16th ICCRTS 2011

"Collective C2 in Multinational Civil-Military Operations"

ELICIT Multistrike: Adapting ELICIT to Study Collaboration and Decision Making for Time-Sensitive Strikes

Topics

Topic 5: Collaboration, Shared Awareness, and Decision Making
Topic 6: Experimentation, Metrics, and Analysis
Topic 4: Information and Knowledge Exploitation

Names of Authors

Nathan Bos, William Fitzpatrick, Nathan Koterba, Jennifer Ockerman, Max Thomas, Steve Carr, and Jim Happel
Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road
Laurel, MD 20723

Point of Contact

Dr. Jennifer Ockerman
Johns Hopkins University Applied Physics Laboratory
240-228-0764
jennifer.ockerman@jhuapl.edu

ELICIT Multistrike: Adapting ELICIT to Study Collaboration and Decision Making for Time-Sensitive Strikes

Abstract

Long distance collaboration is an increasingly important aspect of command and control (C2). 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, 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; these are unfortunately two of the defining features of many C2 tasks involving time-sensitive strikes.

In order to more closely examine high-risk decisions under time pressure, the Experimental Laboratory to Investigate Collaboration, Information-sharing, and Trust (ELICIT; Ruddy, 2007), was altered to emphasize these aspects. 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 multi-decision point, time-sequenced factoid set.

This paper describes the design features and pilot testing of ELICIT Multistrike; examines the differences from the well-known original ELICIT task; and discusses the design of an ELICIT Multistrike experiment to demonstrate the ability to use the ELICIT framework to study collaboration and decision making for time-sensitive strikes.

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 organization 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 (Orasanu and Connolly, 1993):

- Ill-structured problems – require the decision makers to spend significant amount of time understanding the problem and formulating hypotheses to test – e.g., collecting intelligence to improve situation awareness
- Uncertain dynamic environments – 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, civilians, and mission effectiveness
- Action/feedback loops – decisions are on-going and used to string together events that reach for a goal – e.g., the Observe, Orient, Decide, Act (OODA) loop
- Time stress – decisions makers have some level of time pressure – e.g., decisions have life spans of when they can be made and still make a difference
- 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, today's 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 (AOE). Once they have the information, they 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 which 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 and Egido (1990) found that researchers whose offices were located more than 30 meters 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 because the informal contact and collegiality of being nearby 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-mediated communications. There is research showing 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 also 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 changes the behaviors, attitudes, and interaction patterns of groups when restricted to computer mediated versus face-to-face communications (we will still refer to these as 'distributed teams').

Research has identified these 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 (Gutwin and Greenberg, 2004), shared artifacts, and frequent information communication to synchronize work on complex projects (Teasley, et al., 2002). 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 level of workload. Well-functioning teams will often perform 'load balancing' when one team member is overwhelmed, either taking on aspects of that person's workload, performing peripheral functions on their behalf, and 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 (Bos, Gergle, Olson and Wright, 2002). 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 for longer- term collaborations (Rocco, Finholt, Hofer, and Herbsleb, 2001). Another related finding is that when coworkers are separated by distance, they make different psychological attributions about each other's behavior (Cramton, 2001). Assessments of reliability and expertise also change; for example, if a colleague fails to return an email 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' (Ross, 1977), a bias which 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 (Hollingshead, 1998). 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 and

weaknesses, preferences, and learn to interpret their nonverbal communication, and this knowledge allows teams to operate at a higher level of effectiveness (Woolley *et al.*, 2010). Opportunities to develop this type of 'transactive' knowledge of the team are usually much more limited at a distance, due to both narrower information channels and less frequent communication.

Partially distributed teams magnify these problems. These are teams where part of the group are collocated and part is distributed, joining the team as singletons or small clusters. Kiesler, Setlock, Scupelli and Weisband (2004) 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 (Bos, Olson, Nan, Shami, Hoch and Johnston, 2006). Collocated teammates may develop shared references and sparse communications shortcuts that are natural and helpful in fully collocated teams (Clark and Wilks-Gibbs, 1986) but that can unintentionally marginalize distant collaborators.

The effects of geographic distribution on team performance and decision making in a command and control environment is the focus of this paper.

Pilot Study Description and Background

The remainder of this paper describes the efforts that led to a two-trial pilot study being conducted. Each pilot study trial was successful in what it was testing, but no data were analyzed from the pilot study. 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 this objective, the Experimental Laboratory to Investigate Collaboration, Information-sharing, and Trust (ELICIT; Ruddy, 2007), was altered to emphasize the 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 multi-decision point, time-sequenced factoid set.

Finally, a desire to understand the commander's trust in the team recommendations led to a factoid 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 will be stronger for 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 the ones which are most diagnostic and feasible to use for this type of scenario. A table of potential measures is provided in the next section.

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 extension of Natter, Ockerman and Baumgart , 2009). Table 1 summarizes a review of team performance measures. Measures are described that focus on six areas: Decision Making Products, Decision Making Process, Team Situation Awareness (awareness of the task environment), Teamwork Situation Awareness (awareness of team process), Team Workload, and Team Attributes.

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

Focus	Name	Type	Description	Background
Decision Making Product	Response time	Objective-quantitative-SME rated	how quickly the team provides a decision product	Basic experimental human factor's measures from the psychological domain that can provide a standard measure of performance to compare other measures to
	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 (Cooke & Gorman)	Objective-quantitative-SME evaluated	MatLab algorithms to examine communication chains and similarities	(Cooke, Gorman, and Kiekel, 2008) 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, Patterson, and Woods, 2010) 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 (Bos, et al., 2004)
	Decision making strategy surveys	Subjective-qualitative-participant reported	Participants report decision making strategy either through post-exp 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 post-event, an SME observer rates the participant's	Many methods have been proposed and used, the best known is Situation Awareness Global Assessment Technique

			situation awareness	(SAGAT) (Endsley, 1988)
	Individual by participant	Subjective-qualitative-participant reported	Through the use of interviews/probes and/or surveys, during or post-event, an participants self rate their situation awareness	Many methods have been proposed and used, with a popular one being Situation Awareness Rating Technique (SART) (Taylor, 1990)
	Team-based	Subjective or objective-qualitative-SME rated or participant reported	Various ways have been proposed to combine individual SA 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 - lots of 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 (MacMillan, Paley, Entin, and Entin, 2004)
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-TLX (Task Load Index) the standard questionnaire for cognitive workload
	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, 2001 Jarvenpaa and Leidner, 1999
	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 (Kanawattanachai and Yoo, 2002)

	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 (Kanawattanachai and Yoo, 2002)
	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 (Martins, Gilson, and Maynard, 2004)

As described later in the text, a subset of these measures was used in the pilot research.

Prior use of ELICIT

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. 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 (see next section for details).

ELICIT has been growing in popularity, and has been used by a number of different research groups with different purposes. Its original purpose was to study different C2 structures, especially the pros and cons of hierarchical and "edge" command structures for information sharing. On the whole, the edge organization tends to perform better in information sharing tasks, but differences in methods have been found between military and civilian, novice and expert, and younger and older participants (Ruddy, 2007; McEver, Hayes, and Martin, 2007).

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 (1998; 2009) 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 study will increase the time pressure with multiple deadlines but will not be examining the hierarchy versus edge condition; instead we focus on geographic separation.
- Thunholm, Ng, Cheah, Tan, Chua and Chua (2009) investigated the use of text chat with ELICIT. After comparing separate versus combined chatroom, they found that a single large chatroom was the most effective in supporting ELICIT. Despite this

recommendation, our pilot configuration did use multiple chatrooms, as this most accurately simulates field settings we have observed.

- Powley and Nissen (2009) 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, since 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 1) implementing tasks where trust may be relevant because of conflicting information presented by separate groups, and 2) 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 the baseline research task.

1. 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.
2. Multiple overlapping analytical problems. We wanted a study environment where the team's attention was divided between a number of tasks that had to be addressed somewhat in parallel; the team, and the commander in particular, would have to multi-task 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 Center (CICs). Our subject-matter expert also noted similarities to subgroups within a larger Air Force Combined Air 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 ELICIT baseline are compared in the following section.

Our geographic configuration is shown in Figure 1. The team includes seven members. There are two three-member groups of analysis, each team assigned to one "country" (Figure 2) 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.

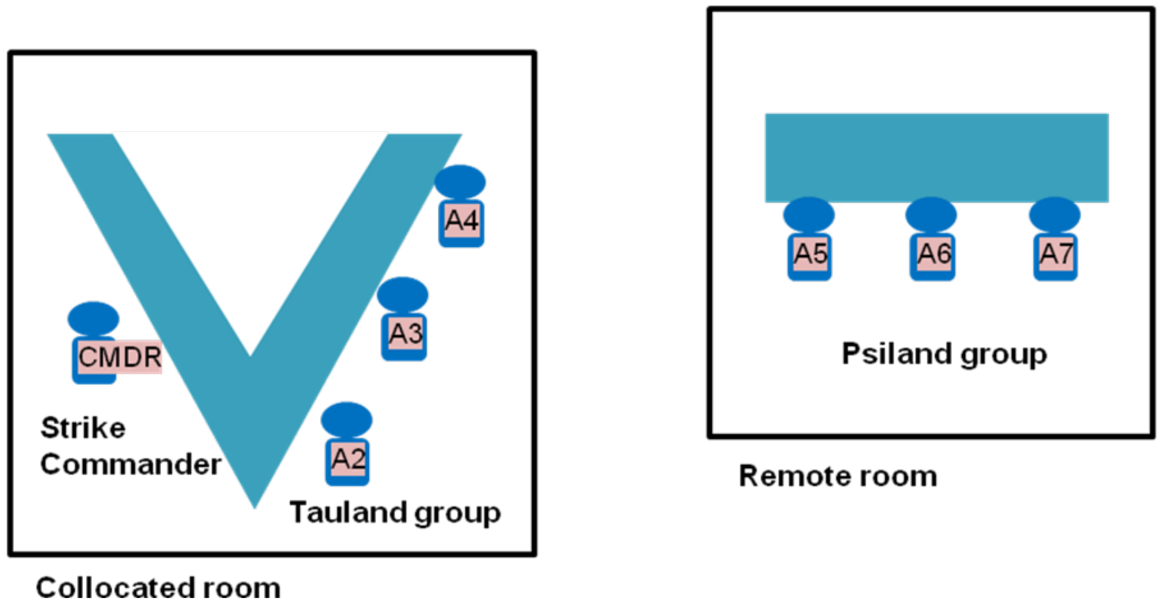

Figure 1. Geographic configuration of teams and commander. One three-person team was collocated with the commander, and was within speaking distance and eye contact range. The other team was located in a 'remote' location.

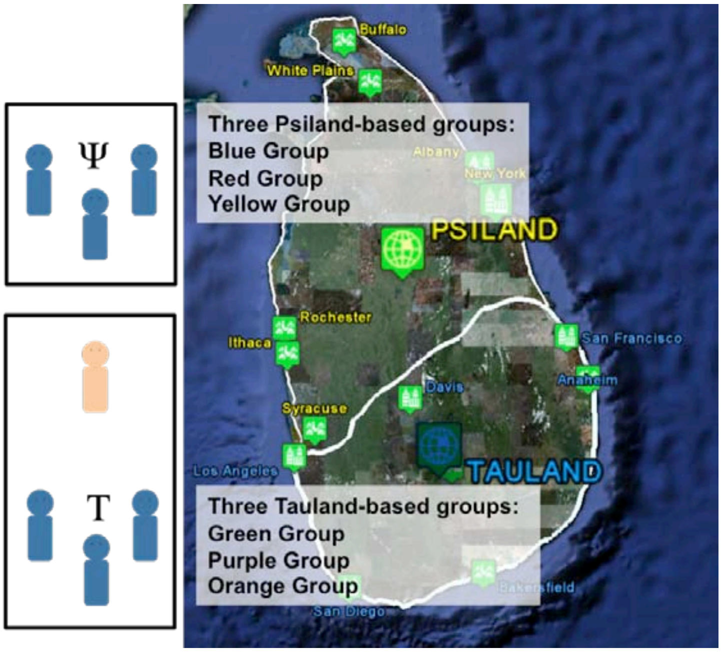

Figure 2. Groups of three analysts were assigned to 'Psiland' and 'Tauland'; each country had three terrorist groups to watch.

Comparison to baseline ELICIT setup

These components of ELICIT Multistrike are the same as the baseline ELICIT task:

- Factoids are distributed among multiple team members (in this case, everyone but the commander); factoids must be shared 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 3 shows how the set of factoids related to one terrorist group meeting (Green) were distributed among the six analysts. The numbers in parentheses indicate which "day" the factoid was released.

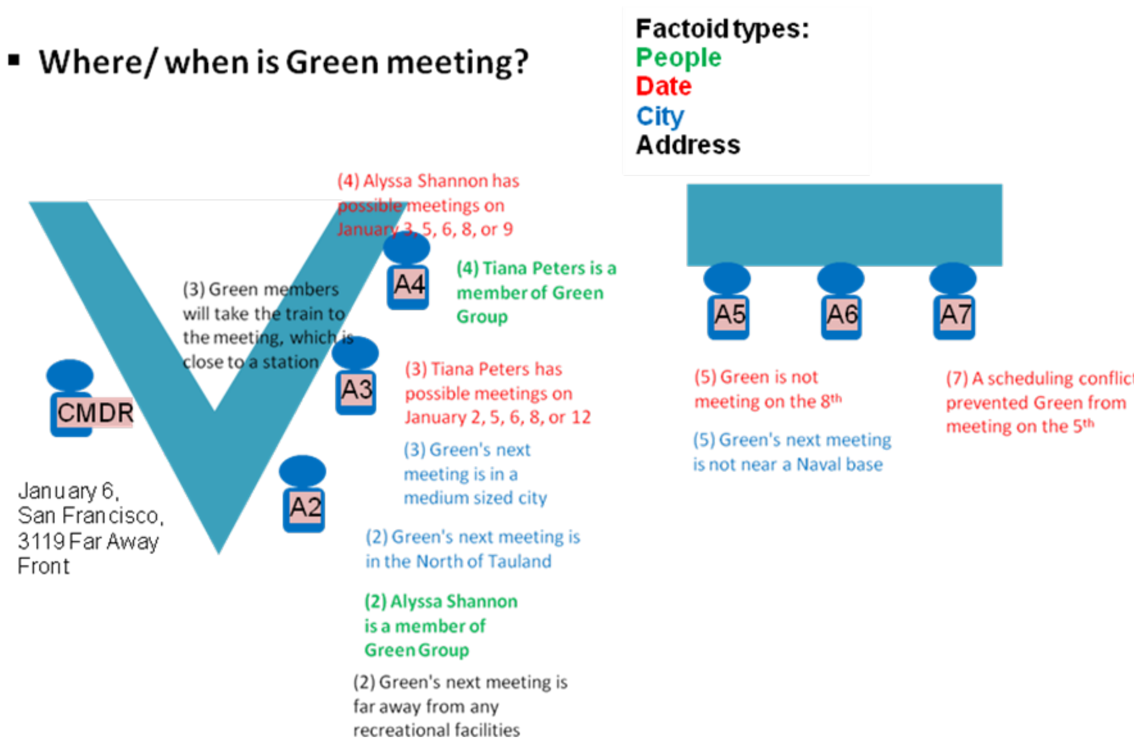

Figure 3. Distribution of factoids necessary to locate and strike Green group

- Factoids are shared via web pages. We used ELICIT's 'Web page' system to allow analysts to post factoids. In our configuration, there were no restrictions; every analyst could read and post to every web page. We did change the names of the pages and added a fourth page (which is supported within the system through configuration files); the four 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 web sites, and when these web sites were accessed by each analyst.

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

- Analysts were trained to use the 'web page' for posting factoids, but not 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; although we did develop plans to use it in future versions to mimic allocation of scarce intelligence, surveillance, and reconnaissance (ISR) resources.

The Multistrike task included these addition 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') with no other members. We anticipated this would be used for within-team analysis. Analysts were also free to communicate verbally with collocated team members, and the commander and Tauland team 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, requiring a fourth 'web page' board called 'people'.
- **Rolling deadlines.** The game was quite fast-paced, with each simulated day lasting only five minutes, as indicated by a countdown timer displayed in the front of the room. The task lasted one hour, or 12 simulated days. Figure 4 shows a timeline of when the terrorist meetings took place over those 12 days. Before the experimental task, teams played a three "day" practice session. Factoids were according to a staggered schedule; all information to strike a meeting was available to the team at least one day before each meeting.

Figure 4. Meeting timeline. A successful strike, with the right date, city, and address, would prevent future group actions.

- **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. This was necessary to allow teams some chance of addressing multiple problems in parallel; the combined tasks were still quite challenging. The tasks also required access to lookup tables for the city and address problems; a section of the address table is shown in Figure 5.

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 5. Lookup table of addresses.

- **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 casualities. 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.

Team Performance Measures Prepared for Pilot Study

Several team performance measures were prepared for the pilot study. Each measure has an expected finding, which corresponds to lower level hypotheses for the study. The subset comprises a complete set of measures to match the categories in the team performance measures listed earlier (see Table 1). Eventually, to meet the objective of determining the most diagnostic and feasible measures, more measures will be used and compared.

Task performance

Team score 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.

Information sharing performance

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

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 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 for the commander and the configuration of both the collocated team and the commander's workspace were designed so that the eyetracker could track when the commander was looking at each chatroom on her/his computer screen, when s/he was looking across at any of the collocated analysts, and when s/he was viewing the large screen data displays.

Expected finding: We anticipate that in a time sensitive task under high workload, the commander will pay more attention to local teammates and less attention to distant teammates (over 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 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 shorter lag time between when information from the collocated team is received and acted upon as compared to the distant team.

Measures of workload and situation awareness

NASA TLX is a widely used survey measure of workload that is completed post-task by participants (Hart, 1988). It measures six types of workload: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort, and Frustration. Situation Awareness Rating Technique (SART), similar to NASA TLX, is a post-task self-report measure of situation awareness (Taylor, 1989). While not as sensitive as assessments done mid-task, 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

Group identity. We used a ten-item scale measuring sense of group identity by Henry, Arrow, and Carini, 1999. 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 on group efficacy over distance adapted from Carroll , Rosson and Zhou (2005). Sample items include:

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

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

Reciprocity scale. We used an 11-item scale related to personal norms or reciprocity, based on Perguni, Gallucci, Presaghi and Ercolani (2002). 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:

1. *I went out of my way to help players who had helped me*
2. *If someone refused to help me, I held a grudge against them*
3. *I was kind and nice if others behave well with me, otherwise it was tit-for-tat*

Expected findings: We anticipated 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 build stronger rapport.

Social network trust and collaboration. Each player rated each other player in the session on five questions, each question 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. Sample items include:

1. *I worked closely with this player*
2. *I trust this player*
3. *I would like to play with this player again*
4. *This player was one of the leaders in the group*
5. *This player was helpful to others*

Expected findings: We would expect to see higher trust and familiarity ratings for those team members that were part of the same 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 volunteers, including training and the 12-day task. We calibrated but did not use eyetracker data; we also only had a few members of the first group of volunteers complete the battery of post-task measures.

In the second pilot run, the team called 3 successful strikes, 1 erroneous strike, and missed 3 meetings, resulting in a final score of zero. In a real run of this task, we would intend to run three or four sessions with the same teams, in distributed and collocated configurations, and expect learning effects to 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 Simile Timeline widget. The timeline in Figure 6 shows the release schedule over time for all factoids related to one terrorist meeting (Purple), as a complement to Figure 3, which focuses on the geographic distribution. The Simile Widgets, including the Timeline, are open-source spin-offs from the original Simile project developed at MIT. The Timeline widget uses Javascript and HTML to display event data from XML or JSON files. The widget is highly customizable and allows multi-band timelines, discrete and spanning events, visual style customization, and a responsive user-interface. For additional information on Simile and the Timeline Widget, visit http://simile-widgets.org/.

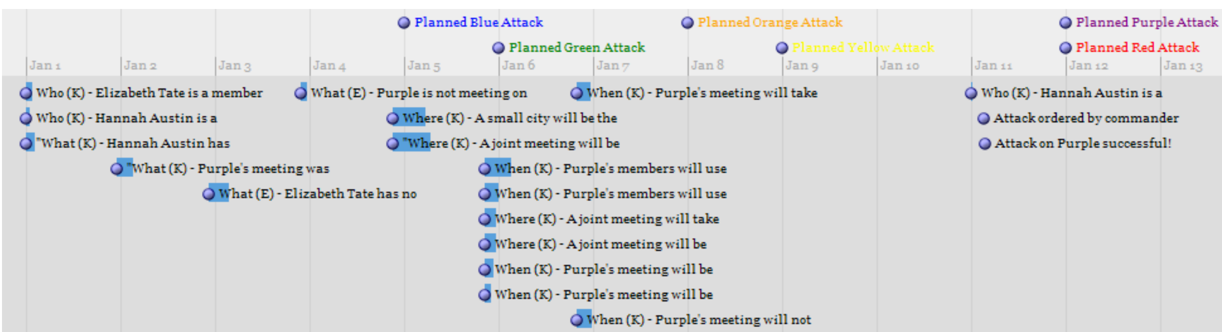

Figure 6. Schedule of factoid releases related to one terrorist meeting (Purple)

Future directions

We hope to secure funding in the near future to conduct a full study using the ELICIT Multistrike environment. 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. It is through the understanding of 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

Armstrong, D., and P. Cole. 2002. Managing distances and differences in geographically distributed work groups. In P. Hinds and S. Kiesler, eds. *Distributed Work*. Cambridge, MA: MIT Press. 167-189.

Bos, N.D., Olson, J.S., Gergle, D., Olson, G.M., & Wright, Z. (2002). Effects of four computer-mediated channels on trust development. In Proceedings of *CHI 2002*. New York: ACM Press.

Bos, N.D., Olson, J.S., Nan, N., Shami, N.S., Hoch, S., Johnston, E. (2006). Collocation blindness in partially distributed groups: is there a downside to being collocated? In Proceedings of *CHI 2006.* New York: ACM Press.

Bos, N.D, Shami, N.S., Olson, J.S., Cheshin, A. & Nan, N. (2004) Ingroup/ out-group effects in distributed teams: an experimental simulation. In Proceedings of *CSCW 2004*. New York: ACM Press. 429-436.

Brehmer, B. (2010, June). Command without commanders. Presented at the 15th *International Command and Control Research and Technology Symposium*, Santa Monica, CA.

Brehmer, B. (1998). Effects of time pressure in fully connected and hierarchical architectures of distributed decision-making. In Y. Waern (Ed.), *Co-operative Process Management*, 131 – 143. Bristol, PA: Taylor & Francis, Inc.

Carroll, J., Rosson, M.B. & Zhou, J. (2005). Collective Efficacy as a Measure of Community. In Proceedings of *CHI 2005*: New York, ACM Press.

Clark, H. H. & Wilks-Gibbs, D. (1986). Referring as a collaborative process. *Cognition*, 22, l-39.

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

Cramton, C. (2001). The mutual knowledge problem and its consequences for dispersed collaboration. *Organization Science, 12,* 346 – 371.

Endsley, M.R. (1988). Situation awareness globabl assessment technique (SAGAT). In Proceedings of *Aerospace and Electronics Conference*, 1988, 789-795. Dayton, OH.

Gutwin, C., & S. Greenberg. (2004). The importance of Awareness for team cognition in distributed collaboration. In E. Salas and S. M. Fiore, eds. *Team Cognition: Understanding the Factors that Drive Process and Performance,* 177-201, Washington: APA Press. Downloaded May 4, 2005 from http://grouplab.cpsc.ucalgary.ca/papers/

Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press. On-line downloaded 1-25-11 http://humansystems.arc.nasa.gov/groups/TLX/tlxpublications.html

Henry, K.B., Arrow, H. and Carini, B. (1999) A Tripartite model of group identity. Small Group Research, (30) 5, 558-581.

Hollingshead, A. B. (1998.) Retrieval processes in transactive memory systems. Journal of Personality and Social Psychology, 74:3, 659-671.

Jarvenpaa, S., & D. Leidner. (1999). Communication and trust in global virtual teams. *Organization Science,* 10, 791-815.

Kanawattanachai, P. and Y. Yoo (2002) Dynamic nature of trust in virtual teams. *The Journal of Stategic Information Systems*, 11(3-4), 187-213.

Kraut, R. E., C. Egido, J. Galegher. (1990). Patterns of contact and communication in scientific research collaborations. In Kraut, R. E., Egido, C., Galegher, J. (Ed.) *Intellectual Teamwork: Social and Technological Foundations of Cooperative*. Hillsdale, NJ: Lawrence Erlbaum. 149-172

MacMillan, J., Paley, M.J., Entin, E.B., and Entin, E.E. (2004) Questionnaires for Distributed Assessment of Team Mutual Awareness. *Handbook of Human Factors*.

Martins, L.L., Gilson, L.L. and Maynard, M.T. (2004) Virtual teams: What do we know and where do we go from here? *Journal of Management*, 30(6), 805-835

McEver, J.G., R.E. Hayes, and D.M. Martin, 2007. *Exemplar Analyses of ELICIT Experimental Data*, Presentation by Evidence Based Research. On-line http://www.dodccrp.org/files/elicit_analysis_final.pdf downloaded 1-25-11.

Nissen, M. E., & Leweling, T. A. (2007, June). Hypothesis Testing of Edge Organizations: Laboratory Experimentation using the ELICIT Multiplayer Intelligence Game. Proceedings International Command & Control Research & Technology Symposium, Newport, RI.

Orasanu, J. and T. Connolly, 1993. The reinvention of decision making, in Klein, G.A., J. Orasanu, R. Calderwood, and C.E. Zsambok (Eds.), *Decision Making in Action: Models and Methods*. Ablex Publishing: Norwood, NJ. 3-20.

Perugini, N.; Gallucci, M.; Presaghi, F.; & Ercolani, A.P. (2002). The Personal norm of reciprocity. European Journal of Personality, 17 (4).

Powley, N., & Nissen, M. E. (June 2009). Trust-Mistrust as a Design Contingency: Laboratory Experimentation in a Counterterrorism Context. Proceedings International Command & Control Research & Technology Symposium, Washington, DC.

Ratcheva, V.; Vyakarnam, S. 2001.Exploring team formation processes in virtual partnerships. Integrated Manufacturing Systems. 12/7. MCB University Press. p. 512-523.

Ruddy, M. 2007. ELICIT – The Experimental Laboratory for Investigating Collaboration, Information-sharing and Trust, *Proccedings of International Command and Control Research and Technology Symposium*.

Rocco, E., Finholt, T.A., Hofer, E.C., & Herbsleb, J.D. (2001, April). Out of sight, short of trust. Presentation at the Founding Conference of the European Academy of Management. Barcelona, Spain.

Ross, L.D. (1977). The Intuitive psychologist and his shortcomings: distortions in the attribution process. In L. Berkowitz (Ed.), Advances in experimental social psychology, 10.

Taylor, R.M. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. Proceedings of the *AGARD AMP Symposium on Situational Awareness in Aerospace Operations*, CP478. Seuilly-sur Seine: NATO AGARD.

Teasley, S. D., L. A. Covi, M. S. Krishnan, & J. S. Olson. (2002). Rapid software development through team collocation. *IEEE Transactions on Software Engineering*, 28(7), 671-683.

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

Woolley, A.W., Chabris, C.F., Pentland, A., Hashmi, N. & Malone, T.W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science* (330), 686-688.

Zelick, D., Patterson, E.S., and Woods, D.D. (2010). Measuring Attributes of Rigor in Information Analysis. In E.S. Patterson and J Miller (Eds.), *Macrocognitioin metrics and scenarios: Design and evaluation for real-world teams*. Aldershot, UK: Ashgate.