2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), Miami Beach, FL

Assessing Team Workload and Situational Awareness in an Intelligence, Surveillance, and Reconnaissance (ISR) Simulation Exercise

James C. Won, *MIT Lincoln Laboratory / Tufts University*
Gary R. Condon, Benjamin R. Landon, Andrew R. Wang, *MIT Lincoln Laboratory*
Daniel J. Hannon, *Tufts University*

***Abstract*—Currently, one of the greatest challenges to the nation's intelligence community is the discovery of clandestine networks, typically in the context of counter-insurgency operations. MIT Lincoln Laboratory has developed an Intelligence, Surveillance, and Reconnaissance (ISR) "Red/Blue" exercise in which two teams compete to discover a complex network within a simulated urban environment. The teams use wide-area persistent surveillance data, as well as decision support and situational awareness tools, to trace relationships between individuals, events, and sites. Information management, analysis, and tool interaction are some of the cognitive factors that contribute to varying levels of the team's workload and situational awareness. This paper describes an initial study of the relationship between workload and situational awareness (SA) in such a setting. The paper also presents observations regarding different roles within the teams and their corresponding workload and SA.**

***Index Terms*—ISR Simulation, Situational Awareness, Team, Workload**

I. Introduction

IN the intelligence community, identifying clandestine networks in cultural "clutter" is one of the most challenging tasks for counter-terror / counter-insurgency operations. Persistent surveillance can provide the raw data needed to support network discovery, but automated algorithms are not able to fully automate the exploitation process. Thus, we recognize that intelligence activities are inherently human-centric, as humans are the ultimate decision makers.

One important component to this decision making process is attaining optimal situational awareness (SA) [1] [2]. There are a variety of factors that can influence SA, but one parameter that has a significant effect is workload. Therefore, the goal then becomes to optimize both SA and workload, and understand the relationship between the two [3]. It is evident that in a high workload situation, a person's SA would be degraded [4]. Research has also shown that decreased SA can result from both high and low workload situations [5], similar to a workload-performance curve inspired by the Yerkes-Dodson inverted-U relationship (see Fig. 1) [6].

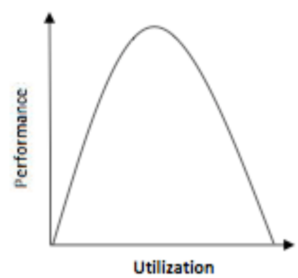

Fig. 1. Notional Workload-Performance Curve [6]

Understanding this relationship has particular relevance to the intelligence community, where teams operate together to make decisions to uncover clandestine networks. Obtaining optimal team workload and SA by establishing appropriate roles and number of analysts could provide valuable guidance in organizing these teams.

One platform to study this relationship is MIT Lincoln Laboratory's Intelligence, Surveillance, and Reconnaissance (ISR) "Red/Blue" exercise. It is designed as a process for discovering new intelligence exploitation concepts, system designs, and algorithms for a broad range of ISR missions through competitive game play. Counter-insurgency operations requires analysts to gain an understanding of a very complex scene; insurgents often conduct activities in the open, relying on the vast amount of urban "clutter" to disappear into background of the general population. This activity requires that humans provide much of the processing capability and understanding to maintain SA and interpret the intent of the activity (e.g. surveillance of a possible target) as opposed to the activity itself (e.g. stopping at a market).

"Red/Blue" experimentation, in general, is a human-in-the-loop process intended to hone tools, workflows, and organizational concepts through a highly instrumented and competitive game, which is focused around a specific

Manuscript received January 15, 2011. This work was sponsored by the Department of the Air Force under Air Force Contract #FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the Department of Defense of the United States Air Force.

J. Won is with MIT Lincoln Laboratory, Lexington, MA 02420 USA and Tufts University, Medford, MA 02144. (phone: 781-981-4264; e-mail: jwon@ll.mit.edu).
G. Condon, B. Landon, and A. Wang are with MIT Lincoln Laboratory, Lexington, MA 02420 USA
D. Hannon is with Tufts University, Medford, MA 02144 USA.

challenge area. Within a problem area, a relevant “Red” scenario is developed, along with a concept of operations (CONOPS) for the resources that the “Blue” teams will have available to them (e.g. specific sensor types). As teams work through the scenario, extensive measurements are taken on the use of the exploitation of tools provided to them, as well as observational data about behavior and organization. This information is then analyzed and fed into the tool development process. This process allows involvement of the user in the development of tools to improve performance, as well as observe the impact of workload and SA.

Many studies have been performed to observe this impact of workload and SA, with respect to adaptive automation [7], types of team displays, [8], and staff level team performance [9].

However, the majority of these and other efforts have been specifically designed to test these particular concepts within a set of design parameters in a laboratory setting. And although the “Red/Blue” exercise is a simulation and not truly a “real-life” scenario, it does emulate a command and control setting in which a team collaborates to solve a realistic problem, and thus provides tangible feedback regarding workload and SA that can be incorporated into real world analyst team formation and workflow.

II. METHODOLOGY

A. Participants

Twelve participants from MIT Lincoln Laboratory, MITRE, and Army Research Laboratory (ARL) took part in this exercise. They were divided up into two teams, six people per team. These are the two “Blue” teams that compete against one another in the context of one common “Red” scenario that has been developed for the game. The participants (mean age = 35.5 years; range 20-50 years) comprised 8 men and 4 women.

B. Design

After an initial overview of the game, teams were allowed to self-organize around the given tools and personnel. Each team was provided a limited set of tools, including four laptops, one projector, one whiteboard, one large geographical map of the city, and office supplies. Each team was given the ability to manually switch the laptop that is projected onto the large screen.

Resources (e.g., the number of laptops) are limited for this exercise. This forces the teams to make organizational decisions: the use of the workstations, how to divide them among the members, how to allocate jobs/tasks between them.

C. *“Red/Blue” Exercise*

The primary tool used to interact with the data is a tool called “BlueStreak” (see Fig. 2), developed at MIT Lincoln Laboratory. This tool is a geospatial and temporal representation of all of the information available in the game. It provides a timeline and a spatial map view of the scene with an overlay of the sensor data. The sensor data shows the movement of vehicles, which can be combined with information about various events occurring in the scene to identify potential hostile sites. Users have the ability to view the data forensically or in real time.

Fig. 2. “BlueStreak” Tool. It provides geospatial and temporal representation the information available in the game, including a timeline, map view, and overlay of sensor data.

An alternate “graph view” is also provided (see Fig. 3). This allows the user to establish nodes and links for the sites of interest in a graph format, ultimately to assist the team in establishing SA of the scenario. The tool is collaborative in nature in that annotations on the map and manipulations on the graph are shared between team members.

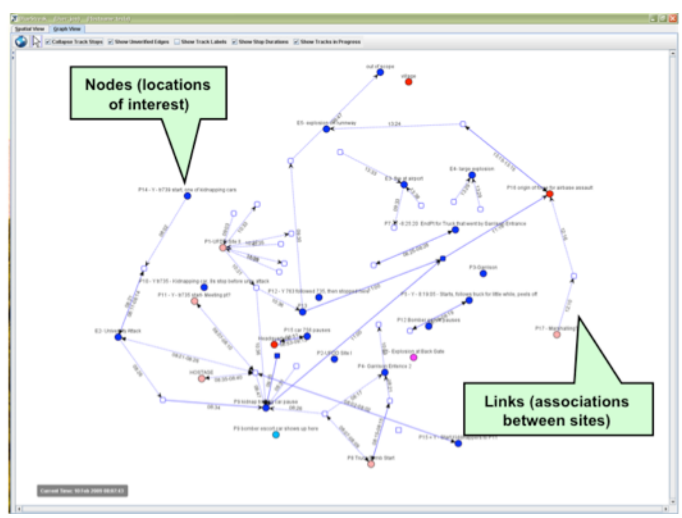

Fig. 3. Graph View. It allows users to establish nodes as locations of interest and links between the nodes to establish associations between sites.

D. *Procedure*

The participants spent one day in training, learning about and interacting with the “BlueStreak” tool, as well as spending time getting acquainted with their teammates. The teams were not given any direction regarding formation of the team, and much of the time was spent organizing the team, establishing and defining roles, and allocating resources.

The teams are self-organized, but typical roles that have emerged are that of leader, vice leader/scribe, graph analyst, and track analyst. There have typically been three track analysts. The leader is responsible for organizing and tasking team members, and maintaining awareness and understanding of events and the evolving situation. The vice leader/scribe's task is in assisting the leader and recording relevant information. The graph analyst's task is in organizing the graph, understanding relationships, and working with the track analysts and leaders to develop linkages between locations. The track analyst's task is in re-constructing the behavior of individual vehicles.

The following day was game day, in which the teams are given their mission. Throughout the three hours of game play, teams were notified of event occurrences from within the "BlueStreak" tool. Events became more frequent as the game progressed, with a variety of expected and unexpected circumstances arising.

At the end of the game, teams were scored on their ability to identify sites as red locations. The game is made intentionally difficult and un-winnable if played alone. This forced team members to collaborate with each other to jointly approach the problem. The difficulty also required team organization to maintain cognition of both what is happening in the game (situational awareness) as well as how the various team members are spending their time (resource allocation).

The two teams participated in the exercise in separate rooms for three hours, and a questionnaire was distributed to all the participants immediately following the exercise. Background questions, as well as subjective questions to assess levels of workload and situational awareness, were posed to the participants. NASA Task Load Index (TLX) and the Situational Awareness Rating Technique (SART) were the methodologies used to assess workload and situational awareness.

E. Workload Evaluation

Evaluating workload is an important aspect of a system evaluation. Subjective workload measurement tools have been developed to quantify the effort a user exerts during the performance of a task. NASA-TLX was the tool that was used for this study. It is a rating procedure that derives an overall workload score based on weighted average ratings on six subscales: Mental Demands, Physical Demands, Temporal Demands, Own Performance, Effort, and Frustration [10].

F. Situational Awareness Evaluation

SART is a measure that indicates the participant's subjective rating of his or her degree of situational awareness. This measure involves three main concepts, Demand and Supply of Attentional Resources, and Understanding. Each concept contains its own sub-concepts. The sub-concepts for Demand are Instability of Situation, Variability of Situation, and Complexity of Situation. The sub-concepts for Supply are Arousal, Spare Mental Capacity, Concentration, and Division of Attention. The sub-concepts for Understanding are Information Quantity, Information Quality, and Familiarity [11] [12] [13].

For the subjective situational awareness ratings, each participant rated each sub-concept on a scale of 7 point rating scale (1=Low, 7=High), and the scores are then combined to form an overall SA score with the formula:

$$SA = \text{Understanding} - (\text{Demand} - \text{Supply})$$

III. RESULTS

The participants' subjective workload ratings from the NASA-TLX questionnaire were collated and averaged across the four roles on the team. The results were then plotted by roles across the six components of the workload rating (see Fig. 5). The plot indicates that the highest rating is attributed to the mental demand of the Vice Leader/Scribe.

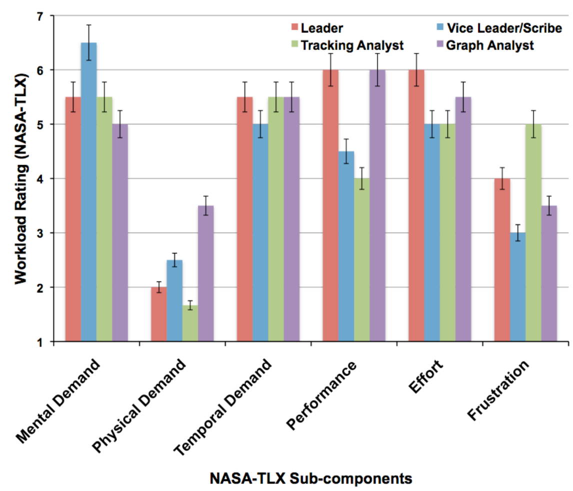

Fig. 5. Workload components by team roles

Likewise, the participants' subjective situational awareness ratings from the SART questionnaire were collated and averaged across team roles (see Fig. 6).

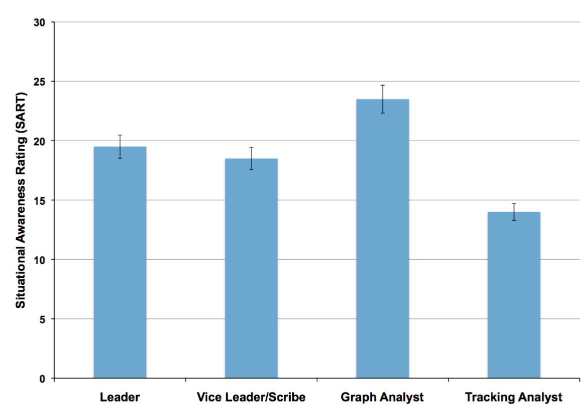

Fig. 6. Situational awareness across team roles

The results indicate that the graph analysts had the highest level of situational awareness, even higher than the team leader.

The results can be broken down further by plotting the roles across the 3 main components of the SART score (see Fig. 7). This breakdown shows that although there was a relatively even distribution of demand, the graph analyst exhibited the highest level of understanding while also having a high level of supply.

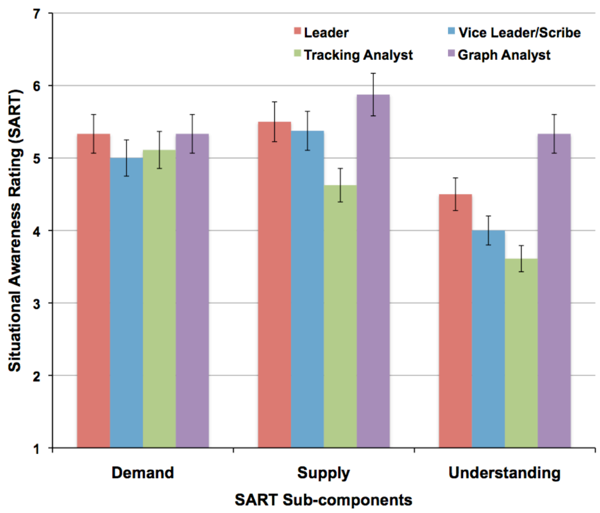

Fig. 7. Situational awareness components by team roles

Lastly, the overall NASA-TLX and SART scores were plotted together to visualize their relationship to each other (see Fig. 8). This allows us to hypothesize how the workload vs. SA curve, or curves, may look like for this "Red/Blue" exercise. A distinction seems to occur between information synthesizing roles (leader, vice leader, graph analyst) and information producing roles (track analyst). As the data seems to cluster by these roles, there could potentially be different curves for each particular role.

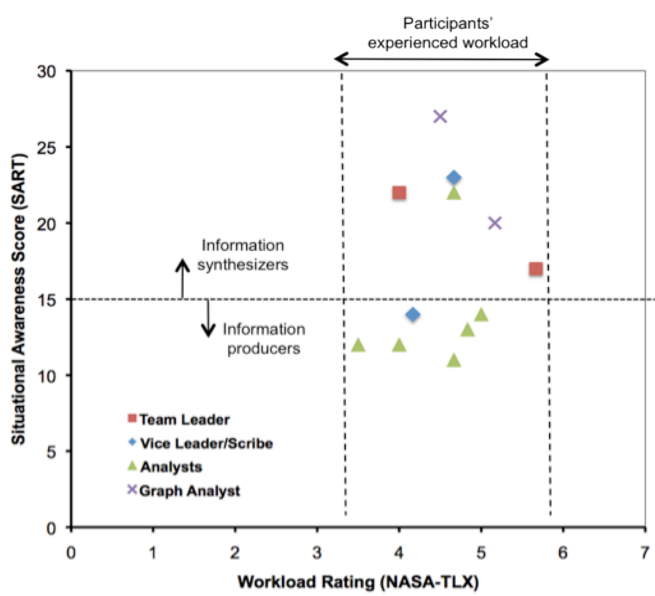

Fig. 8. Workload vs. Situational Awareness. Participants all experienced moderately high workload, while SA differentiated according to role type.

IV. Discussion

Future work in this domain will use the same "Red/Blue" exercise to build upon these initial results. The goal is to examine the constructs of workload and situational awareness, and their relationship with one another.

Regarding workload, the results indicate that all participants experienced similar temporal demand due to the time pressures imposed by this exercise. However, there are differences in the mental demand across the different roles, the highest rating belonging to the vice leader/scribe. Further testing could reveal what factor leads to this increased mental demand, whether it is due to the nature of performing a dual role, or whether it is simply due to the nature of the role of a scribe or a vice leader.

Regarding situational awareness, as mentioned above, the results show that the graph analysts perceived themselves having higher SA than the team leaders. Further testing could explore what effect this has on overall team performance, and whether one dedicated person to monitor SA is an effective method of role allocation in a team. This preliminary result goes against conventional thinking that a team leader should have the highest level of SA on the team. However, the role of a team leader involves many different facets, including coordination of roles, maintaining team morale, and others, which draw more of the team leader's resources, or "available workload". These facets are not part of the graph analysts' responsibilities, allowing them to devote their resources to simply establish good SA of the scenario.

Future work could also address the interaction and relationship between workload and SA. One could hypothesize that both high and low levels of workload lead to low SA, where the operator either has little spare capacity to develop a mental model or is too "out-of-the-loop" to build the necessary knowledge state [5]. Or one could hypothesize that the particular team role could be a significant factor, and further testing could show different shaped curves for different roles.

Further work could build upon the many existing efforts to quantify this type of relationship, in the spirit of the Yerkes-Dodson principle [6]. We could examine how this principle, or perhaps an "extended-U" concept [14] applies to this relationship. Or we could perhaps see that the principle does not apply and that operator SA does not drop with low workload, as was found with regard to operator performance [15]. Ultimately, insights would show that perhaps the relationship is not as simple as a correlation between overall numbers, but a more intricate interplay between the deeper components of these constructs. Different parameters within the "Red/Blue" exercise could be manipulated to examine this interplay. For example, while these initial results show that this experiment did not exercise the full spectrum of workload, future work could use the time factor to increase or decrease temporal demand in a systematic way by varying the frequency of events and time scale at which the game is played. These variations would be intended to reveal the deeper interplay between the different components of workload and situational awareness.

Lastly, the initial results indicate that both NASA-TLX and SART are sufficient methods to subjectively measure workload and situational awareness. Objective measurement techniques were considered, but not implemented for this initial phase in order to maintain game flow and not disrupt the exercise. Future work would aim to gather more data to evaluate whether these measures have the sensitivity required to distinguish the deeper interplay factors, recognizing that the relationship between workload and SA is not entirely straightforward. Deeper investigations into the effect of roles is required, as well as an examination of the components of the measures themselves to study the possible correlations between SART and NASA-TLX.

V. Conclusion

The purpose of this study was to assess the relationship between workload and situational awareness in an ISR simulation exercise, a scenario which is not entirely a "real life" scenario, but at the same time is not a high scripted laboratory experiment designed to measure only workload and situational awareness.

Understanding this relationship and interplay between the constructs of workload and SA allows both developers and users in the ISR realm to identify how a team can attain efficient performance, maximizing situational awareness while minimizing workload.

References

[1] C., McCue, "Data Mining and Predictive Analytics: Battlespace Awareness for the War on Terrorism," *Defense Intelligence Journal,* vol. 13 (1&2), pp. 47-63, 2005.

[2] J. A. Adams, "Unmanned vehicle situation awareness: A path forward," in *Human Systems Integration Symposium*, 2007.

[3] M. R. Endsley, "Situation awareness and workload: Flip sides of the same coin," in *Proceedings of the Seventh International Symposium on Aviation Psychology*, R. S. Jensen and D. Neumeister, Ed. Columbus, OH: Department of Aviation, The Ohio State University, pp. 906-911, 1993.

[4] P. Scerri, S. Owens, L. Sycara, and M. Lewis, "User evaluation of a GUI for controlling an autonomous persistent surveillance team," *Proceedings of SPIE*, vol. 7694, May 2010.

[5] K. C. Hendy, "Situation Awareness and Workload: Birds of a Feather?" *AGARD AMP Symposium on "Situational Awareness: Limitations and Enhancements in the Aviation Environment,"* pp. 21-1 – 21-7, 1996.

[6] M. L. Cummings and C. E. Nehme, "Modeling the impact of workload in network centric supervisory control settings," in S. Kornguth, R. Steinberg, & M. D. Matthews (Eds.), *Neurocognitive and physiological factors during high-tempo operations*, Surrey, UK: Ashgate, 2010, pp. 23-40.

[7] D. B. Kaber, J. M. Riley, K. Tan, and M. R. Endsley (2001). "On the design of adaptive automation for complex systems," *International Journal of Cognitive Ergonomics*, vol. 5 (1), 2001, pp. 37–57.

[8] C.A. Bolstad and M.R. Endsley, "The effect of task load and shared displays on team situation awareness," *14th Triennal Congress of International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* San Diego (July–August 2000).

[9] J. A. Childs, K. G. Ross, and W. A. Ross, "Identifying Breakdowns in Team Situational Awareness in a Complex Synthetic Environment to Support Multi-focused Research and Development,"

[10] S. Hart, and L. Staveland, "Development of NASA TLX (Task Load Index): Results of empirical and theoretical research," in P. Hancock & N. Meshkati (Eds.), *Human mental workload,* Amsterdam: Elsevier, pp. 139-183.

[11] S. J. Selcon, R. M. Taylor, and E. Koritsas, "Workload or situational awareness?: TLX vs. SART for aerospace systems design evaluation", i*n Proceeedings of the Human Factors Society's 35th Annual Meeting*, Santa Monica, CA, USA: Human Factors Society, 1991, pp. 62-66.

[12] R. M. Taylor, "Situational awareness rating technique (SART): The development of a tool for aircrew systems design," in *Situational Awareness in Aerospace Operations (AGARD-CP-478)*, Neuilly Sur Seine, France: NATO-AGARD, 1990, pp. 3/1-3/17.

[13] M. R. Endsley, S. J. Selcon, T. D. Hardiman, and D. G. Croft, "A comparative analysis of SAGAT and SART for evaluations of situation awareness," in *42nd Annual Meeting of the Human Factors & Ergonomics Society*, Chicago, IL.

[14] P. A. Hancock and H. C. N. Ganey, "From the inverted-U to the extended-U:The evolution of a law of psychology. ," *Human Performance in Extreme Environments,* vol. 7, pp. 5-14, 2003.

[15] C. S. Hart, "Assessing the Impact of Low Workload in Supervisory Control of Network Unmanned Vehicles," Thesis, Massachusetts Institute of Technology, 2010.