Title: Measuring Situational Awareness through Analysis of Communications: A Preliminary Exercise

Suggested Track: **Cognitive Domain Issues, C2 Analysis, C2 Experimentation**

Authors:

Shawn Weil
Aptima Inc.,
12 Gill Street, Suite 1400
Woburn, MA 01801
Phone: 781-496-2456
Fax: 781-935-4385
e-mail: sweil@aptima.com

Kathleen M. Carley
Carnegie Mellon University
School of Computer Science
Institute for Software Research
International (ISRI)
Center for Computational Analysis of
Social and Organizational Systems (CASOS)
Pittsburgh, PA 15213
Phone: 412 268 6016
e-mail: kathleen.carley@cs.cmu.edu

Jana Diesner
Carnegie Mellon University
School of Computer Science
Institute for Software Research
International (ISRI)
Center for Computational Analysis of
Social and Organizational Systems (CASOS)
407 South Craig Street, Office 118
Pittsburgh, PA, 15213
Phone: 412-519-7576
e-mail: jdiesner@andrew.cmu.edu

Jared Freeman
Aptima Inc.,
1726 M Street, N.W., Suite 900
Washington, DC 20036
Phone: (202) 842-1548 x316
Fax: (202) 842-2630
e-mail: freeman@aptima.com

Nancy J. Cooke
Cognitive Engineering Research Institute
5810 S Sossaman Rd., Ste 106
Mesa, AZ 85212-6014
Phone: 480-988-2173
Fax: 480-988-3162
e-mail: ncooke@cerici.org

Correspondence:

Shawn Weil
Aptima Inc.,
12 Gill Street, Suite 1400
Woburn, MA 01801
Phone: 781-496-2456
Fax: 781-935-4385
e-mail: sweil@aptima.com

Measuring Situational Awareness through Analysis of Communications: A Preliminary Exercise

Abstract

Network centric warfare promises to increase information sharing and allow distribution of decision making. This will improve military effectiveness, but only if the situational awareness (SA) of warfighters is correctly aligned. Modern natural language processing techniques, such as Network Text Analysis (Carley, 1993), are designed to infer the cognitive states of individuals and groups engaged in cognitive collaboration and measure group SA by exploiting data on the information that team members access and generate. An integrated software application, IMAGES, utilizes AutoMap (Diesner & Carley, 2004) as the primary analysis engine to take advantage of the large amounts of communication and report text that naturally occur in collaborative environments. The text generated in the normal course of work is collected and changed into forms that can be compared and analyzed. A comparison of networks based on text from several individuals or groups yields information about the similarity of their respective mental models. Differences among maps may reflect misalignments of SA, which can be remedied by information sharing and targeted communication. An exercise was conducted to assess the potential of NTA as implemented in AutoMap and IMAGES. The results indicate that NTA will allow analysts to effectively assess SA through passive means.

Networked information systems are widely used to distribute data and intellectual tasks to the best specialists available across an organization and around the globe. Corporations use network technologies to outsource software development, telephone support, and other functions to the most cost efficient labor sources internationally. The Department of Defense uses networked collaboration technologies to coordinate its own forces and its coalition partners in wartime operations in Iraq and peace time recovery operations in Indonesia. Future U.S. Navy missions will leverage FORCEnet – a network-centric concept for organizing, connecting, and empowering warfighters (Alberts & Hayes, 2003). Organizations such as the U.S. Navy's Expeditionary Strike Groups (ESG) are already putting these FORCEnet concepts to use. These organizations use networked information systems to improve the processes and outcomes of cognitive collaboration. In particular, organizations expect this technology to ensure that ***information exchange*** is accurate and efficient (i.e., that people have the right information at the right time) and that team members ***are aligned in their interpretation of that information and the inferences drawn from it***[1]**.** These effects should ensure that teams make better and timelier decisions, that their actions are better synchronized and coordinated, and that the mission effects are powerful and sure.

Network technology, alone, does not ensure these effects. The quality of cognitive collaboration varies widely between teams, whether they work executed face-to-face (McComb, 2005; Isaacs and Clark, 1987; Warner & Wroblewski, 2004), or in sophisticated, networked environments (Cooke, 2005; Entin and Serfaty, 1999). When networks do make a difference, it is sometimes for the worse. Information overload is commonplace in networked workplaces, misunderstandings are frequent, and both decisions and actions are frequently – sometimes tragically – mis-coordinated (Woods, Patterson, & Roth, 2002, Weil et al, 2004). These are the symptoms of flawed situational awareness at the organizational level, and the consequence of failed cognitive collaboration.

To improve cognitive collaboration – particularly distributed, network-centric organizations – leaders must ***measure, monitor, and manage*** it. Doing any of these presents challenges, but measurement is especially difficult. Measures of "return on investment" (e.g., net profits) and mission effectiveness (e.g., number of enemy units destroyed) are often proposed and are easily taken, but they are distal, summative measures of the effects of information systems (and a host of other forces) on organizational performance; they are not measures of cognitive collaboration itself. Measures of information technology performance assess transfer rates, data availability, and the like, but they do not address semantics. They do not shed light, in particular, on the distribution of specific information content (***information exchange***) and the coordination of meaning across an organization (***interpretation and inference****).*

A traditional challenge in measuring cognitive processes is that they are largely unobservable. Laboratory measurement of cognition typically involves highly calibrated procedures that elicit artificial actions (e.g., rapid key-presses) and inferences based on the results. Other measures of cognition, such as eye tracking or neuroimaging techniques, are more direct, but sometimes difficult to interpret.

However, ***cognitive collaboration*** is not internal to the individual, but external. Thus, it is easier to observe and measure in operational settings. As Cooke (2005) has stated, "coordination is team cognition." In a networked environment, one form of coordination – communication – is both accessible and useful for measuring cognitive collaboration. Communications in networked information systems may include email and text chat. The memos, reports, and other documents that team members read and write also are communications – albeit communications made at a slower pace.

We argue that cognitive collaboration can be ***measured*** by (1) tracing the patterns of these communications, which represent information exchange, and (2) analyzing their content, which represent

[1] Under our definition, information networks are intended to improve situational awareness in all three aspects defined by Endsley (2000): perception, comprehension, and prediction.

interpretation and inference. Further, team leaders can ***monitor*** cognitive collaboration***,*** given intuitive, informative visualizations of collaboration patterns and content. This puts leaders in a position to ***manage*** cognitive collaboration.

FORCEnet

The U.S. Navy is undergoing a rapid transformation in the operations it conducts — the enemies it faces, the resources it employs, the capabilities it delivers, the ways it coordinates with other branches of the armed services, and the organizational structure it uses to bring those new resources and capabilities to bear against a new generation of enemies. This revolution began with the development of the "network-centric warfare" concept, and evolved into the definition of FORCEnet as a future Navy organizing principle (again reference to FORCEnet).

FORCEnet is viewed as the operational construct and architectural framework for naval warfare in the information age, integrating warriors, sensors, command and control, platforms, and weapons into a networked, distributed combat force. It is envisioned that FORCEnet will provide the architecture to increase combat capabilities through aligned and integrated systems, functions, and missions. Like network-centric warfare in general, it promises to transform situational awareness, accelerate speed of decision making, and greatly distribute combat power (Figure 1). FORCEnet will harness information for knowledge-based combat operations and increased force survivability. It will also provide real-time enhanced collaborative planning among joint and coalition partners (Alberts & Hayes, 2003; NWDC, 2004; Mayo & Nathman, 2003). FORCEnet, therefore, promises to transform military operations by directly connecting the elements of the military thereby accelerating the speed and accuracy of decision making, and integrating knowledge to dominate the battle space.

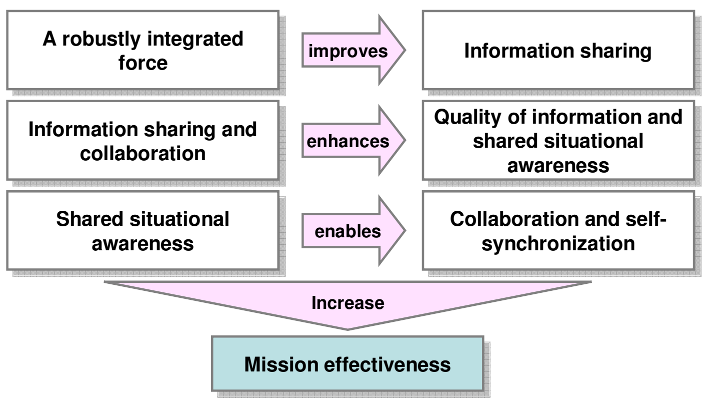

Figure 1. The Tenets of Network-Centric Warfare (Alberts & Hayes, 2003).

Traditional, hierarchical military organizations have evolved over many years, shaped by factors such as physical separation, the availability and bandwidth of analog communication and information exchange channels, and intrinsic human capabilities. In today's world, network connectivity (enhanced through digital satellite transmissions and wireless technology) has alleviated many, but not all, of these constraints. It is now possible to transfer information almost instantaneously between points anywhere in the world. It has become possible, in theory, for anyone anywhere to know anything, if it is happening within the range of networked sensors. There is a widespread recognition that these changes in communication facilitate flatter organizational structures, with higher horizontal communication rates, in contrast to more traditional hierarchical structures based primarily on vertical communication (Fukuyama, 2000, Gilly & Gilly, 1991). Heterarchical (i.e., flat) information structures can surely be reconciled with inherently hierarchical command structures, but it will require an ability to ***measure***, ***monitor***, and ***manage*** the new communication capabilities in the next generation of network-centric C2 organizations.

To achieve the FORCEnet vision, numerous technologies are being integrated into existing naval resources to improve information sharing in command, control, communications, computers, & intelligence (C4I). Among these are tools that are designed to improve tactical situational awareness, such as the ACDS (Advanced Combat Direction System) (reference) and GCCS-M (Global Command and Control System - Maritime) (reference), and those that utilize web-based technologies to integrate mission relevant information, tools, and displays to facilitate group interaction and augment decision-making (e.g., K-Web; Pester-DeWan, et al., 2003). Integrated collaborative environments, such as GVO, are being evaluated to determine the extent to which they can support chat, bulletin-board type discussion, distributed file sharing, synchronous whiteboard and note taking in the military (Bordetsky, 2005). Innovative tools for distributed knowledge sharing, such as E-Wall (Keel, 2005) are being examined by the Navy. Below the radar, standard business collaboration tools have already taken hold: email, chat, and standard Microsoft Office applications that support collaborative editing.

Some of the technologies being developed in support of FORCEnet are currently becoming operational as part of to U.S. Navy's Expeditionary Strike Groups (ESG). The ESG is composed of surface warfare resources augmented by a Marine Expeditionary Unit (MEU), Naval special operations, and surface warfare resources, totaling eight or more coordinated vessels (U.S. Navy, 2004). The ESG is a highly mobile, self-sustaining force, capable of conducting expeditionary warfare operations to support a full range of theater contingencies. Those missions could range from humanitarian and disaster relief to combat operations.

The ESG is a ready example of the need for technology that supports ***measurement***, ***monitoring***, and ***management*** of cognitive collaboration over rich mission data. Many of the missions to which an ESG could be tasked necessitate information exchange and collaboration with other military services, allied nations, and non-governmental organizations (NGO). The sheer volume of information generated during missions present challenges to situational awareness. Differences in organizational and national culture may lead to unexpected conflicts in the interpretation of mission information or decisions based on it. To detect and prevent these problems, ESG leadership will need the capability to measure the distribution and use of information within its network, monitor that activity, and manage it.

Core Concepts: Mental Models and Situational Awareness

Cognitive and organizational psychologists have, for some years, been concerned with the questions of how to measure, monitor, and manage (e.g., teach, influence, manipulate) the knowledge of ***individuals***. This work has produced a rich literature concerning ***mental models*** and expertise (Carley & Palmquist, 1992). In cognitive theory and research, a ***mental model*** has been construed as an internal artifact which may represent a space of logical assertions and inferences (Johnson-Laird, 1983), the physics of interactions between components of machines (Rouse & Morris, 1986), or the social interactions within organizations (Rouse, Cannon-Bowers, and Salas). In general, (Mathieu et al., 2000, p. 274) "mental models…help people to describe, explain, and predict events in their environment." Mental models are conceived to be varied in form and function (as illustrated above), complex -- based on prior experiences, prior perceptions, values, and beliefs (Lambert & Walker, 1995) – and difficult to acquire (Rouse & Morris, 1985). Work in organization theory also discusses the role of team mental models in terms of tacit and actual shared knowledge (Carley, 1997; Klimoski & Mohammed, 1994)

Although the mental model construct seems well-suited for knowledge of the type acquired from textbooks (e.g., physics, mathematics), we argue that it may be less inadequate for dealing with more fleeting knowledge of complex and dynamic situations of the kind found in aviation and military environments. This fleeting knowledge has been associated with ***situational awareness*** (SA)**.** Recently, this work on individual knowledge management has turned towards the knowledge sharing and knowledge building processes of groups. Here, we highlight concepts from this literature that inform the design of methods for improving cognitive collaboration in the ESG, FORCEnet, and similar operational organizations.

The transformation of the military toward network-centric warfare is predicated to a large extent on the belief that properly aligned information ultimately leads to increased mission effectiveness. Simply put, individuals working together within an organization are more effective when they have the information that is essential for accomplishing their individual tasks. Significant resources have been expended on sensors and data transmission to provide information. But information is only the start. The information needs to be properly communicated, perceived, comprehended, and interpreted to contribute to a knowledge state. There have been comparatively few resources expended to assess the impact of this infrastructure on an organization's knowledge, ability to maintain fleeting states of knowledge (i.e., SA), or methods to manipulate it. There are however, some tools that do just this; e.g., AutoMap (Diesner and Carley, 2004; http://www.casos.cs.cmu.edu/projects/automap/) and ORA (%%%http://www.casos.cs.cmu.edu/projects/ora/) which we are using.

Management of group knowledge has some complications that individual knowledge management is able to avoid. It is not enough to assess, monitor, or manage the knowledge of each individual in the group and aggregate the results. For instance, at the level of the organization, two types of SA are generally acknowledged. There is information that needs to be shared among all team members or subsets of them. We will call this ***shared SA***. There is also information that needs to be held by the team as a whole, but not necessarily by every team member. We will call this ***team SA***. These knowledge requirements are based on the responsibilities and tasking of each person, and change over time. The distribution of knowledge requirements across a team or organization depends on the team structure, the degree of heterogeneity or specialization on the team, and the hierarchical diversity among team members. An effective team requires a balance of shared and team SA (see Cooke, Salas, & Cannon-Bowers, 2000). This balance is implicitly recognized by the military; the Department of the Army (1994) has defined SA as the "…ability to have accurate and real-time information of friendly, enemy, neutral, and noncombatant location; a common, relevant picture of the battlefield scaled to specific level of interest and special needs."[2]

The two knowledge constructs, mental models and SA, have been theoretically connected in the team literature. Specifically, a shared understanding of the domain, task, and team is a prerequisite for the more fleeting understanding (SA) that is derived in a dynamic environment. Team research has shown that effective team performance and team SA requires team members to hold common mental models or cognitive representations of tasks and the team (Cannon-Bowers, Salas, & Converse, 1993; McComb, 2005; Graham, 2005). These shared mental models allow team members to coordinate, often implicitly, within their team. As research on the relationship between mental models and team effectiveness has progressed, researchers have delineated the mental models that are important in a team environment (see Rouse, Cannon-Bowers, & Salas, 1992; Converse et al., 1991). Specifically, it is argued that there are at least four main types of mental models that are important in understanding team effectiveness: (1) equipment, (2) task, (3) team (Rouse et al.), and (4) team interaction (Cannon-Bowers, Salas, & Converse, 1993). In addition, these models are argued to be hierarchically related (Rentsch & Hall, 1994). At the lowest level, containing the most basic knowledge is the equipment mental model. This model contains knowledge regarding how the equipment that the team is interacting with works and allows members to predict what the equipment is likely to do and when to make a response. The task mental model, at the next level of hierarchy, contains knowledge relating to the basic attributes of the task and how to accomplish it. Therefore, this model may contain knowledge and beliefs regarding task procedures, goals, strategies, and the interrelationships among this content. The last two mental models in the hierarchy are highly related and have been argued to be the most important to the effective attainment of coordinated team action (see Cannon-Bowers, Salas, & Converse, 1993) and the focus of the current effort. The ***team mental model*** contains knowledge regarding the individual's role in the task as well as the role of others.

---

[2] Endsley has distinguished between ***shared situational awareness*** – "the degree to which team members possess a shared understanding of the situation with regard to their shared SA requirements (Endsley & Jones, 1997)" – and ***team situational awareness*** – "the degree to which every team member possesses the SA required for his or her responsibilities (Endsley, 1995)."

More specifically, this model contains knowledge about team member characteristics, including their task knowledge, skills, abilities, and preferences. Similarly, the ***team interaction model*** also contains information about the team, but the information is related to the individual and collective requirements needed for effective team interaction. In addition, the team interaction model contains information and beliefs regarding how each member's roles and responsibilities intersect with other members on the team. Although the team mental model allows members to form expectations and predict future performance regarding how members are likely to act in a situation, the team interaction model takes it a step further by allowing members to anticipate and sequence their collective actions. Thus the team mental model especially affords team situational awareness.

In expanding the conceptual work concerning the relationship between mental models and team performance and SA, researchers began to realize that although the possession of accurate mental models was a prerequisite for effective team performance and team SA, it might not be sufficient. Specifically, it has been argued that not only must members hold accurate mental models, but that it is the sharing of mutual mental models among members – or ***shared mental models*** – that allows for effective coordinated and adaptive team behavior (see Converse et al., 1991; Orasanu, 1990; Cannon-Bowers & Salas, 1990). Graham (2005) shows that this sharing is particularly critical if military units are to be adaptive. This may be especially difficult for multinational teams because cultural differences can place obstacles to information exchange needed to develop these models. For example, a person from a culture with strong power distance beliefs may not feel comfortable presenting his or her skill set to a supervisor.

Shared and team mental models have been shown to impact team effectiveness in a number of studies (e.g., Mathieu et al., 2000; Bolsted & Endsley 1999). However, defining and measuring team and shared mental models has been the subject of a great deal of research and debate in recent years (Banks & Millward, 2000; Cooke, et al. 2000; Cooke, Salas, Kiekel, & Bell, 2004; Langan-Fox et al., 2001; Espinosa et al., 2001; Warner et al., 2004; Graham 2005). Indeed, shared mental models have engendered some of the same concerns and complications of knowledge distribution raised with team and shared SA. Cooke has developed knowledge metrics specifically for heterogeneous teams that take team position or role into account.

There are several approaches for measuring team mental models that have been evaluated in the organizational theory and psychological literature, including pathfinder analysis, multidimensional scaling, interactively elicited cognitive mapping, and text-based causal mapping (for a review, see Mohammed et al., 2000). Each of these approaches has strengths and weaknesses that make them appropriate for different domains and purposes. Powerful representations of mental models can be created that rely on post-hoc elicitation of the subjective relationships among pertinent concepts. However, though such methods are appropriate for academic inquiries into the meaning of mental models, their operational utility in an automated tool is questionable in the context of network centric warfare, in which military officers require real-time information about the knowledge state of organizational components. Approaches that rely on invasive inquires, such as in-mission probes, are not realistic for operational use. These methods that rely on retrospective responses are also of questionable relevance to the assessment of team SA which is thought to evolve in real-time. Finally, these mental model elicitation methods focus on individual knowledge and its similarity across team members. In network-centric warfare and similar collaborations there is rich information in the interactions among team members that should not be overlooked.

Network Text Analysis

The communications analysis approach taken in the current evaluation, Network Text Analysis (NTA), utilizes the information contained in normal collaboration activities, and extracts from this a series of interlocked networks that contain information that can be used to assess SA. This data can then be assessed using dynamic network analysis (DNA) techniques to estimate SA and to provide insight, at the organizational level, into SA and other aspects of organizational health and risk. In NTA, semantic maps are created for each sub-group within an organization. Semantic maps are networks in which the nodes are the concepts discussed and the connections between nodes signify the strength of association between those concepts, and are enriched by meta-data about the individuals/groups authoring the materials (e.g., time composed, references accessed, actions taken/documented). A comparison of networks based on text from several individuals or groups yields information about the similarity of their respective mental models. Differences among maps may reflect misalignments of SA, which can be remedied by information sharing and targeted communication. However, such misalignments have to be viewed in the context of what work is being doing.

The purpose this initial effort was to explore the potential of NTA as extracted via AutoMap to represent and compare semantic maps derived from intra-team communication. The basic stages required for an integrated tool were identified and followed in an initial demonstration-of-concept exercise. From this exercise, we are able to develop metrics for comparison of texts, and to recognize the avenues for further development and research.

To explore the potential of this approach and technology to assessing situational awareness, representative communications and report data were captured using Groove Virtual Office (GVO) and analyzed using AutoMap (Diesner and Carley, 2004) and ORA (Carley and Reminga, 2004). An initial evaluation exercise was conducted in which several small teams served as surrogates for the command architectures in use or proposed for use in FORCEnet enabled organizations. In a distributed collaborative environment consistent with FORCEnet concepts, participants were presented with the operationally realistic scenario called NEO: Red-Cross Scenario (Warner, Wroblewski, and Shuck, 2003). The task was to develop a course of action (COA) to counteract the events described in the scenario, in this case the evacuation of non-combatants from a non-permissive environment. For this evaluation exercise, participants were divided into several "expert groups", and granted access to different sub-sets of the reference materials designed to assist participants in developing the COA. All collaboration and communication between them was mediated by the GVO, which supports a variety of computer mediated interaction. The occurrence, content, and context of communication among participants were then analyzed to evaluate the planning process and to develop and test specific measures of team and shared knowledge. These results serve as the foundation for the continuing development of an automated toolkit, named ***IMAGES: The Instrument for Measuring and Advancing Group Environmental Situational awareness.***

Methods

Participants

Six employees (3 women and 3 men) of a human-centered engineering consulting firm participated in this exercise. Average age of participants was 37.5 years (range 25-49). Participants had a high level of education (highest degree earned: Baccalaureate: 1; Masters: 3; Ph.D.: 2). None of the participants had experience in the armed forces. All participants received payment for their involvement.

Scenario

The scenario used was the Noncombatant Evacuation Operation (NEO): Red Cross Rescue Scenario developed by the Office of Naval Research (Warner, Wroblewski, and Shuck, 2003). This scenario was ideal in that its supporting materials include both information to be shared among all exercise participants (concerning the locale, enemy forces, and friendly assets) and knowledge that could be uniquely held by a subset of the participants (for intelligence, weapons, and environmental experts). It affords sufficient experimental control while maintaining operational realism and complexity.

In this scenario, three groups of experts are presented with the task of planning a course of action (COA) designed to rescue wounded Red Cross personnel hiding in a church on a war-torn island. There are time pressures due to the worker's medical conditions, as well as logistical constraints due to the political unrest and geography of the island. Text based briefs were available on which to base the COA (Table 1). Team members are then assigned specialized roles – environmental expert, weapons expert, or intelligence expert. All team members were given common background materials (an Initial Brief and

Background Brief) and a set of maps, as well as information specific to their particular expertise (i.e., an Intelligence brief for the intelligence experts, an Environmental Brief for the environmental experts, etc.). The purpose of this manipulation was to replicate the knowledge disparity that could potentially exist in a FORCEnet environment, and to simulate asymmetries in knowledge among team members.

Table 1: Briefs presented in IMAGES Evaluation Exercise.

Brief Name	Size (Words)	Size (Pages)	Presentation Type
Initial Brief	430 Words	One Page	Welcome Screen
Background Brief	1320 Words	Eight Pages	Microsoft Word Document
Enviornmental Expert Brief	151 Words	One Page	Microsoft Word Document
Intelligence Expert Brief	176 Words	One Page	Microsoft Word Document
Weapons Expert Brief	655 Words	Four Pages	Microsoft Word Document

The original scenario was slightly modified to meet the experimental requirements of the current effort. The Initial Brief was expanded slightly, and customized with exercise-specific information (i.e., experimenter contact information, GVO instructions, communications protocols, and guidance regarding time structure). To take maximum advantage of GVO's capabilities, the briefs were reformatted to meet exercise specifications.

Procedure

Team members were distributed at computer terminals in separate rooms. All communication and note-taking was mediated by GVO; there was no voice communication or handwritten personal notes. This facilitated subsequent NTA.

The evaluation exercise was divided in to two phases (Figures 2 and 3). During the Phase 1, the six participants were divided into three groups of two. Each sub-team served as the experts in a particular specialty – Intelligence, Environment, and Weapons. All participants were provided with a common core background brief as well as an expert brief that they alone received. The three expert sub-teams communicated in segregated GVO workspaces; the sub-teams could not communicate with each other. Participants had 75 minutes to complete review the briefs, discuss and agree on a COA, and author a small *Interim COA Report* that described the COA and the assets required. Roughly, participants had ~15 minutes to review the background materials, 45 to develop the COA, and 15 minutes to author the report.

During the Phase 2 of the exercise, participants were re-assigned to two, three-member sub-groups (Figure 2). These Phase 2 sub-groups consisted of a single Intelligence expert, Environmental expert, and Weapons expert. Each participant first independently read and evaluated the *Interim COA Reports* developed by the teams in Phase 1. They then reviewed the background materials and available briefs. In Phase 2, all briefs were available to all participants. As in Phase 1, participants were asked to develop a COA (*Final COA Report*). Roughly, participants had ~15 minutes to evaluate the *Interim COA Reports* and familiarize themselves with the background materials, 45 to develop the COA, and 15 minutes to author the *Final COA Report.*

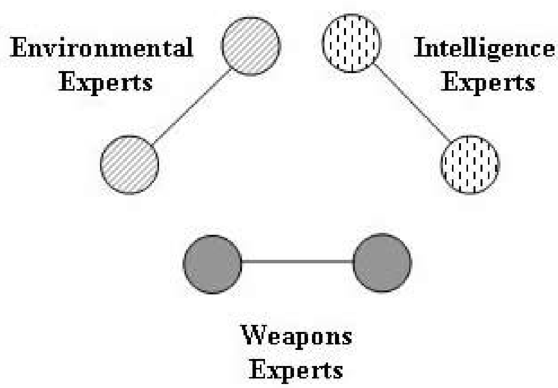

Figure 2: Exercise Design Phase 1. Nodes indicate exercise participants; arcs indicate communication constraints.

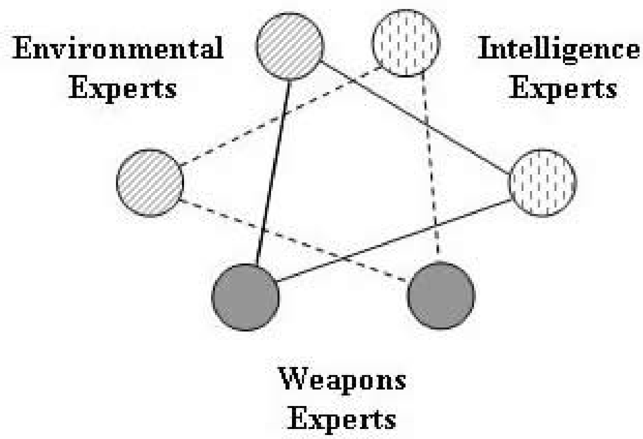

Figure 3: Exercise Design Phase 2. Nodes indicate exercise participants linked in to teams. Solid lines indicate membership in Team A; dashed lines indicate Team B.

All COA reports conformed to a single format. Participants were asked to specify details of the rescue: the Personnel, Transportation, Weapons, and Critical Times for their COA. They were also asked to produce a Detailed Plan in narrative form. The quality of interim and final COAs were measured with reference to the expert solutions accompanying the Warner et al (2003) NEO scenario.

Based on the methodology described above, several texts were generated (Table 2). These were used as data in subsequent semantic map analysis.

Table 2: Material Generated during IMAGES Evaluation Exercise.

Material Source	Material Type
Phase I Environmental Experts	Chat Transcript, Discussion, and Interim COA
Phase I Intelligence Experts	Chat Transcript, Discussion, and Interim COA
Phase I Weapons Experts	Chat Transcript, Discussion, and Interim COA
Phase II Team A	Chat Transcript, Discussion, and Interim COA
Phase II Team B	Chat Transcript, Discussion, and Interim COA

The primary purpose of this evaluation exercise was to examine the potential of semantic maps – based upon inter- and intra-team communication and reporting – to serve as quantifiable proxies for individual, shared, and team mental models. We performed NTA on the corpus consisting of the original source material (Table 1) and the generated materials (Table 2) with the AutoMap software (Diesner and Carley, 2004).

Before AutoMap codes the texts, the analyst has to make coding choices about the identification of the concepts that are relevant for a certain content domain or research question (pre-processing), and the specification of the conditions under which the program links the relevant concepts into statements (statement formation) (Carley, 1993; Carley, 1997a). The pre-processing rules and the statement-formation rules together form the coding scheme. Several types of pre-processing were made, including anaphor replacement, manual header removal, non-content word deletion, and thesaurus application. After these steps were completed, we built a meta-matrix thesaurus that classified 138 concepts relevant for our research question in the pre-processed texts into entity classes of the meta-matrix. The application of a meta-matrix thesaurus enables the ontology-based extraction of social and organizational structure reflected in texts (Diesner and Carley, 2005). The meta-matrix also afforded concept filtering. The domain-specific concepts considered to be the most pertinent for analysis could be isolated and selected for further analyses.

Network Analysis

After pre-processing the data, we specified the statement formation settings in AutoMap that determine how concepts are linked to statements (for detailed information about coding choices in AutoMap and their impact on map analysis results see Diesner and Carley, 2004). We used entire texts as a coding unit and a window size of two. The window size is the maximum distance of concepts that will be linked into statements if they match the pre-processing scheme. We defined rhetorical adjacency for the delete list and the meta-matrix thesaurus. The rhetorical adjacency maintains the original distance of terms by filling gaps that were caused by disregarding terms throughout the pre-processing stages with imaginary place holders. After pre-processing the data and specifying the statement formation settings, a Map Analysis was conducted for the text files generated in the exercise.

From each text, AutoMap extracted one semantic map. The maps were output in DL format in order to visualize them with NetDraw (Borgatti, 2002). Figure 4 shows a representation of the semantic structure within the weapons brief. This figure illustrates the connections among concepts, demonstrates the directionality of the link (i.e., their original order in the text). This text focuses on the Army, Special Forces, and the particular weapons that are available and particularly relevant to this mission.

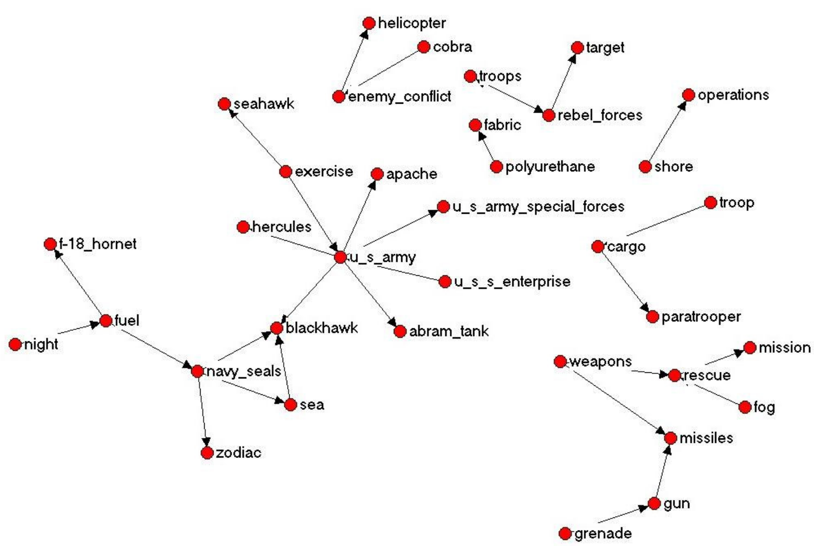

Figure 4: Visualization of Semantic Map of the Weapons Brief.

After extracting the semantic maps, we compared them in order to find out which concepts and statements were shared across the team members and teams, and which were not. CompareMap, a tool connected to AutoMap, supports this functionality (Diesner and Carley, 2004). CompareMap is based on set theory. The tool identifies the union of multiple maps as well as the consensus, represented by the intersection, and dissension (or differences) between maps. AutoMap also produced CSV files which listed the frequency of mapped concepts and statements in each text file. Based on this AutoMap output, the intersection and union of concepts and statements could be determined. We construed the union of multiple maps as a measurement of team mental model (Figure 5, [A ∪ B]). This represents the wealth of information and ideas available to the entire team by virtue of being held by at least one team member. Similarly, the intersection of multiple maps served as the shared mental model (Figure 5, [A ∩ B]). We argue that shared mental models represent the least common denominator of information that all team members or teams possess.

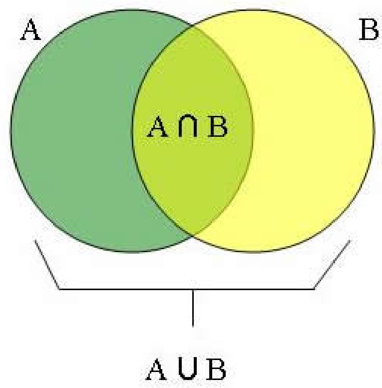

Figure 5: Illustration of the relationship of two semantic maps (A & B). The intersection [A ∩ B] represents the shared mental model; the union [A ∪ B] represents the team mental model.

To compare the texts, we captured the maps representing the team (union) and the shared (intersection) mental models. For each comparison of interest, we calculated the proportion of the intersection to the union of two texts.

$$\frac{A \bigcap B}{A \bigcup B}$$

Equation 1: Proportion of overlap between two texts.

The resulting proportion illustrates the relationship of the shared mental model to the aggregate team mental model, and is normalized to the overall size of the two texts. This proportion was calculated for the unique concepts, total concepts, unique statements, and total statements.

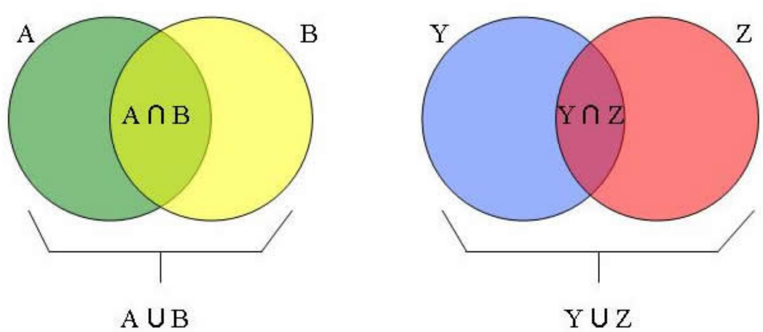

Figure 6: Comparing the overlap between two pairs of text.

Based on the materials given to each team at each exercise phase, we made predictions regarding the relative overlaps of the semantic maps created from the original NEO materials and the materials generated by the teams in each phase (Tables 1 & 2). Our predictions concern the difference between these overlaps in two pairs of texts. That is, do texts A&B overlap more than texts Y&Z (Figure 6), and by how much? This difference was considered meaningful when it exceeded a criterion, *C* (Equation 2). One of the goals of the IMAGES phase I effort was to determine how to choose a meaningful criteria. Based on the limited amount of text and similarity in tasking of the groups, *C* was set at 2.5%.

$$\frac{A \bigcap B}{A \bigcup B} - \frac{Y \bigcap Z}{Y \bigcup Z} \geq C$$

Equation 2: Differences in overlap among sets of texts.

Two analyses were completed, differing by the pre-processing steps taken (Table 3). The motivation for this was to replicate two potential IMAGES use cases. The first analysis used only text manipulation that would not require domain-specific pre-processing, which would be unsustainable in a mature IMAGES tool; the delete list and generalization thesaurus were used, but no anaphor replacement or meta-matrix thesaurus. In the second analysis, all pre-processing steps were completed, which was expected to reduce noise and improve the interpretability of the semantic maps.

Table 3: Parameters used in semantic map analyses.

	Anaphor Replaced	Meta-Matrix Thesaurus	Number of Concepts	Number of Statements
Analysis 1	No	No	1657	4222
Analysis 2	Yes	Yes	137	188

Results

There were three primary predictions, described below. Significant differences are noted in bold. Recall that concepts are words of interest in the texts. Statements were co-locations of two concepts in the text. "Unique" refers to the categories of concepts or statements that are shared between two texts; "Total" refers to the sum of times two texts share a concept. For example, two texts (A & B) share a single concept "Helicopter." Text A has 5 instances of "Helicopter," Text B 3 instances. In this case, the number of unique concepts shared by the two texts is 1, while the number of total concepts is 3 (i.e., the minimum between the two texts).

Prediction 1

We predicted that the sets of materials generated by the Phase I teams of specialists should have greater overlap with the specialized materials they had access to (i.e., the background brief, initial brief, and a single expert brief) than with the materials presented to the other, specialized Phase 1 teams. Each Phase 1 team was given a subset of the NEO materials corresponding to their named specialty (e.g., intelligence materials for the intelligence team), and did not receive two of the extra briefs (e.g., weapons and environmental). As a consequence, there was certain information that they are not privy to, and were consequently unlikely to discuss in their communications or in their COA. The semantic maps will reflect this.

The prediction that the materials generated by the three Phase 1 teams had a larger overlap with the materials each team received – compared to the materials received by the other expert groups – was not supported. Overlaps were largely marginal; the overlap was no larger for the congruent materials than for the incongruent. No consistent differences were found among Unique Concepts or Unique Statements.

One reason for the null results may be the relative size of the specialized background materials, relative to the general briefs. All participant teams received identical Initial and Background briefs, in addition to relatively short expert briefs. The effect of the general briefs may have overwhelmed the effect of the short specialist briefs. Thus, the concepts discussed in each Phase I group would be highly related to the general brief and highly similar. The only exception to this was the Weapons Brief, which was considerably longer than the other expert briefs (Table 1). When the meta-matrix thesaurus was used to constrain analysis 2, the overlap in total concepts between the weapons specialists and the weapons brief was greater than the overlaps between the weapons specialists and the other briefs (2.67% greater)

Prediction 2

We predicted that the materials (chat, discussion, and final COA report) generated by the multi-specialist, Phase II Teams would overlap more with the original materials (Background Brief, Initial Brief, and Expert Briefs) than would the materials (chat, discussion, and interim COA reports) generated by the Phase I teams. Recall that each of the Phase I teams had a different, specialized subset of the total NEO materials; none had access to all materials. These teams could not have discussed certain concepts because they had no knowledge of them. In contrast, both Phase II teams had access to all materials, and thus could discuss all concepts included in the original NEO materials.

The results reveal a trend supporting this prediction. In Analysis 1, the total number of concepts in common for the Phase II teams exceeded those of the Phase I teams, chiefly for Phase II team A. When anaphor replacement and a constrained thesaurus were employed in Analysis 2, this trend weakened. In particular, the Phase I communications and COA had a larger degree of overlap of pertinent concepts and statements than did the Phase II communications and COA. It is unclear if the anaphor replacement is clarifying or strengthening the analyses; for this prediction the results were highly variable when anaphor replacement was employed. No differences were found among Unique Concepts or Unique Statements.

Prediction 3

We hypothesized that the overlap between the materials produced by the two, multi-specialist Phase II Teams would be greater than the overlap between any comparison of the materials generated by the specialized, Phase I teams. The two Phase II teams had access to identical source material, and should consequently have a moderate degree of overlap. In contrast, the three Phase I teams had access to different data sources, and should therefore contain fewer shared concepts.

When the meta-matrix thesaurus was used to focus Analysis 2, we found a larger overlap in total statements between two the Phase II Teams than for any pairing of the Phase I teams. This indicates that the two Phase II teams were discussing the same concepts in close proximity to each other. This is evidence of shared mental models where shared mental models were predicted, and of the validity of this measure and method for measuring such models.

Results and Discussion

The NTA semantic map creation and comparison techniques and the DNA analysis techniques were used to measure the similarity among texts to gain insights into the mental models of participants. In many cases, our predictions concerning the effects of experimental manipulation on those measures were borne out: the overlaps showed the expected pattern. The analyses suggest that the materials generated by multi-specialist teams in Phase 2 displayed a higher degree of overlap with the full set of NEO briefs than did materials generated by specialized Phase 1 teams. We interpret this to mean that the mental models of the Phase 2 team in greater alignment with the source materials than the Phase 1 teams. Similarly, the two Phase 2 teams exhibited higher overlap with each other than most pairings of the Phase 1 experts. This suggests that they had a higher degree of knowledge in common; that is, shared mental models.

These results underscore the potential for IMAGES – and the communications analysis approaches supported by IMAGES, such as AutoMap and ORA, – to **measure**, **monitor**, and **manage** SA based on non-invasive techniques that can be used ubiquitously. Trends indicated that differences could be seen among heterogeneous groups, even in the small scale exercise described above. As the number of individuals interacting increases, tools for extracting networks from communication, like AutoMap, and for assessing the results, like ORA, may be the only feasible approach to assessing SA without negatively impacting mission success by inserting probes into real collaboration activities. However, there were lessons learned in this initial period that will need to be addressed in further research. These include the degree of manual manipulation required for usable analysis, the level of maturity of the metric used, and the need for deeper integration of analysis tools into the collaborative environments in current use.

References

Alberts, D.S., and Hayes, R.E. (2003). *Power to the Edge. Command, Control in the Information Age*. Washington: Command and Control Research Program.

Bakker, R.R. (1987). Knowledge Graphs: Representation and Structuring of Scientific Knowledge. Dissertation. University Twente.

Banks, A.P., and Millward, L.J. (2000). Running shared mental models as a distributed cognitive process. *British Journal of Psychology, 91,* 513-531.

Bordetsky, A. (2005). Update on NPGS Testbed. *Proceedings of the Collaboration and Knowledge Management Workshop 2005*. San Diego, CA.

Borgatti, S.P. (2002). NetDraw: Graph Visualization Software. Harvard: Analytic Technologies.

Burkart, M. (1997). Thesaurus. In M. Buder, W. Rehfeld, T. Seeger, and D. Strauch (Eds.), Grundlagen der praktischen Information und Dokumentation: ein Handbuch zur Einführung in die fachliche Informationsarbeit (4th edition) (pp. 160 – 179). München, Germany: Saur.

Carley, K.M. (1988). Formalizing the Social Expert's Knowledge. *Sociological Methods and Research, 17*), 165-232.

Carley, K.M. (1993). Coding Choices for Textual Analysis: A Comparison of Content Analysis and Map Analysis. In P.V. Marsden (Ed.), *Sociological Methodology 1993*, 23, (pp. 75-126). Cambridge, MA: Blackwell Publishers.

Carley, K.M. (1997a). Network Text Analysis: the network position of concepts. In C.W. Roberts (Ed.), *Text analysis for the Social Sciences* (pp. 79-102). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Carley, K.M. (1997b). Extracting team mental models through textual analysis. *Journal of Organizational Behavior, 18,* 533-558

Carley, K.M. (2002). Smart Agents and Organizations of the Future. In L. Lievrouw and S. Livingstone (Eds.), *The Handbook of New Media*, (pp. 206-220). Thousand Oaks, CA: Sage.

Carley, K.M. (2003). Dynamic Network Analysis. In R. Breiger, K.M. Carley, and P. Pattison (Eds.), *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers,* (pp. 133-145). Committee on Human Factors, National Research Council, National Research Council.

Carley, K.M., and Palmquist, M. (1992). Extracting, Representing, and Analyzing Mental Models. *Social Forces, 70*, 601-636.

Carley, K.M. and Reminga, J.(2004) *ORA: Organization Risk Analyzer*. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report CMU-ISRI-04-101.

Collins, A., and Loftus, E. (1975). A spreading activation theory of semantic processing. *Psychological Review, 82,* 407-428.

Cooke, N.J. (2005). Communication in Team Cognition. *Proceedings of the Collaboration and Knowledge Management Workshop 2005*. San Diego, CA.

Cooke, N.J., (2000). Measuring Team Knowledge. *Human Factors, 22,* 151-173.

Corman, S.R., Kuhn, T., McPhee, R.D., and Dooley, K.J. (2002). Studying Complex Discursive Systems: Centering Resonance Analysis of Communication. *Human Communication Research 28*, 157-206.

Cummins, Denise, D.; Kintsch, W.; Reusser, K; Weimer, Rhonda. (1988) The role of understanding in solving word problems. *Cognitive Psychology*. October 1988. V20, no4. pgs. 405-438.

Department of the Army (1994) TRADOC 525-5. http://www-tradoc.monroe.army.mil/tpubs/pams/p525-5toc.htm

Danowski, James A. (1982). A network-based content analysis methodology for computer-mediated communication: An illustration with a computer bulletin board. In R. Bostrom (Ed.), *Communication Yearbook*, (pp. 904-925). New Brunswick, NJ: Ransaction Books.

Diesner, J., and Carley, K.M. (2004). AutoMap1.2 - Extract, analyze, represent, and compare mental models from texts. Carnegie Mellon University, School of Computer Science, Institute for Software Research International, Technical Report CMU-ISRI-04-100.

Diesner, J., and Carley, K.M. (2005). Revealing Social Structure from Texts: Meta-Matrix Test Analysis as a novel method for Network Text Analysis. In V.K. Naravanan and D.J. Armstrong (Eds.), *Causal Mapping for Information Systems and Technology Research: Approaches, Advances, and Illustrations,* (Chapter 4). Harrisburg, PA: Idea Group Publishing.

Dinan, E. (2004). Groove plays role in Iraq transition. *Mass High Tech*, March 15, 2004.

Endsley, M.R. (1995). Towards a theory of Situation Awareness. *Human Factors, 37,* 32-64.
Endsley, M.R., and Garland, D.J. (2000). *Situation Awareness Analysis and Measurement*. Mahway, NJ: Lawrence Earlbaum Associates.
Endsley, M.R., and Jones, W.M. (1997). *Situation awareness, information dominance, and information warfare.* (AL/CF-TR-1997-0156). Wright-Patterson AFB, OH: United States Air Force Armstrong Laboratory.
Entin, E.E and Serfaty, D. (1999). Adaptive team coordination. *Human Factors*, 41, 312-325.
Espinosa, J.A., Kraut, R.E., Lerch, J.F., Slaughter, S.A., Herbsleb, J.D., & Mockus, A. (2001). Shared Mental Models in Large-Scale, Distributed Software Development. *Proceedings of the Twenty-Second International Conference on Information Systems,* New Orleans, LA.
Fukuyama, F. (2000). Social and organizational Consequences of the information revolution. In the Proceedings of an International Conference, R.O. Hundley, R.H. Anderson, T.K. Bikson, J.A. Dewar, J. Green, M. Libicki, and C.R. Neu (Eds.), *CF-154-NIC*, 2000, RAND and the National Defense Research Institute: Washington, DC.
Gilly, W.B., and Gilly, M.C. (1991). Information Processing and Problem Solving: The Migration of Problems Through Formal Positions and Networks of Ties. *Academy of Management Journal. 34*, 918-928.
Groove Networks. (2003). Use of Groove Software in Austere Environments. White Paper. (http://www.groove.net/pdf/GrooveInAustereEnv.pdf).
Groove Networks, 2004. Groove Networks Announces Role In Newly Announced Homeland Security Information Network. Press Release. (http://www.groove.net/release.cfm?pagename=press_feb26_2004).
Isaacs, E.A., & Clark, H.H. (1987). References in conversation between experts and novices. *Journal of Experimental Psychology: General*, 116, 26-37.
James, P. (1992). Knowledge graphs. R.P.van der Riet and R.A. Meersman (Eds.), *Linguistic Instruments in Knowledge Engineering,* (pp. 97-117). Amsterdam: Elsevier.
Johnson-Laird, P. (1983). *Mental Models.* Cambridge, MA: Harvard University Press.
Keel, P. (2005). EWall: Electronic Card Wall. Web Page. (http://ewall.mit.edu/introduction/).
Kintsch, Walter. (1988) The role of knowledge in discourse comprehension: a construction-integration model. *Psychological Review*. v.95, no. 2 pgs. 163-182.
Klein, H. (1997). Classification of Text Analysis Software. In R. Klar and O. Opitz (Eds.), *Classification and knowledge organization: Proceedings of the 20th annual conference of the Gesellschaft für Klassifikation e.V*, (pp. 255-261). Berlin, Germany: Springer.
Kleinnijenhuis, J., de Ridder, J.A., and Rietberg, E.M. (1996). Reasoning in Economic Discourse: An application of the Network Approach in Economic Discourse. In C.W. Roberts (Ed.), *Text analysis for the Social Sciences,* (pp. 191-208). Mahwah, NJ: Lawrence Erlbaum Associates, Inc
Kleinman, D.L. & Serfaty, Daniel (1989). Team Performance assessment in distributed decision-making. In *Proceedings of the Symposium on Interactive Networked Simulation for Training*, Orlando, FL.
Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor?. *Journal of Management, 20*, 403-437.
Krackhardt, D., and Carley, K.M. (1998). A PCANS Model of Structure in Organization. In *Proceedings of the 1998 International Symposium on Command and Control Research and Technology.* Evidence Based Research. (pp. 113-119). Vienna, VA.
Langan-Fox, J., Wirth, A., Code, S., Langfield-Smith, K., & Wirth, A. (2001). Analyzing shared and team mental models. *International Journal of Industrial Ergonomics, 28,* 99-112.
Lambert, L., & Walker, D. (1995). Learning and leading theory: A century in the making. In L. Lambert, D. Walker, D. Zimmerman, J. Cooper, M. Lambert, M. Gardner, & P. Slack, *The constructivist leader* (pp. 1-27). New York: Teachers College Press.

Manning, C.D., & Schütze, H. (1999). *Foundations of Statistical Natural Language Processing.* Cambridge, MA: The MIT Press.

Mathieu, J.E., Heffer, T.S., Goodwin, G.F., Salas, E., & Cannon-Bowers, J.A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85,* 273-283.

Mayo, R.W., & Nathman, J. (2003). FORCEnet: Turning Information into Power. *Naval Institute Proceedings, Sea Power 21 Series—Part V*, http://www.usni.org/Proceedings/Articles03/PROmayo02.htm.

McComb, S. (2005). Mental Model Convergence (MMC). *Proceedings of the Collaboration and Knowledge Management Workshop 2005*. San Diego, CA.

Mohammed, S., Klimoski, R., & Rentsch, J. (2000). The Measurement of Team Mental Models: We Have No Shared Scheme. *Organizational Research Methods, 3,* 123-165.

Navy Warfare Development Command (NWDC). *FORCEnet: Enabling 21st Century Warfare*, http://www.nwdc.navy.mil/Concepts/IKA/ForceNet.aspx.

Pester-DeWan, J, Moore, R.A., and Morrison, J.G. (2003). Knowledge engineering for command and control transformation at United States European Command. Technical Report 1912, SPAWAR Systems Center San Diego, San Diego, CA.

Popping, R. (2000). *Computer-assisted Text Analysis.* Thousand Oaks, London: Sage Publications

Popping, R. (2003). Knowledge graphs and Network Text Analysis. *Social Science Information, 42*, 91-106.

Popping, R., and Roberts, C.W. (1997). Network approaches in text analysis. In R. Klar and O. Opitz (Eds.), *Classification and Knowledge Organization*, (pp. 381-389). Heidelberg: Springer.

Rouse, William B., Cannon-Bowers, Janis A., and Salas, Eduardo. (1992). The role of mental models in team performance in complex systems. *IEEE Transactions on Systems, Man, and Cybernetics.* 22(6), 1296-1308.

Rouse, W. B., & Morris, N. M. (1985). On Looking into the Black Box: Prospects and Limits in the Search for the Mental Models (DTIC AD-A159080): Georgia Institute of Technology

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box; prospects and limits in the search for mental models. *Psychological Bulletin, 100*, 349-363.

Sowa, J.F. (1984). *Conceptual Structures: Information Processing in Mind and Machine.* Reading, MA.

Tsvetovat, M.; J. Reminga and K.M. Carley, DyNetML: Interchange Format for Rich Social Network Data. *CASOS Technical Report*, Carnegie Mellon University, School of Computer Science, Institute for Software Research International, http://reports-archive.adm.cs.cmu.edu/anon/isri2004/abstracts/04-105.html (2004).

Tsvetovat, M.; J. Reminga and K.M. Carley, DyNetML: Interchange Format for Rich Social Network Data. Proceedings of the North American Association for Computational Social and Organizational Science (NAACSOS) Conference, Pittsburgh, PA (2003).

U.S. Navy (2004) U.S. Navy: Expeditionary Strike Group. http://www.chinfo.navy.mil/navpalib/news/.www/esg.html.

Warner, N.W., Letsky, M., & Cowen, M. (2004). Model of Team Collaboration. Unpublished Report. Naval Air Systems Command, Patuxent River, MD.

Warner, N.W., & Wroblewski, E.M. (2004). The Cognitive Processes Used in Team Collaboration During Asynchronous, Distributed Decision Making. *Proceedings of the Command and Control Research and Technology Symposium, San Diego, CA.*

Warner, N.W., & Wroblewski, E.M. (2005). Collaborative Knowledge in Asynchronous Collaboration (CASC). *Proceedings of the Collaboration and Knowledge Management Workshop 2005*. San Diego, CA.

Warner, N.W., Wroblewski, E.M., & Shuck, K. (2003). Noncombatant Evacuation Operation: Red Cross Rescue Scenario. *Proceedings of Collaboration and Knowledge Management Workshop, San Diego, CA.*

Weil, S.A., Tinapple, D., & Woods, D.D. (2004). New Approaches to Overcoming E-mail Overload. *Proceedings of the Human Factors and Ergonomics Society Conference*. New Orleans, LA.
Woods, D.D., Patterson, E.S., & Roth E.M. (2002). Can we ever escape from data overload? A cognitive systems diagnosis. *Cognition, Technology, and Work, 4*, 22-36.