriented. Tasks are performed through a mission to accomplish goals, but the goals remain relatively constant for the duration of the mission. Based on SA, one makes decisions to do certain tasks to accomplish the high-level goals.[16]

All the principal researchers imply that situational awareness is not a stable phenomenon, but rather a dynamic state. The "situation" changes even as we experience it. And our awareness of the situation will necessarily change as well. So, although we speak of the "state" of someone's situational awareness, we are really speaking about a transitory phenomenon. Our situational awareness is constantly evolving and changing, depending on how the situation itself evolves and on how (and how well) we integrate the available information in an ongoing process of perception, comprehension, projection, and prediction.

---

16. Allen T. Hjelmfelt and Marvin A. Pokrant, *Coherent Tactical Picture,* CNA RM 97-129 (March 1998).

Summarizing, then, we characterize situational awareness as a dynamic mental model of our operating environment and our place in it. We build this model through a process we call situation assessment, which consists of four interwoven subprocesses: perception, comprehension, projection, and prediction. The resulting mental model is inherently subjective, based on integrating acquired information with our own personal structural and situational factors. The quality of our situational awareness may be characterized by the degree to which our mental model—our situational awareness—"accurately" reflects objective reality. Finding ways to measure the "goodness of fit" of our situational awareness is not an easy task. We will discuss this problem from a broader perspective later.

What is shared situational awareness?

With all the imprecision and debate surrounding the basic meaning of the idea of situational awareness, it is hardly surprising that the broader concept of *shared situational awareness* suffers from similar conceptual and semantic difficulties. At times, it's not clear whether the term "shared situational awareness" is being used in the sense of "awareness of a shared situation" or in the sense of "shared awareness of a particular situation." There's a considerable difference.

Used in the sense of "awareness of a shared situation," shared situational awareness implies that we understand that we are in a common, or shared, situation. In contrast, when used in the sense of "shared awareness of a situation," shared situational awareness implies that we all understand a given situation in the same way. The latter is the sense in which the phrase is used most often by the aviation and military communities.

Group or shared situational awareness obviously differs from individual situational awareness because it involves a collection of people, rather than just one individual. In normal circumstances, in any given situation or environment, all of these people would have their own unique, individual situational awareness. But how then can it be shared? To get the members of the group to develop a shared awareness of the situation requires that they

- Build individual situational awareness
- Share their individual situational awareness
- Develop the group's shared situational awareness.

Of these three functions, the first, building an individual awareness of the situation, would seem to be the simplest to attain. After all, we do it all the time.

The second element of the process is essentially a matter of effectively communicating each person's perception of the situation to the other members of the group, so that a "consensus flow" develops. This is probably the most critical issue in creating "shared awareness." It is the essential factor in integrating the individual mental models into a shared model.

However, if we are to share individual situational awareness for the purpose of building a common situational awareness, we will certainly have to establish some *common ground*. In this usage, the term *common ground* is defined as shared knowledge, beliefs, and assumptions, which provide a foundation for effective communication.

> Common ground does not mean strong unification; it does not imply that everyone has the same goals, shares the same view of the world, and acts the same way. A common ground allows for certain diversity and individuality, enables shared views and vocabularies, and tolerates sub-communities, sub-disciplines, and the like. However, there is always a core of common concepts and views.
>
> The common ground is dynamic in nature and therefore is often a matter of explicit negotiation and communication. A common ground can fall apart and eventually can get lost; hence, it needs constant maintenance in order to keep the community, culture, and discipline alive.[17]

Common ground is thus a critical element in building a shared situational awareness. Typically, building this common ground will require some familiarity among team members, a familiarity based on

17. http://www.acm.org/sigchi/chi96/CommonGround.html.

common cultural backgrounds or experiences or developed through training and education.[18]

Another key element in building a shared situational assessment is the communications environment. In a collocated environment team members can talk about the situation and their understanding of it, and can pass information back and forth directly through a variety of printed and electronic media. Understanding—or the lack of same—can be communicated by facial expression, gestures, and body language, tone and hesitation, and the many other cues that regularly supplement conversation.

In a distributed environment, many of these cues and other sources of information and understanding are lost or transformed. Communication by electronic means eliminates much of the context that affects collocated communication; body language, tone, hesitancy, and all the other social conventions that facilitate face-to-face communication. Their absence necessarily makes communication in a distributed team more difficult.

The third element in building shared situational awareness is the integration of the different individual mental models of the situation. Presumably the effectiveness of the group effort will be shaped by the degree to which the members develop a common understanding and common commitment, a "consensus flow" that will lead to a common picture.[19] However, it seems unreasonable to assume that there will be a single "team mental model." As Klimoski and Mohammed observe, "There can be. . . multiple mental models co-existing among team members at a given point. . . ." "These models need not be identical, but they do have to overlap sufficiently to make it possible to perform the mission. Thus, for our purposes we will define *shared situational awareness* as this overlap in the situational awareness of team members.

---

18. For a discussion of some basic ideas in this area, see Helen Altman Klein, Anna Pongonis, and Gary Klein, *Cultural Barriers to Multinational C2 Decision Making*, Klein Associates, 2000.

19. Klimoski and Mohammed, p. 410.

Measuring situational awareness and shared situational awareness

So we have defined situational awareness and shared situational awareness. What, after all, is the point? For our purposes, the point is a practical one: making decisions in complex operational environments rests on a foundation of situational awareness—whether in fast-moving environment of aerial combat or at the snail's- pace (at least comparatively speaking) of coalition consensus building. If we can understand how decision-makers develop their situational awareness and find techniques and technologies to improve our own situational awareness while degrading that of our opponents, then we may be able to get the edge we need to achieve our objectives in minimum time, with minimum commitment of resources and minimum loss of life.

To do that, however, we need to be able to assess the quality of a person's or a group's situational awareness. We also need to be able to measure how different processes, procedures, and technologies may improve or degrade that situational awareness. Ideally, we would like to be able conduct controlled experiments to provide a solid scientific basis for our conclusions. This is difficult.

Most of the various experimental methods that have been proposed to measure SA fall into three general categories, each of which has many variations, not all of which are always applicable to every situation.

1. Subjective. The subject rates his own SA, either by merely being asked to evaluate it or through responses to directed questions.

2. Implicit Performance. This presumes that a subject's performance correlates with SA, on the assumption that improved SA will lead to improved performance.

3. Explicit Performance. Researchers engage in an ongoing effort to "directly probe the subject's SA by asking ques-

> tions designed for that purpose," by suspending the activity being studied for short periods.[20]

All three approaches attempt to quantify certain human behaviors that are difficult (at best) to quantify. They are also all inherently subjective, and are likely to run afoul of the problems that may arise because the very act of attempting to measure some of the important behaviors can sometimes alter fundamentally how the subject will behave, thus making the measurement itself unreliable.

Existing research argues strongly that situational awareness and shared situational awareness are inherently subjective. Because our mental models are representations of a real (however elusive) operational environment, however, they may be subject to a valuation of their "quality," the degree to which they accurately reflect an objective assessment of that reality. Unfortunately, making such assessments in any meaningful way is almost as elusive as defining reality.

One of the reasons for this difficulty is that the elements and nature of situational awareness can differ dramatically—dare I say it?—depending on the situation. My mental model of the situation when I am playing basketball is significantly different from my model while driving my car. Parts of my models are objectively knowable—for example, the positions to the other players on the court or the other vehicles within my range of vision. Parts of my models are objectively testable—for example, if I believe the tractor trailer in front of me is about to turn to his left, that driver's short-term actions will reveal whether, in fact, my model was "accurate."

On the other side of the coin, I am constantly assessing my own model from the subjective viewpoint. Do I have enough information to feel comfortable with my understanding of the situation? Have I encountered a similar situation before? Do I know what I'm doing?

Barry McGuinness proposed a model of how to think about this crossing of objective and subjective measures of situational awareness.

---

20. Hjelmfelt and Pokrant, pp. 15-16.

I have a model of SA measurement which grossly differentiates between SA as revealed by objective queries (e.g., which target is highest priority?) and SA as revealed by subjective measures (e.g., rate your SA on a scale of 1-7). There can be real differences between what an operator "knows" (having assimilated or inferred the relevant items of information) and what he *thinks* he knows overall. Imagine a 2 x 2 square grid. The area to the left of centre = "objective SA good, all relevant info present and correct"; the area to the right of centre = "objective SA impaired, some info missing". The area above the centre = "subjective SA high, subject is confident in info", the area below the centre = "subjective SA low, operator is uncertain."

Measuring Situational Awareness

	Objective Assessment: High	Objective Assessment: Low
Subjective Assessment: High	Objective High: All relevant info present and correct Subjective High: Subject is confident in info	Objective Low: Some info missing Subjective High: Subject is confident in info
Subjective Assessment: Low	Objective High: All relevant info present and correct Subjective Low: Operator is uncertain	Objective Low: Some info missing Subjective Low: Operator is uncertain

Now, the ideal place for your operator to be is the top left square, where (a) he has the right info and (b) he knows it. The worst place to be is top right, where he (a) has lost info but (b) doesn't realise it. If he's in bottom left, he has the right info but doesn't feel sure about it (maybe he's a novice). If he's in bottom right, he's lost info but at least he feels his SA isn't right. Presumably, any operator in the lower two

> boxes is going to go straight into active information-seeking mode (you can see it happen!).[21]

One problem with using such a model in practice, of course, is gathering enough of the right kind of data to place a particular person's situational awareness in the matrix. McGuinness described one possible approach:

> First, as ever I'd recommend approaching the problem from both 'objective' and 'subjective' angles; in other words, a combination of actual knowledge queries and subjective ratings of SA. SA is optimal when objective content is high AND the operator(s) is/are subjectively aware of that. SA is disastrous when the content is off and the operator doesn't realize it. There are other combinations, of course, so it's always best in my book to measure the two together if possible.
>
> For subjective measurement, you could get each operator to rate the quality of SA of both himself and the team as a whole.
>
> I've developed a scale (CARS - crew awareness rating scale) which focuses on 4 dimensions:
>
> 1) awareness of perceptible data
>
> 2) awareness of the big picture
>
> 3) awareness of future developments
>
> 4) awareness of response options.
>
> This was designed originally for the single subject, but can be used team-wise, I think. Also, the scale can be used as a general SA metric or can be addressed to specific tasks of interest, e.g., flight awareness vs. systems awareness.
>
> For objective assessment, our preferred method at present is to identify key items of information pre-trial, then present sets of queries at intervals; somewhat like SAGAT's approach, but different. We use a "situational report"

21. Barry McGuinness, SABRE, October 5, 1998, http://users.ox.ac.uk/~pemb0595/wwwboard/23-sa.html.

(sit-rep) format, which many operators are used to. The subject is given a sit-rep form to complete (in mid-run). For each item, he can either give the requested information (e.g., expected time of arrival) as he is aware of it, or tick a box:

[ ] Information not available

[ ] Information not relevant

[ ] Information not my responsibility[22]

The intrusiveness of this method of collecting data is a good example of the concerns expressed above. Moreover, McGuinness' criteria are not likely to indicate whether there's sufficient information to perform the task or whether it's being performed. Although whether there is enough information available to perform the task is probably measurable using some objective criteria, both of McGuinness's scales essentially focus on the operator's perceptions of the state of his situational awareness. What happens if his confidence in his situational awareness is wrong? Suppose he thinks he's in the upper left quadrant, when in fact he should be in the lower right one? As the saying goes, “It’s not what you don't know that gets you, it's what you know that ain't so.”

The question of how to measure situational awareness becomes even more complex when we consider it in terms of a group or team. Not only do we have to consider measuring the situational awareness of each individual member of the team, but also we must consider how each team member interacts with all the others to form their shared awareness. The problem increases combinatorically as we add more and more members to the team.

So where does that leave us?

---

22. Barry McGuinness, SABRE, November 10, 1998, http://users.ox.ac.uk/~pemb0595/wwwboard/29-sa.html. With regard to the “situational report,” he adds the telling comment, “The last response is particularly interesting in terms of team SA. What if all operators in the team tick it for a particular item!”

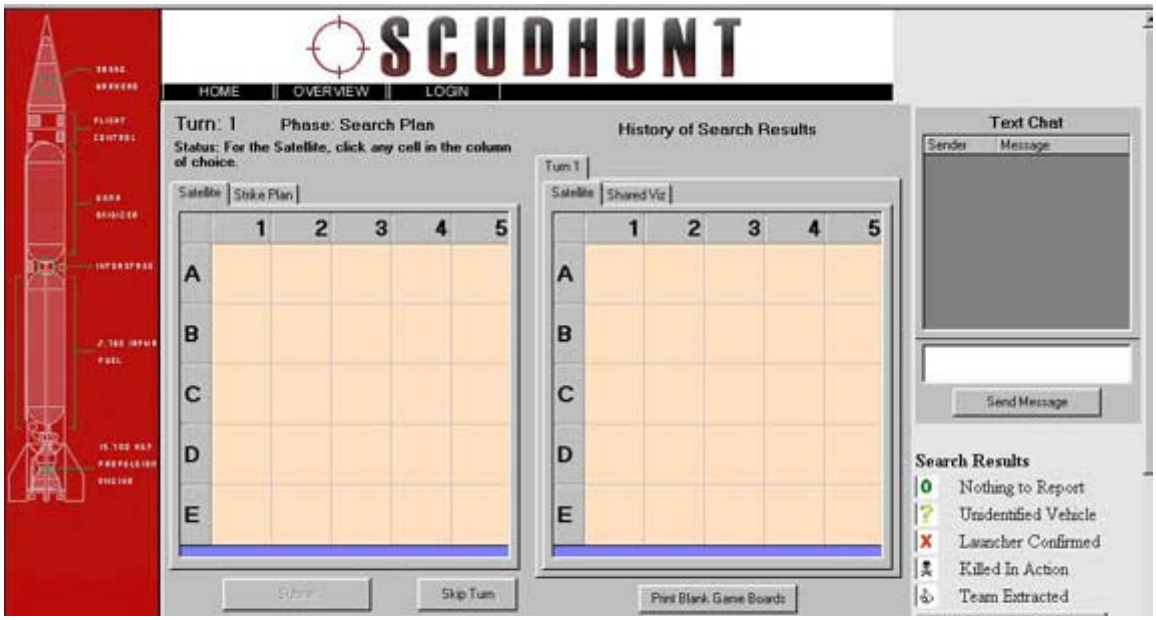
At present, we can find no evidence of a clearly understandable, generally accepted, objective way to measure situational awareness and how it develops, whether for individuals or for groups. As a result, we chose to develop a fresh approach. Our goal was to devise an experimental regime that would allow us to define both an individual's situational awareness and that of a team in clear and understandable terms. In addition, we wanted to minimize the danger that our attempts to measure situational awareness would affect the behavior of the test population.

We quickly focused on the idea of using an abstract game as our test bed, rather than attempting to "instrument" a real-world operation or even a complex simulation of such an operation. The simpler gaming environment allowed us to control how we would define the elements of the situation and how we would measure the situational awareness—and shared situational awareness—of our test population.

The game: SCUDHunt

The first step was to define an environment in which the "situation" could be simplified without making it meaningless. Most previous attempts at using gaming environments in this field employed games in which building situational awareness was only part of the requirement to playing the game.[23] Such an arrangement makes it difficult or impossible to interpret outcomes. Was one player or team more successful than another because they had better situational awareness or shared situational awareness, or simply because they were better at playing the game? Our approach tries to minimize this problem by making situational awareness itself the object of the game.

Simply put, the players must find some number of objects hidden on a grid. Their situation can be defined as the location of the objects. Their situational awareness is then defined as where they believe (or guess) the objects to be.



23. See Bolstad and Endsley, 1999.

The measure of the quality of the team's situational awareness is simply how many of the objects they locate correctly. The process the players use to locate the objects requires them to deploy sensors of various types and various capabilities. Each player controls a unique subset of the sensors available. The players must cooperate, exchanging information about the quality of their sensors, where they deploy them, and what results they obtain in the search.

To provide an operational back story to interest and motivate the players, we cast the game in the light of a joint or combined operation to locate SCUD missile launchers in an unfriendly state which was threatening to attack a neighboring, friendly state, but prior to the outbreak of any open hostilities. At the end of each turn of the game (nominally representing one day of real time), the players would be asked to evaluate their information and provide recommendations for which target squares they believed were most likely to contain the SCUDs. Thus, their recommendations would represent their individual situational awareness. The degree to which the recommendations of the team members overlapped would be the measure of their shared situational awareness. See Appendix A for more details on the game.

As designed and tailored for this experiment, SCUDHunt is played by a single team of four players. The positions of the target SCUDs is determined at random before the game begins, and, once placed, they do not move. The goal of the game is for team members to work together, sharing information, and then correctly decide where they should plan a strike for the SCUDs. There are three such targets deployed on a board composed of a 5 x 5 grid.

The information the team members share is based on search results of the different sensors the team members control. These sensors include: a reconnaissance satellite, manned reconnaissance aircraft, reconnaissance-equipped unmanned air vehicles (UAVs), human intelligence (HUMINT, or a "spy"), Navy Seals, Joint Special Operations teams, and communications intelligence resources (COMINT). Each of these sensors has different capabilities for searching the grid, characterized by their availability, and the probabilities of returning information when they search a square. We provide the players a

detailed description of their systems and their availability. We provide them only a verbal assessment of the ability of their sensors to make an accurate search of a square and of the potential vulnerability of their assets to destruction by the enemy forces.

Each team member controls one or more sensors, and no team members share their sensors. Players allocate or move their sensors (depending on the capability of each) to different squares on the game board. For each square successfully searched (that is, in which the sensor survives any possible threats to it), the sensor's owner (and only that player) receives information about the results of the search. Players may then share this information with their team members, taking into consideration the accuracy of their sensor device. The types of search results that may occur include: Nothing Significant to Report, Possible Target (some type of vehicle detected), and Probable Target (launchers detected). Note that it is possible for all of these results to be false. (That is, a Probable Target may be indicated when no target is actually presen, as well as Nothing Significant to Report when a target is, indeed, in the square.) The frequency with which these results occur is based on the reliability of each specific sensor as defined by its result probabilities.

SCUDHunt is a fairly simple game. Its simplicity makes it possible to monitor the game and gather information about the situational awareness of the individual team members as well as the team's shared situational awareness. The next section describes our experimental construct in more detail.

The SCUDHunt experiment

We designed an experimental environment for using SCUDHunt to demonstrate its potential—and the potential for the general approach it embodies—to develop experimental data useful for exploring situational awareness and shared situational awareness. The experimental environment is web based; each game of SCUD-Hunt in this experiment is played by a distributed team using a version of the game designed for cooperative play over the Internet.

We defined the objectives of the experiment (beyond its value as a demonstration) in terms of measuring the effects of various types of collaboration tools on the ability of the players to build their shared situational awareness.

Data and measures

As our primary data for analysis, we collect the situational awareness of each individual team member (defined as their mental model of the location of the SCUDs) by asking the team members to provide individual recommended target locations (strike plans). We ask them to submit the fewest number of possible target squares they are comfortable with as possibly containing a valid target. We then compare these individual lists and calculate a measure of overlap. This calculation is the ratio of the total number of target squares designated by the players in the team, divided by the total number of distinct squares designated. For example, if each of the 4 players designates 3 squares as their recommended targets, the total number of squares is 12. Suppose those players choose the following target squares:

Player 1: A1, A2, A3

Player 2: A1, B1, B2

Player 3: A1, B2, B5

Player 4: A1, A3, B4.

Of the total of 12 squares, 8 of them are unique (for example, square A1 is counted only once even though it appears on all four target lists).

This team's score for that turn would thus be 12/8 = 1.5. Using this measure, a perfect shared situational awareness would equal the number of players, or 4 for the example shown above. The poorest score would have no shared targets, and there would be as many unique squares as there are total target squares. The resulting score would be 1. Thus, our measure of shared situational awareness will range between 1 and the number of players involved. This measure of the team's shared situational awareness not only captures their "mental model" of where they believe the targets to be, but also has the added value of including their various degrees of uncertainty—team members who are less certain about where the targets may be are likely to include more squares in their target recommendations.

At the end of each turn, we calculate the score the team achieves that turn, but we do not reveal that score to the players. At the end of five turns, we calculate the score as usual, but this time we tell the players the result in terms of how well they did identifying the locations of the SCUDs. The game ends at that point. As a final measure of the team's overall shared situational awareness for that game, we average the results of each turn.

Underlying experimental model

For the purposes of this experiment, we focused our interest on the effects of collaboration tools. We considered two types of tools: communications and shared visualization.

We defined three levels of communications: no communications among the players at all (essentially having each play a solitaire game controlling only his own sensor); communications via internet text chat; and communications via voice in a telephone conference call.

In addition, we defined two levels of shared visualization: no shared visualization, and a shared visualization tool that consisted of a

picture of the grid with all search results displayed at the end of each turn. The source of the individual search results was not provided, so the players still had to communicate that information among themselves.

The combination of three levels of communications and two of shared visualization gave us a total of six *factorial treatments*. Because we were concerned with the possibility of learning effects as players became more familiar with the game, as well as with inherent differences among the teams themselves, a Latin Square experimental design seemed most appropriate. Such a design allows us to distinguish the effects of the differences among teams and the effects associated with the order of play of games using the different treatment combinations from the effects of the treatments themselves. In our case, because we had six treatments (combinations of communications and shared visualization), we would need to use six teams. Each of the teams would play the game using a different sequence of treatment types. The order in which each team would play these different types is arranged so that when laid out in the form of a matrix, with teams representing rows and the order of play columns, each treatment would appear exactly once in each row and each column.

Design matrix

	Game 1	Game 2	Game 3	Game 4	Game 5	Game 6
Team 1	B	E	A	C	F	D
Team 2	**D**	**A**	**E**	**B**	**C**	**F**
Team 3	**E**	**B**	**C**	**F**	**D**	**A**
Team 4	A	F	D	E	B	C
Team 5	**F**	**C**	B	D	A	E
Team 6	C	D	F	A	E	B

A = No Comms, No shared vis
B = No Comms, Shared vis
C = Text chat, No shared vis
D = Text chat, Shared vis
E = Voice, No shared vis
F = Voice, Shared vis

Statistical analysis

This Latin Square experimental design with factorial treatments is a standard linear statistical model, and its analysis is well known. Our primary interest lies in the effects of the treatments—the combinations of options among communications and shared visualization. For purposes of this prototype experiment, we consider these effects as fixed, rather than random. Simply put, this means that we can characterize the effects of the treatments on the scores achieved by the teams as unknown constants added to an overall mean.

The types of testable hypotheses available to us in this model are again well known. The standard *analysis of variance* techniques for the model allow us to test whether the results show statistically significant differences

- Among the teams (the *rows* of the design matrix)
- Resulting from the sequence in which the teams play the games (the *columns* of the design matrix)
- As a result of the different treatments (the *letters* in the design matrix).

If the evidence points to differences among the treatment effects, we can further explore those differences. We can directly test differences among the communications modes and differences between the visualization modes. Using *orthogonal contrasts*, we can further refine the tests of treatments by examining the difference between no communications and some communications, and between text chat and voice. Finally, we can test whether there is a statistically significant interaction effect between communications and shared visualization. (See Appendix B for a mathematical description of the underlying model and analysis.)

To carry out this analysis, we collected the turn-by-turn shared awareness score of each team as they played each game. At the end of each individual game, we averaged these scores. This procedure results in a 6 x 6 matrix of data, corresponding to our design matrix.

Shared Awareness Scores

Team	Game 1	2	3	4	5	6	Total
1	1.416	3	1.332	2.42	3	3.48	14.648
2	2.534	1.35	3.2	2.84	2.674	3.42	16.018
3	2.11	1.972	2.58	3.1	2.448	1.32	13.53
4	1.314	1.852	2.92	2.434	3.046	3.284	14.85
5	2.808	2.75	3.178	3.01	1.426	2.322	15.494
6	1.282	1.358	1.358	1.444	1.318	2.144	8.904
Total	11.464	12.282	14.568	15.248	13.912	15.97	83.444

Based on this data, we conducted the standard analysis of variance calculations.

Statistical results

Analysis of Variance

Source of variation	SS	df	MS	F	P
Teams (rows)	5.60	5	1.12	4.83	**.001< p <.005**
Sequence (columns)	2.52	5	0.50	2.17	p ~ .10
Treatments (comms/vis)	6.78	5	1.36	5.86	**.001< p < .005**
--Communications	3.19	2	1.60	6.89	.005 < p < .01
--Visualization	1.92	1	1.92	8.31	.005 < p < .01
--Interaction	1.66	2	0.83	3.58	.025 < p < .05
Error	4.63	20	0.83		
Total	19.52	35			

Orthogonal Contrasts		MS	F	P
Comm	Comm – No Comm	3.17	13.67	**.001 < p < .005**
	Text chat – Voice	0.03	0.12	.50 < p
Vis	Shared vis – No shared vis	1.92	8.31	.005 < p < .01

The *p values* shown in the rightmost column of the results matrix indicate the strength of the evidence against the null hypotheses that the various sources of variation have no effect. The smaller the *p value*, the stronger the evidence. Typically, p values of .05 or .01 or less are considered *statistically significant*. As we can see from the above results, we have very strong evidence of differences among the teams and

among the treatments. We have strong evidence of differences among both the communications treatments and the visualization treatments, with somewhat weaker evidence of an interaction effect between communications and visualization. Further exploring the differences, we see strong evidence that the difference between no communications and any communications is the source of most of the differences associated with communications, while there is little evidence that the mode of communications played much of a role.

In simplified form, a summary of our hypothesis tests is as follows:

Hypothesis tests

- Do communications and shared visualization affect building a shared picture? YES
- Do communications matter? YES
- Does mode of communications matter? NO
- Does shared visualization matter? YES
- Is there interaction between comms and shared visualization? Probably

Ancillary data and observations

In addition to the principal data described above, we also collected a set of ancillary data from the teams before, during, and after the experimental runs. This information falls into three categories: basic game data, collected automatically in a Microsoft Access database; voice communications tape-recorded during those games making use of voice; observations made by the control team.

Ancillary data

The main elements of these sources of data are described below.

The automated Microsoft Access database collected the following information:

- Game identifiers (team and codes for the treatment combinations)
- The time required to complete the game
- Location of all 3 SCUDs (row and column)
- Number of turns played
- Type of game (method of communication/visualization)
- Game players
- Roles.

In addition, for each search asset and for each turn of play, we collected the asset identity, where it was placed, and the results of its search. This allowed us to reconstruct the actual sequence of game events. Similarly, we identified each recommended target square selected by each player each turn. (This would allow us to do follow-on analysis of team shared awareness on a turn-by-turn basis.)

Finally, the automated data collection recorded all text chat (identified by game identification number and the player) in chronological

order. If a message was not sent to the entire team, the identity of the designated recipients was recorded as well.

We also tape-recorded all games during which the players used teleconferencing (voice) as the mode of communication. (All players were aware of this and agreed to its use ahead of time.) In addition, after each team had completed all six iterations of the game, we conducted a Hot Wash via teleconference and recorded all this voice traffic.

We collected additional information from the players in the form of an online background questionnaire and a subjective assessment questionnaire. These instruments are described in Appendix C.

Observations

Besides our statistical analysis of the data, we derived a number of interesting results from observing the games. During each game, at least one analyst controlled it (setting up the type of game, coordinating player roles, and managing the activity) and at least one additional analyst acted as an observer. The following details some of the subjective observations we made during the course of the experiment.

1. *The fact that players were playing a game appeared to promote bonding and trust.*

In four of the six teams, SCUDHunt players had never met, and they had no idea of the background of their fellow teammates. The literature dealing with virtual teams suggests that the best way to form an effective virtual team is to have them meet face-to-face initially to establish a common set of goals and objectives and to build a shared trust. When we observed the games played by the teams that did not know each other, we were surprised at how quickly they appeared to become a tight, bonded team. One of the players commented that he felt that they came together faster as a team because they were not competing on an organizational or political level. He wrote “the lack of conflicting personal agenda or career goals allowed for exceptional team performance.” It might be that teams that play a game together, even if that game is distributed, might build trust

faster than teams that do not play a game. This observation deserves further research.

2. *Some female players appeared to be more concerned than other players about building a consensus.*

Three of the six teams included both male and female players. In these teams, it appeared that some of the female players were more concerned that the team reach a consensus about what to do than other, male, players appeared to be. Those players strove to extend the discussions until players agreed on what their individual target recommendations were. Could there be a relationship between the mix of male and female players and which teams have a higher degree of shared situational awareness? This question also deserves further investigation.

3. *Teams that established a specific process for playing the game appeared to have a better shared awareness.*

As we observed the play of the game, we noted that some teams developed a well-defined process for how they would place their assets and search the game board for SCUDs. Those teams that employed such a repeatable process appear to have had higher SSA scores. Another observable benefit from teams that had a repeatable process was that they did better when they played the game in which they used a shared visualization device but had no direct communications. When they saw the shared visualization, they knew, based on previous experience, which assets had been placed in which locations. A comment from one of the players that is related to this phenomenon was, “Because this was the fifth game, we didn't need as much comm capacity: I knew where (a player) would start the SEALS and JSOTF, and I had a pretty good idea of where the other sensors would go.”

4. *Teams that started with the no communications game seemed to take longer to achieve higher scores in subsequent games.*

We did not have time to analyze the data fully on this subject, but preliminary observations indicate that having some form of communication (text or phone) when a team is young and not fully familiar with each other is beneficial. For the SCUDHunt teams who played the no

communications game early on, it appeared to be harder for them to develop a rhythm as a team.

Discussion and conclusions

The SCUDHunt experiment demonstrates our ability to design and produce simple yet rich gaming environments. Such environments can serve as testbeds for statistically designed experiments to produce quantitative data, allowing us to conduct rigorous statistical tests of hypotheses focused on well-defined research issues.

The specific results of the experiment as described above support the importance of communications and shared visualization tools as aids in the building of shared situational awareness among members of distributed teams. The results also indicate that voice communications may not be superior to real-time text communications in the kind of environment represented by our use of SCUDHunt in this experiment. (For example, there was no time pressure involved in the game as we played it.) These results are certainly preliminary and based on a specific set of tools, teams, and circumstances. Nevertheless, these indications are worth pursing in future research focused on specific issues associated with them.

In the course of the experiment, we collected far more information than we could analyze or report on here. That information includes demographics about the players and results of surveys taken after each game played. The full set of data from each game includes shared situational awareness scores after each turn, as well as details of recommended target squares and how accurate the recommendations were relative to the actual target locations. This broader data set would allow us to explore issues as diverse as learning effects (both within games and across the sequence of games) and possible effects of personality on the ability of the teams to build their shared situational awareness. One obvious area we had insufficient time and resources to explore is the relationship between “good” shared situational awareness and “accurate” target locations. To what extent did the players work together to improve their overall accuracy, and to what extent did their collaboration lead to the Pied Piper syndrome

as a single strong personality drove the entire team off the cliff of poor target location?

This experiment essentially allowed the players unlimited time to make their decisions. What would happen under time pressure? What is the overall relationship between the different modes of communication and visualization and the time required to play the game effectively? Does the use of text chat, for example, require more time for the players to accomplish their task? We could easily conceive of future experiments during which we limit the overall game time and measure directly the effect of time pressures on developing shared situational awareness.

The experimental experience as a whole also taught broader lessons. Chief among these is the extraordinary logistical difficulties involved in conducting structured experiments with human subjects, particularly when those subjects are unpaid volunteers recruited from across the United States. Coordinating schedules to get groups of four players to collaborate in real time using telephone and internet connectivity proved as daunting a challenge as any we faced in this effort. It is little wonder that more experimentation of this type does not take place.

As a result of our experience with these difficulties, we speculate on the prospects for a new approach to this sort of experimentation, particularly in the early phases of exploring a particular set of issues. Our idea is to develop autonomous game-playing agents to conduct the initial experiments. Could we build such agents and endow them with human qualities of interest (for example, Myers-Briggs personality types) in a manner designed to reflect the effects of those qualities on the play of the game? There is evidence that agent-based technology could work here. After preliminary, agent-based experiments identified likely parameters of high interest, we could design follow-on experiments to use human subjects to focus on those parameters.

One example of the general approach we envision may be seen in the book *Simulating Society.*[24]

> *Simulating Society* explores the basis for social and economic behavior. Using the methodology of computer simulation, the authors model various factors that are involved in a system of individuals (or agents) who interact socially and economically with one another. Computer simulations are extremely useful in the social sciences. It provides a laboratory in which qualitative ideas about social and economic interactions can be tested. This brings a new dimension to the social sciences where 'explanations' abound, but are rarely subject to much experimental testing.[25]

Games such as SCUDHunt could serve as flexible testbeds for exploratory research, particularly to practitioners in situational awareness and related fields. It is easy to see immediate applicability to Gary Klein's work on Cultural Lens, for example.[26] We could incorporate individual aspects of cultural differences (such as power distance and risk avoidance) into agent behaviors and explore how those concepts affect command and control in a coalition effort quantitatively and in detail. Similarly, we could use SCUDHunt or a similar game to do a detailed assessment of Mica Endsley's four-point model of team situational awareness, exploring how team members share required information, use devices available for sharing information, develop shared mental models or other mechanisms for sharing situational awareness, or employ processes specifically effective at sharing such information.[27]

Solid, quantitative research into these areas can address the problems associated with a field in which "'explanations' abound, but are rarely subject to much experimental testing." They will provide the sort of

---

24. Richard J. Gaylord and Louis J. D'Andria. *Simulating Society: A Mathematica Toolkit for Modeling Socioeconomic Behavior.* New York: Springer Verlag, 1998.

25. Taken from the web site describing *Simulating Society,* http://www.telos-pub.com/catalog/MATHEMATICA/SimSoc.html.

26. Klein, Pongonis, and Klein, *Cultural Barriers . . .*

27. See, once again, Bolstad and Endsley, "Shared Mental Models."

firm foundation for developing tools and approaches to understanding key elements of future operational environments, in which team-based decision-making, or "network-centric warfare" may come to play increasingly important roles. The more we understand about how and why our own decision processes work (or not), the greater our ability to improve them. The more we understand about how the decision processes of real or potential adversaries work, the greater our ability to degrade them.

Appendix A: SCUDHunt[28]

Objective

SCUDHunt is a simple, short, abstract game of command, control, communications, intelligence, surveillance and reconnaissance (C3ISR) played by a team of variable size (typically 7 ± 2). The objective of the game is to explore and document some variables that may facilitate or obstruct the development of shared cognition or "situational awareness" in a team.

More specifically, the game will enable structured comparisons of process and outcomes between collocated teams (based on face-to-face interaction) and distributed or "virtual" teams (based on online interaction). The game requires group decision-making and allocation of scarce resources under conditions of time pressure and uncertainty.

Components

The game design is intended to support two different implementations:

- Tabletop game that can be played by a collocated team and control group (using either a computer and display system, or a map, dice, and cardboard counters)
- Online game hosted on a web server that can be played asynchronously by a distributed team using standard web browsers. The control function may be either manual or automated.

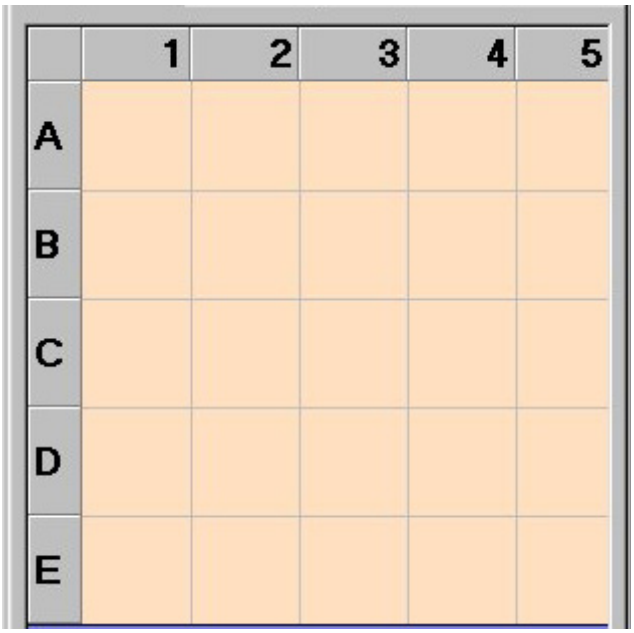
28. Disclaimer: SCUDHunt is an unclassified product. Game scale, timelines, and force structure are notional and abstract. SCUDHunt is not intended to reflect actual capabilities of U.S. or threat systems or real-world U.S. policy, doctrine, or command relationships.

Background

Prior to the start of play, all players read or view a situation briefing containing essentially the following information (maximum of one page).

The rogue state of Korona has acquired mobile ballistic missiles and weapons of mass destruction. Korona is threatening a U.S. ally, Kartuna, located (off-map) across the narrow Gulf of Sabani. Your team mission is to locate the missile launchers, using various ground, space, air, and intelligence assets.

The elite fanatical Koronan Revolutionary Guard Special Artillery Regiment, with a number of mobile missile launchers, has deployed from its depot to a secret hide site. This deployment is supported by deception operations that may confuse our sensors somewhat.



Korona is divided into 25 grid squares identified by columns numbered from 1 to 5 and rows lettered from A to E. Row E is the coastline of the Gulf of Sabani. Each of the three targets is randomly placed in a different grid square at the start of play. (Alternatively, a standard 8 x 8 chessboard, or a grid of arbitrary size overlaid on real or

imaginary terrain may be used: this will allow a longer or more challenging game.)

In “Basic” SCUDHunt, targets are stationary, but future versions may implement automated or controller-mediated mechanics allowing targets to move if they are repeatedly detected in a given grid square.

Sequence of play

Each turn represents one day of real time. Players are assigned roles, which give them control of one or more intelligence, surveillance, and reconnaissance “assets.” During a turn in the face-to-face version, players may freely discuss their search strategy with the other players, up to a pre-set time limit. In the online versions, players may read messages from other players, send a message to another player (or set of players), or broadcast a message to all players, up to a pre-set time limit. Players execute the decision by placing assets on the board (by clicking on a grid square, row or column, or placing a counter on the map). A player is never required to place an asset.

After all players have completed their actions for the turn, the day's search results for each asset are displayed to the owning player (only), along with the status of any assets that malfunctioned or aborted a mission before completing a search task. In the face-to-face game, this might be a paper report handed to each player. In the online game it would be a graphic display, which could be printed or saved from turn to turn. Players must, therefore, communicate what they learn each turn in order to create a shared tactical picture.

Players may (optionally) be informed in advance, in general terms, about system reliability or risks for their own asset(s), but not those of other players.

There are three basic search results:

0 - nothing significant to report (grid square seems empty)

? - vehicles detected (may be launchers, deception operations, or routine civilian traffic)

X - launchers detected.

Reported search results may be true or erroneous, depending on the number and type of assets assigned to search a given grid square. (An erroneous detection of a launcher indicates a successful deception operation by the Koronans.) The team mission is a success if the real missile launchers are detected and targeted. The team mission is a failure if at least one real missile launcher is not targeted.

After a variable number of turns determined by Control (players might not be told game length in advance), players are informed that clear signs indicate an imminent Koronan missile launch.

Players are asked to designate some (hopefully small) number of grid squares (representing the players' best guess of the locations of the launchers) that will be targeted for attack. At the conclusion of the game, Control reports back to the players how well they did at identifying actual targets and avoiding errors.

Player assets

Recon Satellite: Searches one entire column each turn. It has a high probability of confirming the absence of vehicles, but cannot reliably distinguish between launchers and dummies.

Manned Aircraft: May fly only along the Gulf (Row E) outside Koronan airspace. It searches the coastal grid squares with excellent reliability and two rows inland (Rows C and D) with reduced reliability. The manned aircraft must “rest” at least one turn between flights due to crew fatigue and maintenance requirements.

UAV: May enter Koronan airspace to search any row. It has good search reliability. For each grid square it enters, there is a chance that it will crash or be shot down, which aborts any further search on that turn. There is a variable probability that a lost UAV will be replaced the next day.

COMINT: Searches any grid square. Cannot reliably distinguish launcher from deception operations.

HUMINT: Searches any grid square with excellent reliability. The agent has limited mobility; after initial placement on any square, he may only remain in the same square or move to an adjacent grid square. Each turn the agent is on the board, there is a chance he will be caught and executed. He cannot be replaced.

Spec Ops: May be inserted to search any grid square with excellent reliability. Can reliably distinguish between deception and launchers. Each turn that the Spec Ops team is in play, there is a chance the team will be compromised and forced to perform an emergency extraction. If extracted, the team will be unavailable for 1 or more days to rest and refit. (Beginning on the 2nd rest day there is a chance the team will return to available status).

Navy SEALS: Similar to Spec Ops, but may be inserted only into coastal grid squares.

Additional assets may be added to accommodate more players or to reflect special interests of a target audience, such as sea-based ELINT platforms, different kinds of space-based sensors, or information operations.

Player roles

The number of players is variable; determining the best set of player roles for a given team size will require some thought and experimentation.

In addition to the general situation briefing given to all players, each player should receive a capsule description of his own player role, including motivations, command relationships, assets, and special abilities or limitations.

Possible players might include:

Space Asset Manager: controls recon satellite

Electronic Intelligence Manager: controls COMINT, SIGINT assets

Manned Air Asset Manager: controls manned air assets

UAV Manager: controls UAV

SpecOps Manager: controls Special Operations teams

Spy Master: controls HUMINT.

Search results

All probability calculations for search results will be performed by Control (or by the computer in the on-line game). Players do not know the exact probabilities, only a general description of their own asset's system reliability, detection phenomenology, accuracy of information, and timeliness of information. Grid squares searched by multiple assets will report each asset separately (players may be given a simple formula or utility to compute joint probabilities based on shared information). The performance of each sensor system in the game is defined by a table of the following form.

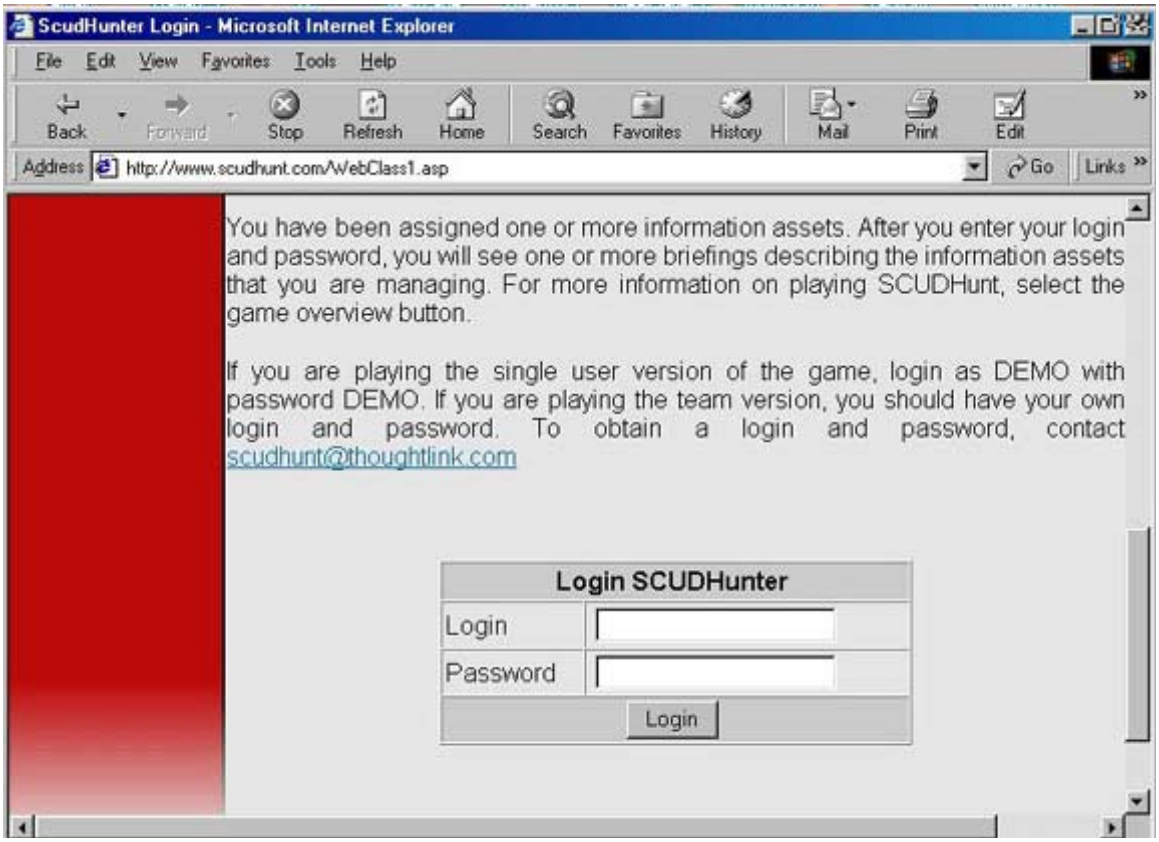
Sensor

	Real-World Situation		
Report	0	?	X
0	.7	.2	.1
?	.2	.6	.2
X	.1	.2	.7

Web implementation

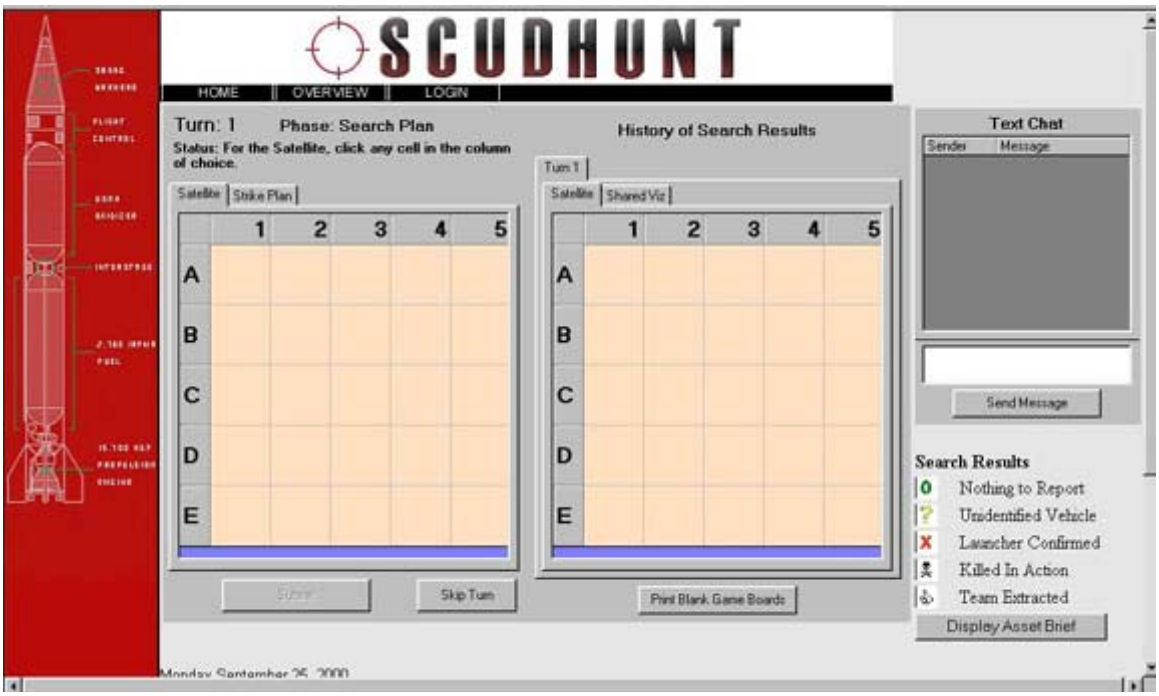
The web-based version of SCUDHunt is implemented in Visual Basic 6.0. We use Web Classes on the server for game sequencing and database access. We use ActiveX controls on the client-side browser for the game board and interactions. The game instrumentation data is stored in a Microsoft Access database. SCUDHunt comprises two applications, one for the Player and one for the Controller.

The Player application comprises three screens—Sign In, Asset Briefing, and Game Board. The Sign In screen permits entry of a username and password for authentication.

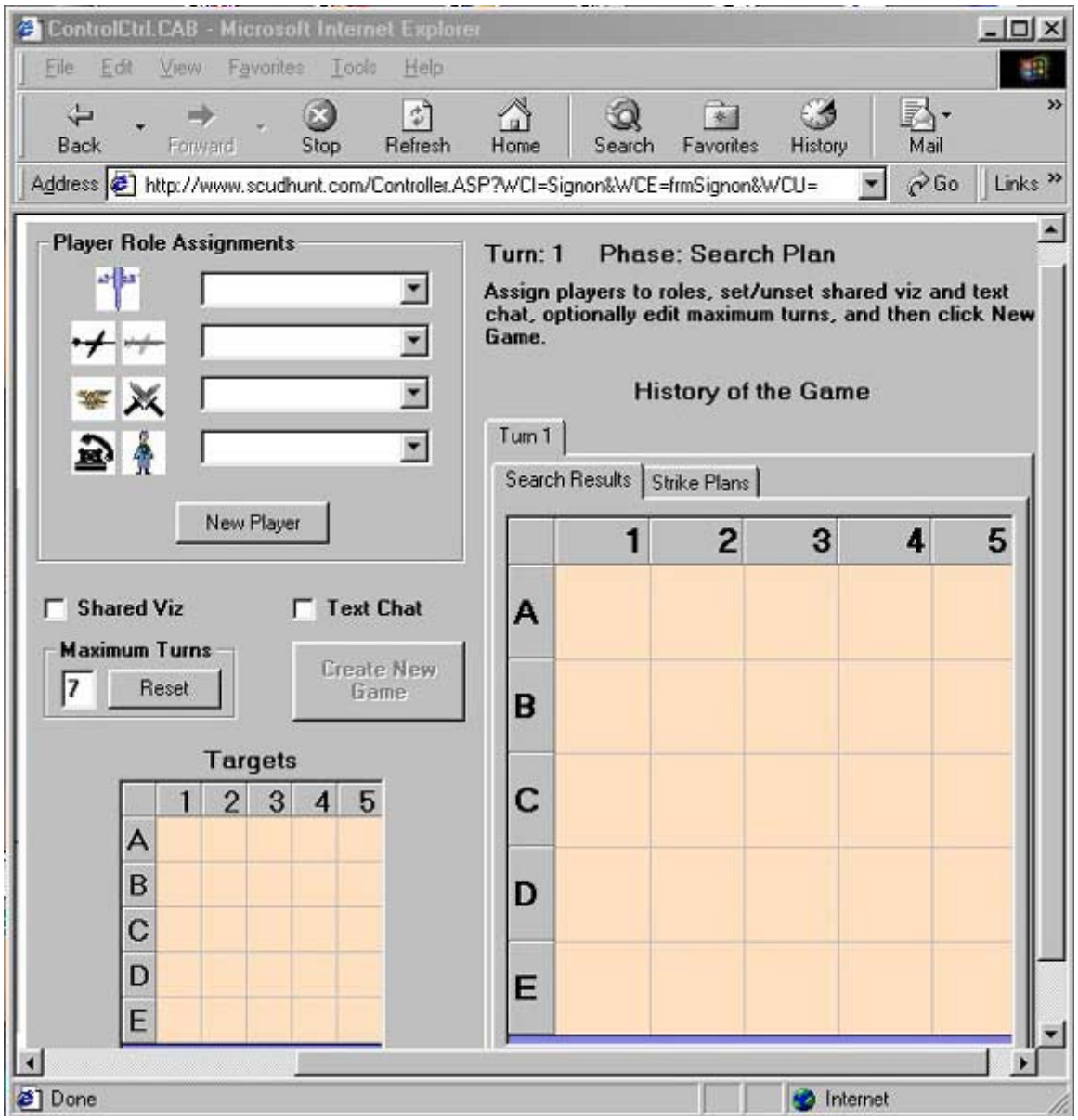


The Asset Briefing screen provides details on the capabilities of the assets managed by that player. Here is an example, the screen for the Special Operations (SpecOps) manager.

The Game Board screen is divided into five areas—Text Chat control (if enabled for that running of the game); a legend of Search Results; Status giving the current turn, phase, and action; the interactive Game Boards (one for each Search Plan of this player's assets and one for the Strike Plan); and a set of History game boards (one for each turn for each of this player's assets and a Shared Visualization showing the Search Results for all players' assets (if enabled for that running of the game).



The Controller application comprises two screens—Login and Main. The Controller Login screen authenticates the Controller name and password and is similar to the player's log-in screen. The Main screen lets the Controller set up a new game and watch its progress.



The Controller chooses the players and assigns them assets, enables or disables shared visualization and text chat, and sets the maximum number of turns for the game. The game program automatically generates a new random target placement after the Controller clicks the New Game button. The Controller can then watch the progress of the game by monitoring Status information and game boards for each turn, which display aggregates of the Search Results and Strike Plans for all the players.

Appendix B: The experimental model

The SCUDHunt experiment was designed as a 6 x 6 Latin Square linear model. The different teams appear as rows. The sequence of games appears as columns. The treatment combinations appear exactly once in each row and column.

Design matrix

	Game 1	Game 2	Game 3	Game 4	Game 5	Game 6
Team 1	B	E	A	C	F	D
Team 2	**D**	**A**	**E**	**B**	**C**	**F**
Team 3	**E**	**B**	**C**	**F**	**D**	**A**
Team 4	A	F	D	E	B	C
Team 5	**F**	**C**	B	D	A	E
Team 6	C	D	F	A	E	B

A = No Comms, No shared vis
B = No Comms, Shared vis
C = Text chat, No shared vis
D = Text chat, Shared vis
E = Voice, No shared vis
F = Voice, Shared vis

The standard model for such a Latin Square design is:

$$Y_{ij(k)} = \mu + \rho_i + \kappa_j + \tau_k + \varepsilon_{ij(k)} \ ,$$

where

$Y_{ij(k)}$ is an observation (whose mean we normally symbolize as $\mu_{ij(k)}$) ,

μ = an overall constant ,

ρ_i, κ_j, and τ_k are the effects associated with the rows, columns, and treatments, respectively ,

$\varepsilon_{ij(k)}$ are independent $N(0,\sigma^2)$,

$i = 1,2. . . 6$; $j = 1,2 . . . 6$; $k = 1, 2 . . . 6$.

Note that the subscript *k* is parenthesized in the equation to indicate that not all values of k appear with each combination of *i* and *j* because the treatments occur only once in each row (*i*) and column (*j*).

In the standard analysis of variance, our testable hypotheses take on the following forms:

For row effects:

H_0: all $\rho_i = 0$

H_1: not all $\rho_i = 0$.

For column effects:

H_0: all $\kappa_j = 0$

H_1: not all $\kappa_j = 0$.

For treatment effects:

H_0: all $\tau_k = 0$

H_1: not all $\tau_k = 0$.

In this experiment, each of the treatments—which we have symbolized as letters in the design matrix but identified by their ordinal number for the purposes of the equations (e.g., A = 1, D =4)—is actually composed of two crossed factors, defined as follows:

τ_1 = A = No communications, no shared visualization

τ_2 = B = No communications, shared visualization

τ_3 = C = Text chat, no shared visualization

τ_4 = D = Text chat, shared visualization

τ_5 = E = Voice, no shared visualization

τ_6 = F =Voice, shared visualization.

Testing the effects of the component factors in these six treatments requires the use of orthogonal contrasts. A contrast is simply a linear combination of factor means. A contrast is usually symbolized by the letter L and defined

$$L = \sum_{j=1}^{r} c_j \mu_j \quad ,$$

where the c_j are coefficients that meet the criteria

$$\sum_{j=1}^{r} c_j = 0 \quad .$$

Orthogonal contrasts are pairs of contrasts L_1 and L_2 such that

$$L_1 = \sum_{j=1}^{r} c_{1j} \mu_j \qquad L_2 = \sum_{j=1}^{r} c_{2j} \mu_j \quad ,$$

where

$$\sum_{j=1}^{r} c_{1j} = 0 \qquad \sum_{j=1}^{r} c_{2j} = 0$$

and

$$\sum_{j=1}^{r} c_{1j} c_{2j} = 0 \quad .$$

In our case, we use such contrasts to reflect the effects of communications by comparing No Communications to the average of Text Chat and Voice. We can choose these contrasts in such a way that they

are also each orthogonal to the contrast between Shared Visualization and No Shared Visualization, so that these three mutually orthogonal contrasts represent a complete decomposition of the main treatment effects.

Appendix C: Questionnaires

Below are copies of the pre-game and post-game questionnaires we asked all the SCUDHunt players to complete.

Pre-game questionnaire

This survey is being given to all SCUDHunt participants as part of a study supporting DARPA's *Wargaming the Asymmetric Environment* program.

The purpose of the survey is to determine your background with computers, collaboration tools, computer games, and the military.

This information will be kept confidential. It may be used in two ways.

It may be aggregated in order to characterize the different SCUDHunt teams.

Individual answers may be used as quotes in the study's final briefing. If so, they will be presented as anonymous quotes, and the individual will not be identified.

1. Name (first, last):

2. Organization:

3. E-mail address:

4. Please assess your overall level of computer expertise:

Low: Seldom use computers

Moderate: Use computers often and are comfortable with them

High: Use computers a lot and feel very confident about your abilities

5. How you ever used any of the following collaboration tools:

5a. E-mail: Yes No

If Yes, how often:

Often: Weekly or more

Frequently: Monthly or more

Rarely: Yearly or less

5b. Web browsers: Yes No

If Yes, how often:

Often: Weekly or more

Frequently: Monthly or more

Rarely: Yearly or less

5c. Text Chat: Yes No

If Yes, how often:

Often: Weekly or more

Frequently: Monthly or more

Rarely: Yearly or less

5d. Video Teleconferencing: Yes No

If Yes, how often:

Often: Weekly or more

Frequently: Monthly or more

Rarely: Yearly or less

6. Do you like playing games: Yes No

7. Have you played computer games before? Yes No

If Yes, how often:

Often: Weekly or more

Frequently: Monthly or more

Rarely: Yearly or less

8. Have you played web-based computer games before? Yes No

If Yes, how often:

Often: Weekly or more

Frequently: Monthly or more

Rarely: Yearly or less

9. Have you played multi-player web-based games before? Yes No

If Yes, how often:

Often: Weekly or more

Frequently: Monthly or more

Rarely: Yearly or less

Please list the games you most often play (up to 10):

10. Do you have military or defense-related experience? Yes No

If Yes, list organizations and or service:

How many years:

11. Have you ever worked with people from your organization who are geographically separated from where you work? Yes No

If Yes, how often:

Often: Weekly or more

Frequently: Multiple times a year

Rarely: Once or twice

Never:

Describe some of your experiences working as a member of a distributed team and the tools that you used to communicate?

12. Have you ever worked as part of a team composed of people outside your organization? Yes No

If Yes, how often:

Often: Weekly or more

Frequently: Multiple times a year

Rarely: Once or twice

Never

Describe some of your experiences working as a member of a distributed team and the tools that you used to communicate?

Post-game questionnaire

SCUDHunt Post-Game Questionnaire

Name:

Date:

Your Myers-Briggs type (if you know it and if you don't mind sharing it):

(To determine this, you can take an online test at

http://www.onlinepsych.com/public/Mind_Games/ptt/pttframe.htm

or

http://www.keirsey.com/cgi-bin/keirsey/newkts.cgi)

How well did you know the other team members prior to today's session?

Do not know them

Slightly

Well

Very well

During the games in today's session,

- Was there a leader? (Yes/No):

- If so, who?

- Did it change over time?

- Did team members take on roles? (e.g., leader, brainstormer, facilitator, kept track of details) If so, please explain:

Which game (or set of communication/visualization conditions) was easiest and why?

Which game was hardest and why?

Game 1: Text chat, No shared visualization

On a scale of 1-5 (1 = Not At All and 5 = Very Much), was this game fun?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your team's shared picture of where the SCUDs were at the end of the game?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your ability to visualize where the SCUDs were on the game board?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your ability to communicate with your teammates?

Did your team try to come to consensus on decisions on sensor placement? (Yes/No):

Did your team try to come to consensus on decisions on launcher locations? (Yes/No):

If your team tried to come to consensus, on a scale of 1-5 (1 = None/Very Little and 5 = A Lot/Very Much)

-- To what extent was there tension between you and other team members?

-- To what extent did you and the others agree?

-- To what extent did you feel that you and the other team members were equals?

What are your overall comments about Game 1:

Game 2: Text chat, Shared visualization

On a scale of 1-5 (1 = Not At All and 5 = Very Much), was this game fun?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your team's shared picture of where the SCUDs were at the end of the game?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your ability to visualize where the SCUDs were on the game board?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your ability to communicate with your teammates?

Did your team try to come to consensus on decisions on sensor placement? (Yes/No):

Did your team try to come to consensus on decisions on launcher locations? (Yes/No):

If your team tried to come to consensus, on a scale of 1-5 (1 = None/Very Little and 5 = A Lot/Very Much)

-- To what extent was there tension between you and other team members?

-- To what extent did you and the others agree?

-- To what extent did you feel that you and the other team members were equals?

What are your overall comments about Game 2:

Game 3: Voice, Shared visualization

On a scale of 1-5 (1 = Not At All and 5 = Very Much), was this game fun?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your team's shared picture of where the SCUDs were at the end of the game?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your ability to visualize where the SCUDs were on the game board?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your ability to communicate with your teammates?

Did your team try to come to consensus on decisions on sensor placement? (Yes/No):

Did your team try to come to consensus on decisions on launcher locations? (Yes/No):

If your team tried to come to consensus, on a scale of 1-5 (1 = None/Very Little and 5 = A Lot/Very Much)

-- To what extent was there tension between you and other team members?

-- To what extent did you and the others agree?

-- To what extent did you feel that you and the other team members were equals?

What are your overall comments about Game 3:

Game 4: No comm, No shared visualization

On a scale of 1-5 (1 = Not At All and 5 = Very Much), was this game fun?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your team's shared picture of where the SCUDs were at the end of the game?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your ability to visualize where the SCUDs were on the game board?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your ability to communicate with your teammates?

Did your team try to come to consensus on decisions on sensor placement? (Yes/No):

Did your team try to come to consensus on decisions on launcher locations? (Yes/No):

If your team tried to come to consensus, on a scale of 1-5 (1 = None/Very Little and 5 = A Lot/Very Much)

-- To what extent was there tension between you and other team members?

-- To what extent did you and the others agree?

-- To what extent did you feel that you and the other team members were equals?

What are your overall comments about Game 4:

Game 5: Voice, no shared visualization

On a scale of 1-5 (1 = Not At All and 5 = Very Much), was this game fun?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your team's shared picture of where the SCUDs were at the end of the game?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your ability to visualize where the SCUDs were on the game board?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your ability to communicate with your teammates?

Did your team try to come to consensus on decisions on sensor placement? (Yes/No):

Did your team try to come to consensus on decisions on launcher locations? (Yes/No):

If your team tried to come to consensus, on a scale of 1-5 (1 = None/Very Little and 5 = A Lot/Very Much)

-- To what extent was there tension between you and other team members?

-- To what extent did you and the others agree?

-- To what extent did you feel that you and the other team members were equals?

What are your overall comments about Game 5:

Game 6: No comms, shared visualization

On a scale of 1-5 (1 = Not At All and 5 = Very Much), was this game fun?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your team's shared picture of where the SCUDs were at the end of the game?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your ability to visualize where the SCUDs were on the game board?

On a scale of 1-5 (1 = Poor and 5 = Excellent), how would you grade your ability to communicate with your teammates?

Did your team try to come to consensus on decisions on sensor placement? (Yes/No):

Did your team try to come to consensus on decisions on launcher locations? (Yes/No):

If your team tried to come to consensus, on a scale of 1-5 (1 = None/Very Little and 5 = A Lot/Very Much)

-- To what extent was there tension between you and other team members?

-- To what extent did you and the others agree?

-- To what extent did you feel that you and the other team members were equals?

What are your overall comments about Game 6: