Exploring Cognitive Aspects of Command and Control

Jennifer J. Ockerman, William B. Fitzpatrick, Nathan T. Koterba, Nathan D. Bos, Max J. Thomas, Stephen S. Carr, and L. James Happel

*C*ommand and control (C2) cognitive systems engineering science and technology covers many areas of investigation. In the past, studies at APL have focused on the design of text messaging and TouchTable interfaces, on the impact of fatigue on cognitive performance, and on how the quality of data and the geographic distribution of the team impact collaboration and decision making in C2 environments. In addition, there have been studies investigating the use of physiological sensors to assess cognitive workload and situation awareness in an objective manner. All of these studies have furthered our understanding of how humans cognitively interact with systems and how these interactions affect performance. This article provides a brief overview of these past studies to increase readers' knowledge of previous work and then delves more deeply into the mechanics of recent work to study geographically distributed teams completing a C2 task.

INTRODUCTION AND BACKGROUND

Many command and control (C2) cognitive systems engineering science and technology studies have been completed at APL in the past 5 years. All of these studies have added to the knowledge of how humans are impacted at the cognitive level (i.e., thinking, reasoning, and decision making) by system design, their environment, and their own cognitive state, as well as how these factors impact total system performance. Following is a brief overview of some of these studies.

CollabSpace

CollabSpace was an APL internal research and development (IR&D) study completed in 2006 that investigated how aspects of a text messaging, or chat, environment affect performance on a C2 task. The chat environment enhancements that were examined included providing a task list, providing object links, and combining all chat "rooms" into a single window. A task list pulls all the messages addressed to the user into a separate window and allows the user to respond

directly through the task list. The object links are hyperlinks in the chat messages for any object the integrated geographical display is aware of, allowing a user to quickly locate a discussed object on the geographical display. Instead of putting each chat room in a separate window, messages from all rooms were listed chronologically in a single window to potentially provide context and allow threads to be followed more easily. In addition, the rate at which chat messages arrived was varied during the experimental session, although the rate at which the messages addressed directly to the experimental participants arrived was not varied. The results indicate that each of the enhancements improved performance on at least one type of experimental task but not necessarily on others. In addition, the rate at which chat messages arrived did impact some of the tasks, even when the messages contained information irrelevant to the participant.[1,2]

Fatigue and Situation Awareness

Additionally, in 2006, a study was conducted to investigate the relationship between fatigue and situation awareness. The study had Navy personnel perform a sonar task. The participants were kept awake for 36 hours, during which they were monitored to detect their levels of sleepiness, workload, situation awareness, and task performance. Measurement techniques included surveys, reaction-time tests, batteries of cognitive-ability tests, eye-tracking records, and task performance. The results include correlations between the measures that show that fatigue does affect situation awareness; however, this impact is very individualized, and thus there is a need for dynamic assessment and future work in this area.[3]

Interactive Multi-Touch Visualization

Interactive Multi-Touch Visualization was an IR&D study performed in 2009. This study sought to discover the most effective and/or efficient ways of using the TouchTable technology. A comprehensive understanding of the TouchTable's capabilities and limitations allows potential application designers and system integrators to make informed decisions on the appropriate application of the TouchTable and similar technologies. To gain a better understanding of the TouchTable's capabilities, a pilot study was conducted. The pilot study allowed the experimentation team to exercise and examine the system's critical human factors aspects, which included speed, accuracy, ergonomics, proxemics (i.e., study of spacing between people), collaboration, tactile interaction, visual orientation, occlusion (i.e., blocking), crowding, and clutter, as well as multiperson and multifinger interaction. The pilot study allowed trials for all conditions to be successfully set up and run. The results of

the pilot study indicated that the TouchTable's interface was somewhat intuitive and user acceptance was high. However, great care should be taken when implementing on touch interfaces applications that require speed and accuracy and do not tolerate errors.[4]

Data Quality Versus Decision Quality

As part of APL's Command and Control Cross Enterprise Initiative (C2 CEI) in 2007, a study was conducted to investigate the impact of data quality on C2 decision making. The data quality varied across three dimensions: consistency, timeliness, and relevancy. The study consisted of a team of six participants playing the different roles in a global strike engagement scenario. Final results of the study, though preliminary, showed that data consistency, and data consistency interacting with timeliness, had the most impact on decision accuracy, with more consistent and timely data leading to more accurate decisions. In addition, it was shown that the mental workload ratings of mental demand and temporal demand were higher when data consistency was lower and that frustration increased as the timeliness and the timeliness–consistency interaction decreased.[5]

SmartProxy

During the following year of APL's C2 CEI effort, an experiment was performed that combined an APL-developed technology called SmartProxy, which has the ability to represent and manipulate a network's bandwidth and connectivity while inserting mitigation strategies for disadvantaged users (i.e., users with occasional connectivity, limited bandwidth, or both), with human performance measures. SmartProxy technology seeks to optimize information flow to disadvantaged users (e.g., submarines) by prioritizing the types of messages received by disadvantaged users during times of limited bandwidth or connectivity. For example, if a boat has occasional connectivity, SmartProxy could be set to filter the messages the boat receives to only those coming from a specific area of interest so that the boat is not inundated with data from distant locations having no bearing on its current mission. This approach could help ensure that the limited data the boat receives are more likely to be pertinent. Participants performed a route-planning task under different network conditions, and the human factors team collected cognitive measurements including physiological (e.g., electroencephalogram or brain waves, electrocardiogram or heart rate, galvanic skin response or sweating, and eye tracking), psychological, and subjective performance data. As shown by the analysis of the cognitive measurements, the vast majority of participants using the SmartProxy technology demonstrated improved

performance (McKneely, J., and Frey, T., "Smart Proxy/Augmented Cognition (ProxyCog) Experiment Update–15 Sept 2008," APL internal presentation).

Hidden Profile Tasks

In addition, during that same year (2008), a study was conducted to examine information sharing and collaboration in a C2 environment. The study used hidden profiles tasks, which are tasks in which members of a team must solve a problem using information that is distributed among team members in such a manner that no single team member has all of the information; team members must collaborate and coordinate to solve the problem. Using a simulated carrier strike planning task, with coordination between three team members, the study suggested that the hidden profile protocol supports the understanding and analysis of C2 information sharing and collaboration.[6]

Summary

As one can see from these brief descriptions, the science and technology efforts aimed at understanding the cognitive aspects of C2 are quite varied in terms of their objectives and results; however, they all use the scientific method of generating one or more hypotheses, developing a study (and often a study environment) along with an experimental design to test those hypotheses, collecting data, and then analyzing the collected data to determine how well they support the hypotheses. The rest of this article describes, by way of a case study of recent work that continued the use of the hidden profile method, how the early steps in this process are implemented.

Under APL's IR&D program, a project was conducted to create and pilot-test a test environment that included a larger team with multiple decisions to be made under time pressure. The project's goals were to increase understanding of collaboration and group decision making in a geographically distributed team, as well as to explore methods of measuring C2 collaboration and decision making.

CASE STUDY BACKGROUND

Long-distance collaboration has become an increasingly salient aspect of C2 as net-centricity has provided the ability to reach farther and faster to support C2 efforts. It is not always possible or desirable for all relevant experts and decision makers to be collocated during time-sensitive missions. However, geographic distance and the associated problems of narrow communications bandwidth with limited expression channels, lack of trust, and biased attention have not always been taken into account in the design of C2 systems. Time pressure and high-risk decisions exacerbate the problems of coordination and trust over long distance; unfortunately, these are two of the defining features of many C2 tasks.

C2 and Decision Making

Much important work is accomplished by teams rather than individuals; yet our understanding of team decision making lags behind our understanding of individual cognition. Group decisions are affected by organizational structures, task constraints, time constraints, and geographic distribution in ways that are complex enough that experimental research is needed to untangle the overlapping influences.

C2 is a complex team endeavor that exhibits all eight dimensions described for naturalistic decision-making environments:[7]

- **Ill-structured problems:** These require the decision makers to spend a significant amount of time understanding the problem and formulating hypotheses to test (e.g., collecting intelligence to improve situation awareness).

- **Uncertain dynamic environments:** These provide the decision maker with incomplete and imperfect information that changes over time (e.g., almost any real-world environment with multiple actors and natural occurrences).

- **Shifting, ill-defined, or competing goals:** Decision makers must accommodate multiple goals, which change with the environment and might be conflicting (e.g., trading off risk to own troops or civilians and mission effectiveness).

- **Action/feedback loops:** Decisions are ongoing and are used to string together events that reach for a goal (e.g., the observe–orient–decide–act loop).

- **Time stress:** Decision makers have some level of time pressure (e.g., decisions have life spans during which the decisions must be made in order to still be relevant).

- **High stakes:** Decisions have outcomes of real significance to the decision makers and others (e.g., life-and-death decisions).

- **Multiple players:** Decisions are made by multiple people at different levels or collaboratively (e.g., decisions made at different echelons need to be coordinated).

- **Organizational goals and norms:** Decision makers must operate within an organization that has goals and rules beyond the personal preferences of the decision makers (e.g., military culture and rules of engagement).

In addition to these eight characteristics, C2 teams are often distributed, with technology bringing together their expertise from wherever it is located to wherever it is needed. For example, teams planning global strike missions need to be able to effectively reach forward, out, and back to regional knowledge, munitions, and intelligence experts having unique information about the targets and area of engagement. Once they have the information, the teams must combine it in useful and meaningful ways to support decision making.

Geographically Distributed Collaboration

To maximize effectiveness, a team should be able to reach back to the best-available expertise, regardless of distance. A commander should be able to put together a team across vast distances, bringing together staff across forward-deployment sites, reaching back to expertise across the world, and getting data from on-site human information sources. This is becoming more and more technologically feasible. However, there is a well-known set of human factors issues related to long-distance collaboration that must also be considered but have received little research attention.

Research on long-distance collaboration has shown that even very small differences in location and availability can have large implications for how teams function. For example, in an office research environment, Kraut et al.[8] found that researchers whose offices were located more than 30 m apart were less likely to collaborate, and the effect of distance was a stronger predictor of collaboration than having similar research topics. This seems to be the case because the informal contact and collegiality of being close together makes formal collaboration easier and more likely to be initiated.

There are two overlapping factors at work when collaboration takes place at a distance: the first is the distance itself, and the second is the necessary use of computer- and/or telephone-mediated communications. There is research showing that the mere perception of distance has effects on behavior, including making different types of behavioral attributions to distant collaborators. Time zone and circadian rhythm differences can affect both availability and other social aspects. The two factors also interact when distance causes communications to be delayed or unreliable. However, most of the effects of distance are due to the limitations of the communications channels, not the distance itself. Most laboratory research on distance collaboration (including this study) does not attempt to simulate time zones or choppy communications but focuses instead on the way that mediated communications change the behaviors, attitudes, and interaction patterns of groups that do not have access to face-to-face communications and are restricted to computer-mediated communications (we shall still refer to these groups as "distributed teams").

Research has identified the following key problems that often occur when groups interact at a distance using mediated communications:

- **Coordination difficulty:** Distributed teams may have more difficulty coordinating work. Well-functioning collocated teams rely on a high level of workplace awareness,[9] shared artifacts, and frequent information communication to synchronize work on complex projects.[10] Distributed teams often must use "loosely coupled" rather than "tightly coupled" coordination strategies, i.e., relying on formalized roles to divide and conquer problems. For some tasks, this inhibits effectiveness.

- **Load-balancing failure:** Teams that are collocated have a relatively high level of awareness of team members' current states and levels of workload. When one team member is overwhelmed, well-functioning teams will often "load balance" by taking on aspects of that person's workload, performing peripheral functions on their behalf, and/or reducing distractions. Workload is more difficult to perceive and interpret at a distance (often coming across simply as silence), and distant colleagues often have fewer options for offloading work, which can lead to more frequent breakdowns.

- **Failure to develop trust:** Teams that do not know each other well develop trust more slowly, and the level of trust tends to be more fragile.[11] The mechanism by which face-to-face contact facilitates trust is still somewhat mysterious, but multiple studies have shown that trust is one of the most difficult aspects of teamwork to develop at a distance. Lack of informal, social contact that usually corresponds with collocation is a contributor to lack of trust for longer-term collaborations (Rocco, E., Finholt, T. A., Hofer, E. C., and Herbsleb, J. D., "Out of Sight, Short of Trust," presentation at the Founding Conference of the European Academy of Management, Barcelona, Spain, April 2001). Another related finding is that when coworkers are separated by distance, they make different psychological attributions about each other's behavior.[12] Assessments of reliability and expertise also change; for example, if a colleague fails to return an e-mail or phone call, people are more likely to make a situational attribution for a local colleague ("they must have been busy") but a dispositional attribution for a distant colleague ("they are unreliable"). This finding is a special case of the "fundamental attribution error,"[13] a bias that has been shown in many other settings and can be a strong determiner of affinity and trust.

- **Lack of transactive knowledge:** Transactive knowledge is knowledge about the skills and abilities of other people.[14] Team members tend to have a less accurate map of the capabilities of their distant collaborators. Collocated teams that work together over time develop sophisticated maps of their team members' skills, strengths, weaknesses, and preferences and learn to interpret their nonverbal communication; this knowledge allows teams to operate at a higher level of effectiveness.[15] Opportunities to develop this type of transactive knowledge of the team are usually much more limited at a distance because of both narrower information channels and less frequent communication.

Partially distributed teams magnify these problems. These are teams where part of the group is collocated and part is distributed, joining the team as singletons or small clusters. Fussell et al.[16] found that collaborators had difficulty managing time and attention equitably across projects with different geographic configurations. When involved in both collocated and distributed collaborations, participants favored tasks with collocated partners despite equal importance of tasks. Experimental studies of partially distributed teams have shown similar effects, where collocated team members have a strong communication bias toward paying attention to local colleagues.[17] Collocated teammates may develop shared references and sparse communications shortcuts that are natural and helpful in fully collocated teams[18] but can unintentionally marginalize distant collaborators.

The effects of geographic distribution on team performance and decision making in a C2 environment was the focus of the work presented in this article.

PILOT STUDY DESCRIPTION AND BACKGROUND

The remainder of this article describes the efforts that led to a two-trial pilot study being conducted. The purpose of the trials was to demonstrate data collection techniques, to identify issues associated with selected scenarios, and to provide data for refinement of follow-on, detailed studies. The overall objectives being designed to were the desire to understand how team performance and collaboration are affected by team member distribution during a high-risk, high-time-pressure scenario, as well as which human performance metrics are most diagnostic in this environment.

To facilitate these objectives, the Experimental Laboratory to Investigate Collaboration, Information-sharing, and Trust (ELICIT),[19] a virtual online environment, was altered to emphasize high-risk decisions under time pressure. The resulting environment, ELICIT Multistrike, uses the ELICIT environment as the information-sharing and collaboration tool, a text messaging system

as the communication tool, and a multidecision point, time-sequenced factoid set.

Finally, a desire to understand the commander's trust in the team recommendations led to a factoid design (i.e., an information distribution design) that forced the commander to take the advice of the distant team members over the advice of the collocated team members.

Thus, the overarching study hypotheses were:

- Team performance, collaboration, and information sharing would be poorer when the team was geographically distributed.

- Trust would be stronger among collocated team members.

Unfortunately, both team performance and trust are difficult attributes to measure. Therefore, the study was also designed to examine several different measures to determine which ones are most diagnostic and feasible to use for this type of scenario. Table 1 provides potential measures.

MEASURING TEAM PERFORMANCE

As part of our research preparation, we conducted a review of team performance measures that might be relevant to distributed C2 tasks (this is an adaptation and extension of Natter et al.[20]). Table 1 summarizes a review of team performance measures that focus on six areas: decision-making product, decision-making process, team situation awareness (awareness of the task environment), teamwork situation awareness (awareness of team process), team workload, and team attributes.

As described later in this article, a subset of these measures, and others that came to light as the pilot study was developed, were prepared for the pilot study.

PRIOR USE OF ELICIT

As mentioned earlier, to support our study of geographically distributed collaboration, we developed a variant of the ELICIT task. ELICIT is built and supported by Evidence Based Research, Inc., at the direction of the DoD Office of the Assistant Secretary of Defense/Network and Information Integration (OASD/NII) Command and Control Research Program (CCRP). The original ELICIT is an intelligence analysis task with a single problem that the entire group works on under moderate time pressure. Our adaptation includes multiple tasks ("strikes") with rolling deadlines that compete for both group members' and leaders' attention. A detailed description of our adaptation can be found in the next section, *ELICIT Multistrike Task and Setting.*

ELICIT has been growing in popularity and has been used by a number of different research groups for different purposes. Its original purpose was to study dif-

Table 1. Partial list of team performance measures used in C2 and related research

Name	Type	Description	Background
Decision-making product			
Response time	Objective/quantitative/ subject-matter expert (SME) rated	How quickly the team provides a decision product	Basic experimental human factors measures from the psychological domain that can provide a standard measure of performance with which to compare other measures
Accuracy	Objective/qualitative/ SME rated	How accurate the team's decision product is	
Completeness	Objective/qualitative/ SME rated	How complete (as in reasoning) the team's decision product is	
Decision-making process			
Information or communications flow analysis: chains and patterns[21]	Objective/quantitative/ SME evaluated	MATLAB algorithms to examine communication chains and similarities	Cooke et al.[21] proposed and used these algorithms for teams of three for a relatively constrained and repetitive task
Rigor measure adapted to team problem solving	Subjective/qualitative/ SME rated	Provides a framework to rate the rigor of analysis used by an intelligence analyst	Zelick et al.[22] proposed and tested an individual measure of analysis rigor
Decision-making strategy traces	Objective/qualitative/ SME completed and compared	Trace of the information flow from injection through completion of decision	Often done in team collaboration studies[23]
Decision-making strategy surveys	Subjective/qualitative/ participant reported	Participants report decision-making strategy either through postexperience interviews or on a survey	Many methods and needs to be tailored to domain
Team situation awareness			
Individual by SME observer	Objective/qualitative/ SME rated	Through the use of interviews/probes and/or surveys, during or postevent, an SME observer rates the participant's situation awareness	Many methods have been proposed and used, the best known is the Situation Awareness Global Assessment Technique (SAGAT)[24]
Individual by participant	Subjective/qualitative/ participant reported	Through the use of interviews/probes and/or surveys, during or postevent, participants self-rate their situation awareness	Many methods have been proposed and used, with a popular one being the Situation Awareness Rating Technique (SART)[25]
Team based	Subjective or objective/ qualitative/SME rated or participant reported	Various ways have been proposed to combine individual situation awareness measures to arrive at a team measure	There is still a significant amount of debate within the human factors community as to how this should be done—there is much room for further research
Teamwork situation awareness			
Teamwork analysis by SME observer or team participants	Objective or subjective/ qualitative or quantitative/SME rated or participant reported	SME observes team and reports, or participants respond to interview questions and/or surveys, on various issues with team members understanding the role of other members and/or their current workload	One team participant method of data collection was developed by Aptima researchers[26]
Team workload			
Individual task-performance based	Objective/quantitative/ SME evaluated	Use of the task performance to estimate cognitive workload	Both primary and secondary task performance measures have been used for years to measure cognitive workload
Individual questionnaire based	Subjective/quantitative/ participant reported	Use of questionnaires to allow participants to self-report workload	NASA Task Load Index (TLX), the standard questionnaire for cognitive workload

Table 1.   Partial list of team performance measures used in C2 and related research—*continued*

Name	Type	Description	Background
Team workload—*continued*			
Distribution of workload	Subjective/quantitative/ SME evaluated	Workload distribution charts or diagrams based on individual workload estimates	
Team attributes			
Team trust communications	Objective/quantitative/ participant reported	Communication intensity and ability to cope with technical and task uncertainty	Ratcheva and Vyakarnam[27] and Jarvenpaa and Leidner[28]
Cognition-based trust questionnaire	Subjective/quantitative/ participant reported	Ask participants to report on cognition-based trust of other team members (e.g., "I see no reason to doubt my teammates' competence and preparation for the job")	Cognition-based trust has been shown to positively impact performance of virtual teams[29]
Affect-based trust questionnaire	Subjective/quantitative/ participant reported	Ask participants to report on affect-based trust of other team members (e.g., "I can talk freely to my team about difficulties I am having")	Affect-based trust has not been shown to positively impact performance of virtual teams[29]
Team cohesion questionnaires	Subjective/quantitative and qualitative/participant reported	Focus on presence of a clear, valued, and shared vision; instances of conflict (in messages/dialog); time lag in resolving conflict; mechanism of conflict resolution (resources used, type of persuasion); tone of communication (formal, informal, task based)	Cohesiveness has been shown to positively impact performance of virtual teams working on complex tasks[30]

ferent C2 structures, especially the pros and cons of hierarchical and "edge" (i.e., flat with equal distribution of information to all members) command structures for information sharing. On the whole, the edge organization tends to perform better on information-sharing tasks, but differences in methods have been found between military and civilian, novice and expert, and younger and older participants.[19, 31]

Three aspects of our current study (i.e., time pressure, use of text chat with ELICIT, and exploration of trust issues) have been explored previously using versions of ELICIT.

- Brehmer[32, 33] studied the effects of time pressure and the interaction of time pressure and hierarchy; he found that without a central node (a central leader in this case), local commanders communicate with each other and can outperform a hierarchy under time pressure. Our pilot study aimed to increase the time pressure with multiple deadlines but did not examine the hierarchy versus edge condition; instead, we focused on geographic separation.

- Thunholm et al.[34] investigated the use of text chat with ELICIT. After comparing separate chatrooms with a combined chatroom, they found that a single large chatroom was the most effective in support-

ing ELICIT. Despite this recommendation, our pilot configuration did use multiple chatrooms because this most accurately simulates field settings we have observed.

- Powley and Nissen[35] examined whether trust was important for ELICIT performance and whether interventions to increase trust would improve performance; generally they found trust to be unrelated to performance. This is not too surprising because trust is not relevant to every information-sharing or coordination task. ELICIT's use of reliable information (no false factoids) also eliminates source credibility as a factor, making trust less important. Our study is designed to explore the role of trust further by implementing tasks where trust may be relevant because of conflicting information presented by separate team members and implementing geographic separation, which is known to influence trust.

ELICIT MULTISTRIKE TASK AND SETTING

We sought to develop a version of ELICIT with somewhat different task characteristics than those of the baseline research task.

- **Time-sensitive decision making.** The original ELICIT does have some time pressure in that groups must try to solve a single analytical problem in an hour, but we sought to increase the time pressure and make this more of a factor in both the individual's and commander's decision making.

- **Multiple overlapping analytical problems.** We wanted to study an environment in which the team's attention was divided among a number of tasks that had to be addressed somewhat in parallel; the team, and the commander in particular, would have to multitask and prioritize in addition to solving the analytical problems presented.

There were two reasons for these alterations. First, we wanted to reproduce some of the realistic task constraints of C2 groups such as Air Force time-sensitive targeting groups, Army Tactical Operations Centers (TOCs), or Navy Command Information Centers (CICs). Our SME also noted similarities with subgroups within a larger Air Force Combined Air and Space Operations Center (CAOC). These groups track multiple targets and address issues in parallel, sometimes facing extreme time pressure and tightly grouped deadlines.

The second reason for these additional constraints is that they make the study environment more vulnerable to the known problems of long-distance collaboration. As described in the literature review, many problems of distance result from poor allocation of attention across separated groups, especially under time pressure. We expect that a commander collocated with one section of a team and distant from another may have particular tendencies toward biased attention and skewed decision making.

The adjustments resulted in the ELICIT Multistrike configuration. ELICIT Multistrike and the ELICIT baseline are compared in the following section.

Our geographic configuration is shown in Fig. 1. The team includes seven members. There are two three-member analysis groups (one group assigned to each "country," Tauland and Psiland) and three associated terrorist groups. There is also a single commander who has sole authority to call a strike on targets in either country. The commander was collocated with the Tauland group.

COMPARISON TO BASELINE ELICIT SETUP

The following components were the same for both ELICIT Multistrike and the baseline ELICIT task:

- Factoids are distributed among multiple team members (in this case, everyone but the commander); factoids must be shared (initially a factoid is given to a single team member) and used together to solve analytical problems. There is no specialization of knowledge; any analyst may receive factoids related to any aspect of any terrorist group. Factoids were delivered according to a preset schedule. Figure 2 shows how the set of factoids related to one terrorist group meeting (Green) were distributed among the six analysts.

- Factoids are shared via webpages. We used ELICIT's "Webpage" system to allow analysts to post factoids. In our configuration, there were no restrictions; permissions given to analysts in ELICIT Multistrike were similar to those given to edge organizations in the ELICIT baseline—every analyst could read and post to every webpage. We did change the names of the webpages and added a fourth webpage (which is supported within the system through configuration files); the four webpages in our set were People, Date, Cities, and Addresses.

- ELICIT log files were used to support analysis of when factoids were distributed to individual analysts, when they were posted to publicly accessible websites, and when those websites were accessed by each analyst.

We did not use the following components of the baseline ELICIT task:
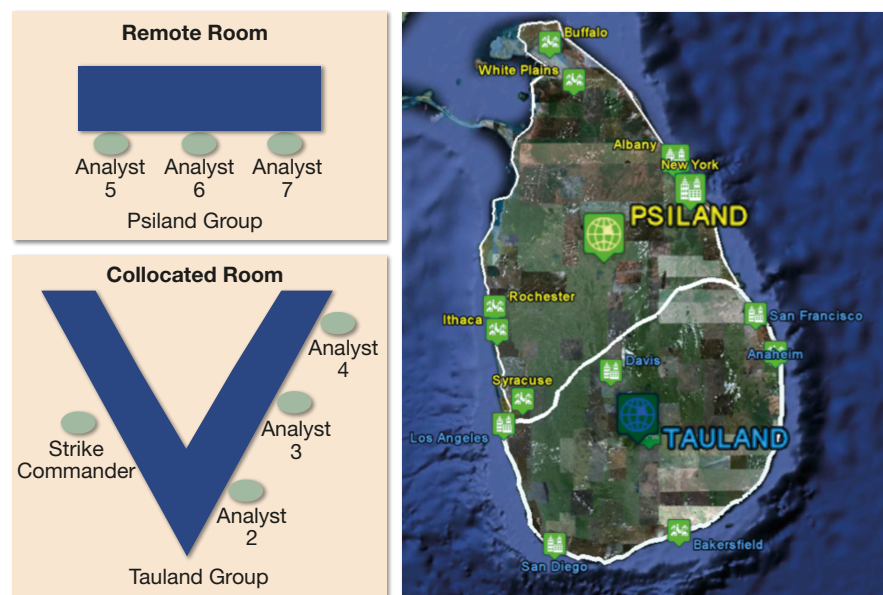

Figure 1. Geographic configuration and assignment of groups and commander. One three-person group was collocated with the commander and was within speaking distance and eye-contact range. The other group was in a "remote" location.
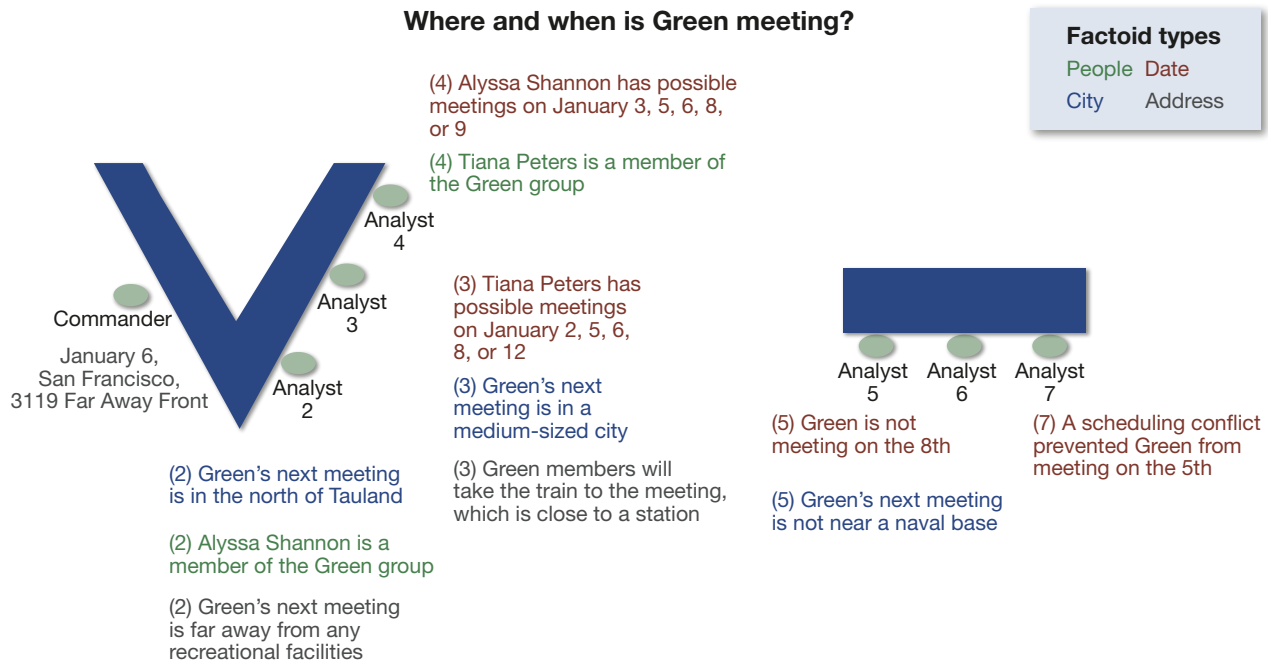
Where and when is Green meeting?

Factoid types
People   Date
City   Address

(4) Alyssa Shannon has possible meetings on January 3, 5, 6, 8, or 9

(4) Tiana Peters is a member of the Green group

Analyst 4

(3) Tiana Peters has possible meetings on January 2, 5, 6, 8, or 12

Commander

Analyst 3

(3) Green's next meeting is in a medium-sized city

January 6,
San Francisco,
3119 Far Away Front

Analyst 2

Analyst 5   Analyst 6   Analyst 7

(5) Green is not meeting on the 8th

(7) A scheduling conflict prevented Green from meeting on the 5th

(3) Green members will take the train to the meeting, which is close to a station

(2) Green's next meeting is in the north of Tauland

(5) Green's next meeting is not near a naval base

(2) Alyssa Shannon is a member of the Green group

(2) Green's next meeting is far away from any recreational facilities

Figure 2. Distribution of factoids necessary to locate and strike the Green group, which is scheduled to meet 6 January in San Francisco. Numbers in parentheses indicate which analyst received the factoid.

- Our analysts were trained to use the webpages for posting factoids but were not trained to use the one-to-one sharing capability. This capability was not disabled, however, and some analysts did use it occasionally.

- We did not use the query capability, which allows the analysts to request additional information; however, we did develop plans to use it in future versions to mimic allocation of scarce intelligence, surveillance, and reconnaissance resources.

The Multistrike task included the following additional features:

- **Chatroom communications.** Use of Internet Relay Chat (IRC) channels is ubiquitous in many real-world C2 environments, and we wanted to provide this capability as well. There were three chatrooms. All seven team members were logged into the commander's chatroom, which was primarily used for communication from and to the commander. Each three-person group also had their own dedicated chatroom (Tauland and Psiland) that included no other members. We anticipated that these group chatrooms would be used for within-team analysis. Analysts were also free to communicate verbally with collocated team members, and the commander and Tauland group could also talk.

- **Multiple parallel analytical problems.** The task facing the teams was this: they were to try to find and disrupt meetings of the six terrorist groups before they could carry out terrorist attacks. They could strike a meeting when they could identify its date, city, and address. Part of the analytical task also required identifying some terrorist group members, necessitating the fourth webpage called People.

- **Rolling deadlines.** The game was fast paced, with each simulated day lasting only 5 min, as indicated by a countdown timer displayed in the front of the room. The task lasted 1 h, or 12 simulated days. Figure 3 shows a timeline of when the terrorist meetings took place over those 12 days. Before the experimental task, teams played a 3-"day" practice session. Factoids were released to the analysts according to a staggered schedule; all information needed to strike a meeting was available to the team at least 1 day before that meeting.

- **Custom factoid sets.** We designed our own factoid sets using the conventions and software provided by ELICIT. The individual analytical tasks were somewhat easier than the original ELICIT problems, requiring less inference and following some regular patterns between problems. It was necessary to make the individual analytical tasks somewhat easier to allow teams some chance of addressing multiple problems in parallel; the combined tasks were still quite challenging. The city and address problems within the tasks also required access to lookup tables; a section of the address table is shown in Fig. 4.

				Blue meeting (Psiland)	Green meeting (Tauland)		Orange meeting (Tauland)	Yellow meeting (Psiland)		Red/Purple joint meeting (Psiland/ Tauland)	
1 Jan 2011	2 Jan 2011	3 Jan 2011	4 Jan 2011	5 Jan 2011	6 Jan 2011	7 Jan 2011	8 Jan 2011	9 Jan 2011	10 Jan 2011	11 Jan 2011	12 Jan 2011

Figure 3. Meeting timeline. A successful strike, with the right date, city, and address, would prevent future group actions.

- **Team scores.** Teams received 100 points every time the commander called a successful strike, which required identifying the date, city, and address of a meeting any time before the end of the simulated day when the meeting was to occur. Teams lost 75 points every time they failed to strike a meeting, which would result in a successful terrorist attack the next day. Teams lost 75 points if they called a strike but had one of the details wrong, resulting in civilian causalities. Strikes were called by the commander using the chatroom; results of successful and unsuccessful strikes and missed meetings were communicated by the game administrator via the commander chatroom. Later review of the pilot study by SMEs has indicated that the scoring rubric may need to be adjusted in future studies to increase realism (e.g., a larger penalty for missing an opportunity to strike).

TEAM PERFORMANCE MEASURES PREPARED FOR THE PILOT STUDY

Several team performance measures were prepared for the pilot study. Each measure has an expected find-ing that corresponds to lower-level hypotheses for the study. The subset comprises a complete set of measures to match the team performance measures categories listed in Table 1, although a few measures of team attributes were not included in the original measure review. Eventually, to meet the objective of determining the most diagnostic and feasible measures, more measures will be used and compared.

Task Performance

Team score, as described above, was used as a measure of decision-making product performance.

Expected Finding

We expect to be able to link specific aspects of team process and attributes, such as information sharing and team cohesion, to differences in team performance. For example, collocated teams might have higher cohesion, more efficient information sharing, and better task performance than a geographically distributed team. With a laboratory task with a known correct answer, the task performance can be used to correlate other measures and determine which measures are most useful.

ID	City	Name	Type of Building	Closest Recreational Facility	Near a Major Highway?	Near a Construction Site?	Age of Structure (years)	Nearest Public Transportation
18	New York	2112 Hopper Street	House	swimming pool	yes	yes	12	Bus stop
19	New York	897 Oak Tree Road	Hotel	none	yes	no	15	Metro
20	New York	553 Runner Circle	Public Building	swimming pool	no	yes	20	Railroad
21	New York	7880 Whiskey Bottom	House	city park	no	yes	25	Railroad
22	New York	4343 Lincoln Lane	House	city park	no	yes	33	Railroad
23	New York	8899 Merganzer Drive	House	swimming pool	no	no	30	Bus stop

Figure 4. Lookup table of addresses.

Information-Sharing Performance

This measure sought to examine how efficient analysts were at posting factoids they received to the correct webpage and calculated the lag time between receiving each factoid and posting it to a webpage.

Expected Finding

We expect that analysts may prioritize factoids related to their own area of operation (Psiland or Tauland) ahead of factoids relevant to the other country group. A measure of this expectation would be the time delay between receiving and posting factoids; a significant difference between own-group-relevant and other-group-relevant factoids might indicate an information-sharing bias.

Commander's Attention (Eyetracking)

A faceLAB eyetracking device was positioned in front of the commander, and the configurations of both the collocated team's and the commander's workspaces were designed so that the eyetracker could track when the commander was looking at each chatroom on his or her computer screen, when the commander was looking across at any of the collocated analysts, and when the commander was viewing the large-screen data displays.

Expected Finding

We anticipate that in a time-sensitive task with high workload, the commander will pay more attention to local teammates and less attention to distant teammates (over a chatroom), even when information from the distant team may be equally or more time critical.

Commander's Attention (Text Analysis)

A second measure of attention would be the amount and quality of chatroom communication (both groups) and verbal communication (collocated group only) between the commander and each of the groups. In addition, the time lag between when the commander receives information from analysts and when the information is acted upon can be measured.

Expected Finding

Again, we anticipate an attention bias by the commander, which would lead to more and higher-quality interaction with the local team and a shorter lag time between when information from the collocated team is received and when it is acted upon as compared with the distant team.

Measures of Workload and Situation Awareness

NASA TLX is a widely used survey measure of workload that is completed by participants after they have completed their task.[36] NASA TLX measures six types of workload: mental demands, physical demands, temporal demands, own performance, effort, and frustration. SART, which is similar to the NASA TLX, is a post-task self-reported measure of situation awareness.[25] Although not as sensitive as assessments done midtask, SART has proven its usefulness in operational settings.

Expected Finding

We anticipate that as workload increases beyond a comfortable level, situation awareness (both team and teamwork) will decrease. In addition, we hypothesize an interaction between distance and workload/situation awareness such that when workload increases and situational awareness diminishes, as measured by NASA TLX and SART, the biasing effects of distance will increase. This will be driven by the "narrowing of attention" under high workload.

Other Measures of Group Function and Affect—Team Attributes

Group Identity

We used a 10-item scale, developed by Henry et al.,[37] that measures sense of group identity. This measure had three subscales: affective (emotional), behavioral, and cognitive identity.

Expected Finding

Higher scores on group identity measures would be expected to correlate with better group performance. Geographic separation tends to disrupt identity formation or lead to strong group identity of local subgroups at the expense of the larger group. Analysis would focus on whether the aggregate identity measure does correlate with performance and whether there is a difference in effect between the subscales, which might imply different importance.

Group Efficacy

We used a three-item scale, adapted from Carroll et al.,[38] on group efficacy over distance. Sample items include:

- Our group worked well together.
- Despite the fact that some people were remote, we worked well together.
- Our group was good at coordinating longer orders.

Expected Findings

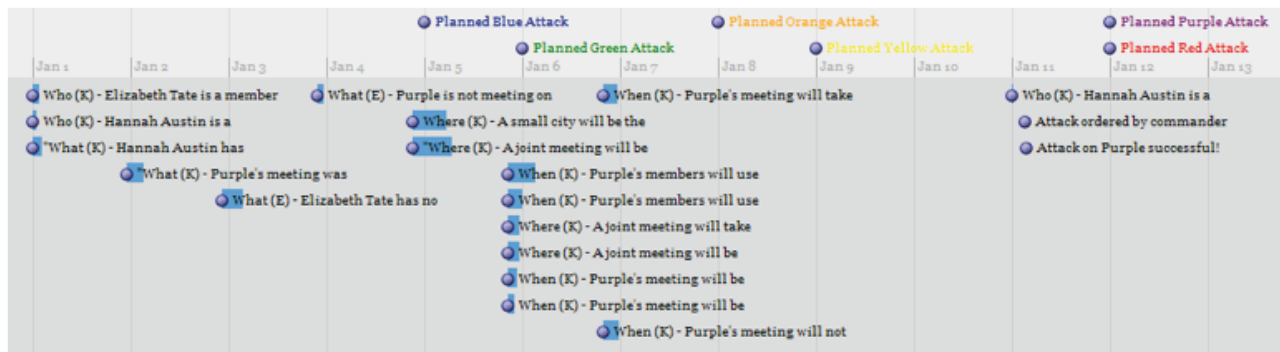Effective teams would be expected to have higher self-reported group efficacy.

Figure 5. Schedule of factoid releases related to one terrorist meeting (Purple).

Reciprocity Scale

We used an 11-item scale, based on the work of Perugini et al.,[39] related to personal norms of reciprocity. When groups have strong reciprocity norms, group members expect to be "repaid" when they provide assistance or share information with teammates. Generally, strong reciprocity norms are not optimal for teams, especially in information-sharing tasks, where rapid and free distribution of information is critical for success. Sample items include:

- I went out of my way to help players who had helped me.

- If someone refused to help me, I held a grudge against them.

- I was kind and nice if others behaved well with me; otherwise, it was tit for tat.

Expected Findings

We anticipated that there might be higher reciprocity norms between the groups, and possibly the commander and the distant group, and lower reciprocity norms within the groups as they built stronger rapport.

Social Network Trust and Collaboration

Each player rated every other player on five items, and each item was rated using a five-point Likert scale. This measure was adapted from other social network surveys. Prior research using these scales showed that trust and familiarity tend to vary by location. The five items were as follows:

- I worked closely with this player.

- I trust this player.

- I would like to play with this player again.

- This player was one of the leaders in the group.

- This player was helpful to others.

Expected Findings

We would expect to see higher trust and familiarity ratings for team members that were part of the same country group.

Data from Pilot Study

We were able to conduct two pilot runs of the ELICIT Multistrike task with seven independent volunteers for each run. All volunteers had some familiarity with intelligence analysis and C2 in a military context. We completed the entire experimental protocol with these groups, including training and the 12-day task. We calibrated but have not yet analyzed the eyetracker data; only a few members of the first group had completed the battery of post-task measures when we realized that the list of measures we were using was too long for future use and would need to be condensed.

In the second pilot run, the team called three successful strikes and one erroneous strike and missed three meetings, resulting in a final score of zero. In a real run of this task, we would run three or four sessions with the same teams, in distributed and collocated configurations, and we would expect that learning effects would allow teams to score much better in the later sessions.

We are using the pilot data to experiment with different analyses and visualization methods. One promising direction uses timelines generated with the Massachusetts Institute of Technology SIMILE Timeline widget. The timeline in Fig. 5 shows the release schedule over time for all factoids related to one terrorist meeting (Purple) as a complement to Fig. 2, which focuses on the geographic distribution. This visualization of timeline data, which are shown only in table format in ELICIT, allows event timing and relation to other contextual factors to be reviewed more quickly.

FUTURE DIRECTIONS

The pilot trials were instrumental in identifying issues to be addressed in future work as well as in

confirming that the logistics were feasible and that the experimental task was at an appropriate level of difficulty to stress teams but not overwhelm them. Through a full study, we hope to better understand how geographic separation can impact C2 team decision making and performance in high-stress, multiple-task scenarios. In addition, we would like to determine which team performance measures have the most diagnostic capability while still being easily used and potentially automated. It is through understanding team cognitive behavior and using diagnostic measures that mediations and solutions can be suggested to change technology, techniques, training, and behaviors to improve C2 team performance in a variety of scenarios. The task data can also be made available to other research groups who would like to run ELICIT Multistrike or a variant on their own.

REFERENCES

[1]Tzeng, N., Ockerman, J., Cantu, O., and Koterba, N., "Creation and Use of a Command and Control Collaboration Testbed," in *Proc. International Command and Control Technology and Research Symp.*, San Diego, CA, pp. 1–13 (2006).

[2]Ockerman, J., Tzeng, N., Koterba, N., and Cantu, O., "Experimenting with Enhanced Chat," in *Proc. American Society of Naval Engineers (ASNE) Human Systems Integration Symp. (HSIS)*, Annapolis, MD, pp. 1–16 (2007).

[3]Baumgart, L., McKneely, J., and Comperatore, C., "Using Neural and Physiological Correlates to Detect Loss of Situation Awareness," in *Proc. 4th Int. Conf. on Augmented Cognition*, Baltimore, MD (2007).

[4]Fitzpatrick, W. B., "Interactive Multi-Touch Visualization (IMTV) Pilot Study: A Human Factors Perspective," Technical Memorandum STI-10-010, JHU/APL, Laurel, MD (2009).

[5]North, P. D., Forsythe, S., Frey, T., Iny, M., Ockerman, J., et al., "Experimentation Via the Use of an Executable Workflow Model to Evaluate C2 Decision Quality," in *Proc. International Command and Control Research and Technology Symp.*, Seattle, WA, pp. 1–21 (2008).

[6]Natter, M., Bos, N., Ockerman, J., Happel, J., Abitante, G., and Tzeng, N., "A C2 Hidden Profile Experiment," in *Proc. International Command and Control Research and Technology Symp.*, Washington, DC, pp. 1–27 (2009).

[7]Orasanu, J., and Connolly, T., "The Reinvention of Decision Making," in *Decision Making in Action: Models and Methods*, G. A. Klein, J. Orasanu, R. Calderwood, and C. E. Zsambok (eds.), Ablex Publishing, Norwood, NJ, pp. 3–20 (1993).

[8]Kraut, R. E., Egido, C., and Galegher, J., "Patterns of Contact and Communication in Scientific Research Collaborations," in *Intellectual Teamwork: Social and Technological Foundations of Cooperative*, R. E. Kraut, C. Egido, and J. Galegher (eds.), Lawrence Erlbaum, Hillsdale, NJ, pp. 149–172 (1990).

[9]Gutwin, C., and Greenberg, S., "The Importance of Awareness for Team Cognition in Distributed Collaboration," in *Team Cognition: Understanding the Factors that Drive Process and Performance*, E. Salas and S. M. Fiore (eds.), APA Press, Washington, pp. 177–201 (2004).

[10]Teasley, S. D., Covi, L. A., Krishnan, M. S., and Olson, J. S., "Rapid Software Development Through Team Collocation," *IEEE Trans. Softw. Eng.* **28**(7), 671–683 (2002).

[11]Bos, N. D., Olson, J. S., Gergle, D., Olson, G. M., and Wright, Z., "Effects of Four Computer-Mediated Channels on Trust Development," in *Proc. CHI 2002 Conf. on Human Factors in Computing Systems*, Minneapolis, MN, pp. 1–6 (2002).

[12]Cramton, C., "The Mutual Knowledge Problem and Its Consequences for Dispersed Collaboration," *Organ. Sci.* **12**(3), 346–371 (2001).

[13]Ross, L. D., "The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process," in *Advances in Experimental Social Psychology*, L. Berkowitz (ed.), p. 10 (1977).

[14]Hollingshead, A. B., "Retrieval Processes in Transactive Memory Systems," *J. Pers. Soc. Psychol.* **74**(3), 659–671 (1998).

[15]Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., and Malone, T. W., "Evidence for a Collective Intelligence Factor in the Performance of Human Groups," *Science* **330**(6004), 686–688 (2010).

[16]Fussell, S. R., Kiesler, S., Setlock, L. D., and Weisband, S., "Effects of Instant Messaging on the Management of Multiple Project Trajectories," in *Proc. CHI 2004 Conf. on Human Factors in Computing Systems*, Vienna, Austria, pp. 191–198 (2004).

[17]Bos, N. D., Olson, J. S., Nan, N., Shami, N. S., Hoch, S., and Johnston, E., "Collocation Blindness in Partially Distributed Groups: Is There a Downside to Being Collocated?" in *Proc. CHI 2006 Conf. on Human Factors in Computing Systems*, Montreal, Quebec, Canada, pp. 1313–1321 (2006).

[18]Clark, H. H., and Wilks-Gibbs, D., "Referring As a Collaborative Process," *Cognition* **22**(1), 1–39 (1986).

[19]Ruddy, M., "ELICIT – The Experimental Laboratory for Investigating Collaboration, Information-sharing and Trust," *Proc. International Command and Control Research and Technology Symp.*, Newport, RI, pp. 1–74 (2007).

[20]Natter, M., Ockerman, J., and Baumgart, L., "Review of Cognitive Metrics for C2," *ITEA J.* **31**(2), 179–209 (2010).

[21]Cooke, N. J., Gorman, J. C., and Kiekel, P. A., "Communication as Team-Level Cognitive Processing," in *Macrocognition in Teams: Theories and Methodologies*, M. Letsky, N. Warner, S. Fiore, and C. A. P. Smith (eds.), Ashgate Publishing Ltd., Hants, UK, pp. 51–64 (2008).

[22]Zelick, D., Patterson, E. S., and Woods, D. D., "Measuring Attributes of Rigor in Information Analysis," in *Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams*, E. S. Patterson and J. Miller (eds.), Ashgate, Aldershot, UK, pp. 65–84 (2010).

[23]Bos, N. D, Shami, N. S., Olson, J. S., Cheshin, A., and Nan, N., "In-Group/Out-Group Effects in Distributed Teams: An Experimental Simulation," in *Proc. 2004 ACM Conf. on Computer Supported Cooperative Work*, Chicago, IL, pp. 429–436 (2004).

[24]Endsley, M. R., "Situation Awareness Global Assessment Technique (SAGAT)," in *Proc. Aerospace and Electronics Conf.*, Dayton, OH, pp. 789–795 (1988).

[25]Taylor, R. M., "Situational Awareness Rating Technique (SART): The Development of a Tool for Aircrew Systems Design," in *Proc. AGARD AMP Symp. on Situational Awareness in Aerospace Operations* (1990).

[26]MacMillan, J., Paley, M. J., Entin, E. B., and Entin, E. E., "Questionnaires for Distributed Assessment of Team Mutual Awareness," Chap. 51, *Handbook of Human Factors*, N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. W. Hendrick (eds.), CRC Press, Boca Raton, FL, pp. 484–494 (2004).

[27]Ratcheva, V., and Vyakarnam, S., "Exploring Team Formation Processes in Virtual Partnerships," *Integr. Manufactur. Syst.* **12**(7), 512–523 (2001).

[28]Jarvenpaa, S., and Leidner, D., "Communication and Trust in Global Virtual Teams," *Organ. Sci.* **10**(6), 791–815 (1999).

[29]Kanawattanachai, P., and Yoo, Y., "Dynamic Nature of Trust in Virtual Teams," *J. Stateg. Infor. Syst.* **11**(3–4), 187–213 (2002).

[30]Martins, L. L., Gilson, L. L., and Maynard, M. T., "Virtual Teams: What Do We Know and Where Do We Go from Here?" *J. Manage.* **30**(6), 805–835 (2004).

[31]McEver, J. G., Hayes, R. E., and Martin, D. M., *Exemplar Analyses of ELICIT Experimental Data*, http://www.dodccrp.org/files/elicit_analysis_final.pdf (accessed 25 Jan 2011).

[32]Brehmer, B., "Effects of Time Pressure in Fully Connected and Hierarchical Architectures of Distributed Decision-Making," in *Cooperative Process Management*, Y. Waern (ed.), Taylor & Francis, Inc., Bristol, PA, pp. 131–143 (1998).

[33]Brehmer, B., "Command Without Commanders," in *Proc. 14th International Command and Control Research and Technology Symp.*, Washington, DC, pp. 1–14 (2010).

[34]Thunholm, P., Ng, E. C., Cheah, M., Tan, K. Y., Chua, N., and Chua, C. L., "Exploring Alternative Edge versus Hierarchy C2 Organizations using the ELICIT Platform with Configurable Chat System," *Int. C2 J.* **3**(2) (2009).

[35]Powley, N., and Nissen, M. E., "Trust-Mistrust as a Design Contingency: Laboratory Experimentation in a Counterterrorism Context," *Proc. International Command and Control Research and Technology Symp.*, Washington, DC, pp. 1–40 (June 2009).

[36]Hart, S. G., and Staveland, L. E., "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," in *Advances in Psychology 52: Human Mental Workload,* P. A. Hancock and N. Meshkati (eds.), Elsevier Science Publishers B.V., Amsterdam, pp. 139–178 (1998).

[37]Henry, K. B., Arrow, H., and Carini, B., "A Tripartite Model of Group Identification," *Small Group Res.* **30**(5), 558–581 (1999).

[38]Carroll, J., Rosson, M. B., and Zhou, J., "Collective Efficacy as a Measure of Community," in *Proc. CHI 2005 Conf. on Human Factors in Computing Systems*, Portland, OR, pp. 1–10 (2005).

[39]Perugini, N., Gallucci, M., Presaghi, F., and Ercolani, A. P., "The Personal Norm of Reciprocity," *Eur. J. Pers.* **17**(4), 251–283 (2002).

The Authors

Jennifer J. Ockerman is a senior cognitive systems engineer in the Communications and Networking Systems Group of the Asymmetric Operations Department. Dr. Ockerman served as the Principal Investigator for the Precision Engagement Business Area IR&D project described in this article. **William B. Fitzpatrick** is a senior human factors engineer in the Command and Control Group of the Force Projection Department. Mr. Fitzpatrick served as the Co-Principal Investigator and the Project Manager for the IR&D project described in this article. **Nathan T. Koterba** contributed to the use and recording of the chat messaging system used for the pilot study and performed some initial data analysis. Mr. Koterba is currently employed by V3 Ventures, Inc. **Nathan D. Bos** is a senior research associate in the Research and Exploratory Development Department. He does research in computer-supported collaborative work, with a focus on long-distance collaboration, as well as modeling studies of human social and cultural behavior. **Max J. Thomas** is a member of the Research and Development Department and contributed to the development of software and experimental conditions for the project described in this article. **Stephen S. Carr** is the Branch Supervisor of the Sensor and Communication Branch of the Air and Missile Defense Department. Dr. Carr provided subject-matter expertise for the project described in this article and is currently working on several human factors engineering issues and their concomitant impacts on autonomous systems. **L. James (Jim) Happel** is an APL Principal Professional Staff member and Group Supervisor of the Systems Engineering and Assessments Group of the Force Projection Department. He has served as a special assistant for command, control, communications, computers, intelligence, surveillance, and reconnaissance (C4ISR) for the Global Engagement Department. Mr. Happel has been a proponent of researching the effects of cognitive bias in collaborative C2 operations as well as its effect on the process of engineering systems. For further information on the work reported here, contact Jennifer Ockerman. Her e-mail address is jennifer.ockerman@jhuapl.edu.

The *Johns Hopkins APL Technical Digest* can be accessed electronically at **www.jhuapl.edu/techdigest.**