Assessment for Operator Confidence in Automated Space Situational Awareness and Satellite Control Systems

Joe Gorman

Charles River Analytics

Martin Voshell

Charles River Analytics

Amy Sliva

Charles River Analytics

1. INTRODUCTION

The United States is highly dependent on space resources to support military, government, commercial, and research activities. While satellites operate at great distances, on-line observation capacity is limited, and operator actions and observations can be significantly delayed. Safe and timely operation requires decision support systems that are designed to provide situational understanding, enhance decision making, and facilitate collaboration between human operators and system automation both in-the-loop, and on-the-loop.

As decision supporting systems (DSS) grow in complexity, their autonomous capabilities and the automated processing of their inputs multiply the number of sources used, the volume of data analyzed, and the rate at which outputs can be updated. Improved DSS autonomy has the potential to improve situational estimates and response estimates. However, the heuristic and probabilistic nature of analysis performed by these systems is difficult to verify and validate (V&V). The complexity of the systems and the lack of formal V&V reduces the confidence of human operators and decision makes teaming with them. Our previous research proposed a novel automated process assessment method that would allow developers and system evaluators to use subjective evaluation criteria defined by users (operators and decision makers) that relay on the DSS to increase their confidence [1]. This prior work raised a critical systems engineering need, in that such approaches did not readily address *how* to identify realistic test conditions for use during V&V that would promote requisite operator confidence and decision support. Joint cognitive systems engineering (JCSE) provides a rich set of methods for analyzing and informing the design of complex systems that include both human decision-makers and autonomous elements as coordinating teammates [2]. While, JCSE-based systems can enhance a system analysts' understanding of both existing and new system processes, JCSE activities typically occur outside of traditional systems engineering (SE) methods, providing sparse guidance about how systems should be implemented. In contrast, the Joint Director's Laboratory (JDL) information fusion model [3] and extensions, such as the Dual Node Network (DNN) technical architecture [4], provide the means to divide and conquer such engineering and implementation complexity, but are loosely coupled to specialized organizational contexts and needs. The novel Dual Node Decision Wheel (DNDW) concept extends the DNN to integrate JCSE analysis and design with the practicalities of system engineering and implementation using the DNN [5] [6]. This approach combines insights from Rasmussen's JCSE Decision Ladders align system implementation with organizational structures and processes [7] [2] [5].

In the current work, we present a novel approach to assessing joint human automation system performance based on patterns occurring in operational decision-making that are documented by JCSE processes as traces in a decision ladder (DL). In this way, joint human-automation system assessment is closely tied not just to system design, but the design of the joint cognitive system that includes human operators, decision-makers, information systems, and automated processes that must work together to accomplish operational goals. Such operationally relevant and integrated testing provides a sound foundation for operator trust in system automation that is required to safely operate satellite systems.

The following section describes the evolving DNDW concept. When next describe the concept of of traces in decision ladders, followed by an illustration how traces are integrated for a robust human-system assessment. Finally, we present conclusions and recommendations for future work.

2. DUAL NODE DECISION WHEELS

The DNDW architecture [6] was originally developed to bridge the gap between information fusion approaches for sensemanking and decision support methods for response planning and plan execution. The DNDW combines the structural elements of the dual node network (DNN) technical architecture for data fusion and resources management [4] with the cognitive engineering elements of decision ladders [2] [8] and decision wheels [9], helping alleviate challenges in decision making processes of operators and decision makers and the new "big data" challenges [9] through more fluid coordination of support technologies within organizational and cognitive decision-making processes.

Individually, the DNN technical architecture and cognitive decision making models each provide a solution to only part of the problem facing users of DSSes. The DNN is an enhancement of the revised JDL data fusion model [3] [4], which provides a general framework for describing and designing information fusion and response planning applications. The DNN extends this framework to explicitly capture the relationship between data and information fusion products and response management to communicate dynamic information needs to the underlying fusion process as illustrated in Figure 2. Using this duality between data fusion and resource management processes, the DNN provides a technical architecture that supports effective assessment and management of the expanded portfolio of data sources, resources, entities of interest models, and algorithms. The DNDW leverages this flexible, interconnected structure to organize technical fusion and analysis components; however, this structure alone still regards decision making as exogenous to the fusion and resource-management process. DNDW takes this framework a step further by using concepts from JCSE [2] [10] to develop a formalism that can integrate fusion and decision making to provide a holistic view of an organizations decision making processes.

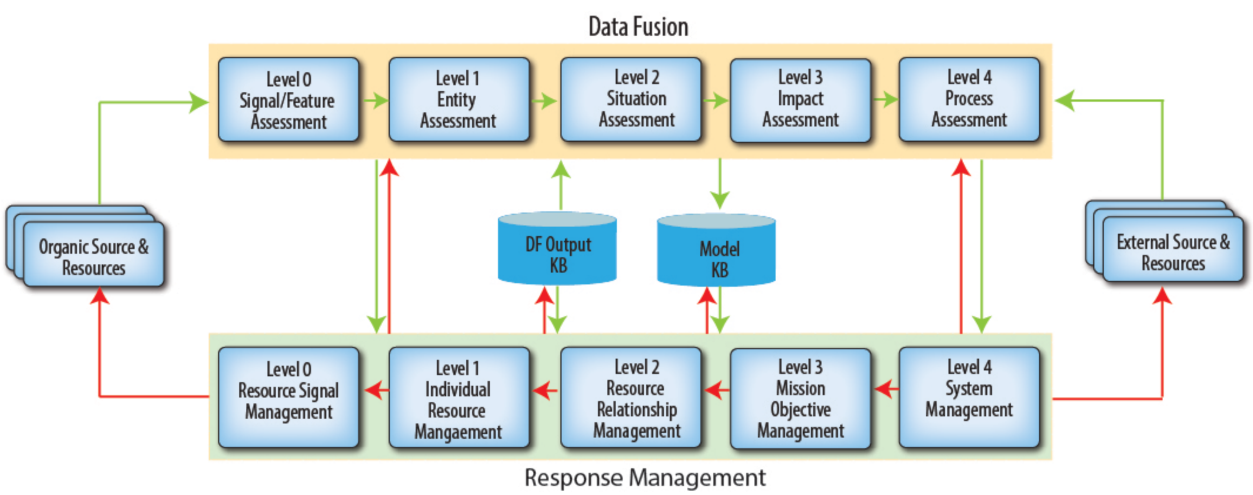
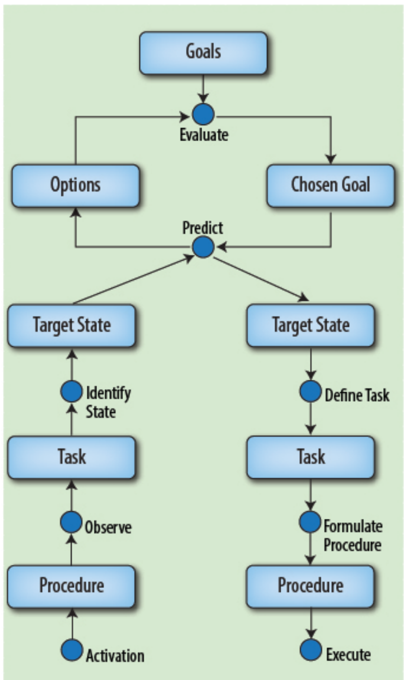
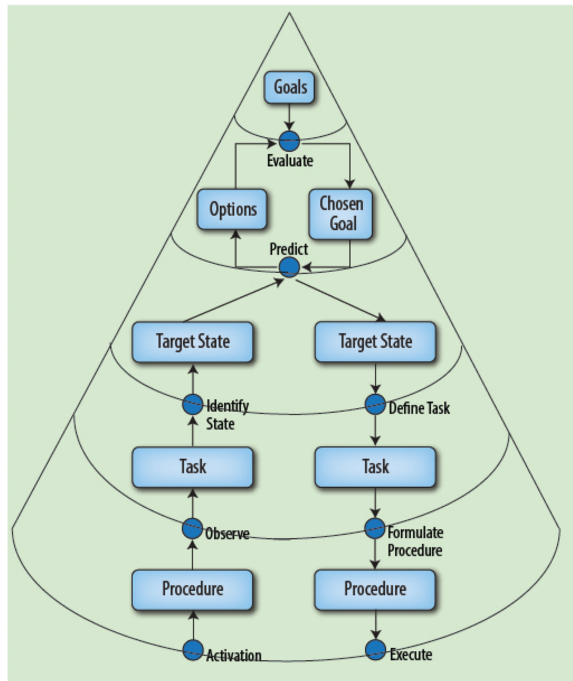


Figure 1: Conceptual information exploitation system implemented using the DNN framework.

JCSE is an approach to designing human and technical systems that leverages models of human cognitive and organizational processes, such as decision making, to identify the optimal system design in support of accomplishing individual and organizational work goals [2] [10]. Rasmussen's decision ladder, pictured in Figure 2(a), provides a template for describing decision-making processes and associated control and task requirements. However, in many contexts we are not designing information systems for individual processes, but for integrated processes across an organization. For example, in space operations intelligence analysis and command and control (C2), multiple coordinating agencies must effectively collaborate to support remote commanders as distributed staffs and agencies across multiple echelons that continuously make decisions based on available information. Conducting information analysis in these environments entails supporting the cognitive work by operators and automated analytics across the organization. The basic decision ladder can be extended for larger organizational and team processes through the JCSE decision wheel model [11], where an individual decision ladder maps to a single "slice" of a decision wheel representing a larger team (Figure 2 (b)). Within an enterprise or distributed organization, multiple such wheels interact similar to interlocking mechanical gears.



(a)



(b)

Figure 2: Concepts from cognitive systems engineering that represent human cognitive work and decision-making processes: (a) decision ladder for individuals, and (b) one slice of a decision wheel of interrelated decision ladders for organizational or team decision making. Decision ladders are typically used as an analysis tool to describe either existing or newly designed decision making processes. There is no distinction between automation and operator supported process.

The DNDW combines insights from structural data fusion in the DNN [4] and decision-making models from CSE [2] [10], representing a fundamentally new perspective on both information fusion and organizational decision making, viewing these not as distinct phases in an operational pipeline, but as highly interconnected processes. Figure 3 illustrates the components of the DNDW architecture, combining aspects of both the DNN and cognitive decision models to align and coordinate situational assessment, planning, and response execution processes. DNN provides a means of organizing and implementing data fusion, planning and control processes, while cognitive decision models describe the decisions being made and the processes leading up to them, indicating how the data fusion system must fit into the organization.

The bottom portion of the diagram is essentially a DNN, capturing the processing and exploitation of multi-INT data through the detection of features of interest, observations of entities (e.g., high value targets), and situational understanding of relationships between entities. Communication between these sensemaking fusion processes (i.e., fusion nodes) and processes supporting C2 tasks (i.e., response management nodes) ensure alignment and coordination. The DNN is interwoven with a decision ladder, indicating the processes and tasks—which may be either automated or human cognitive processes—that occur as data flows through and operations occur in the DSS.

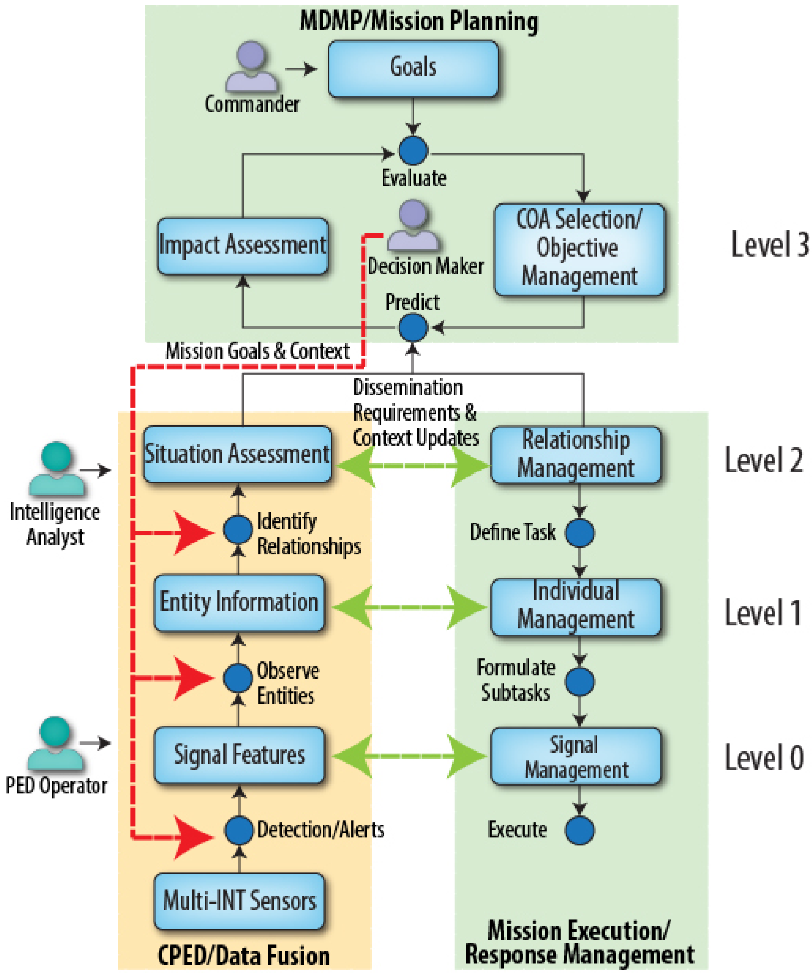


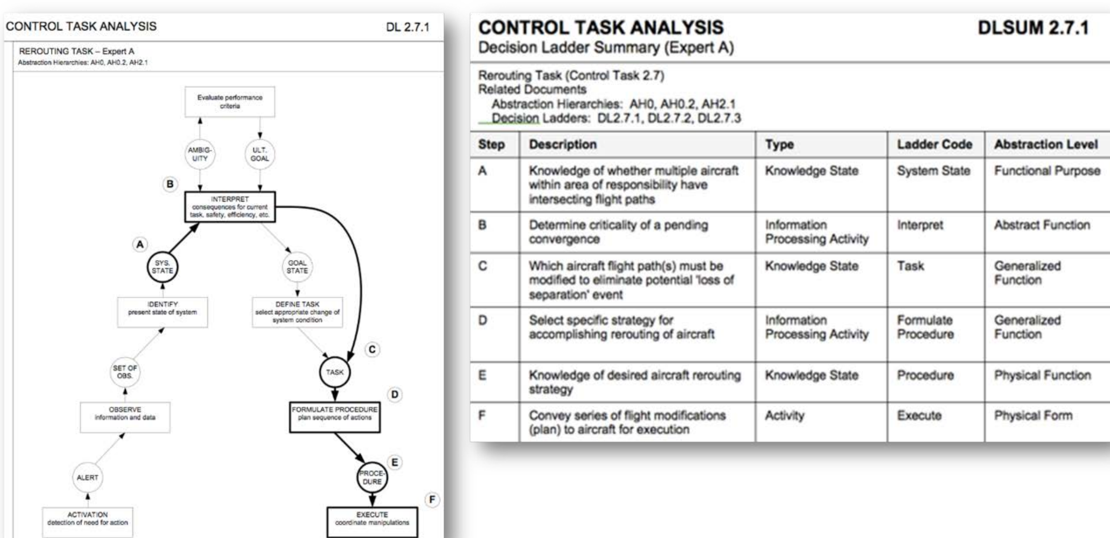
Figure 3. Dual Node Decision Wheels architecture.

The DNDW concept represents a new, integrated approach to information fusion and organizational decision making. There is alignment between goals and information that produces DSS designs that more accurately represents high-level fusion and the required feedback between C2 and fusion processes necessary to support effective data driven decisions. A DNDW system will be a collection of services (i.e., multi-INT sensors, processing and exploitation, decision support) linked according to communication standards to facilitate the alignment of intelligence products and C2 requirements. Using a service oriented architecture (SOA) in this is known to produced systems that are flexible and able to integrate analytic processes in a "plug and play" manner to incorporate the most effective tools or system chains. This implementation agility combines with an organizational view of DDS that makes DNDW suitable for integrating processes across any organizational structure and scalable to information flows across echelons and enterprise systems.

3. DECISION LADDERS AND ACTIVITY TRACES

While JCSE-based analysis and design approaches have been developed to address the challenges of system design, for the most part, Cognitive Task and Work analysis type evaluations are rarely strongly empirical. While developing new automated displays for first-of-a kind Navy battle management systems Rosseau, et al derived the concept of decision-centered testing (DCT) [12]. DCT was an attempt to show how a JCSE investigation can work as part of an empirical system validation and complement current guidance with regard to testing practices addressing software validation, human performance, and usability analysis. Using the analogy of how a mechanical engineer can employ sets of advanced analytics and mathematical modeling tools to identify where key 'stress points' are in a structure, the analogy for HSI evaluation, the evaluator should not have to touch everywhere in the system, but rather focus on key system stress points to focus on and elicit analytic information. Central to the DCT concept is the use of a system decomposition (e.g., functional analysis) combined with well-defined support requirements necessary for any goal-directed task. The identification of such support needs then can guide the design of particular test scenarios. The more robust and detailed the work analysis, the more accurately specific places can be targeted to “pressure” with testing. For example, in the authors’ recent anti-submarine warfare (ASW) research, the use of a preliminary functional analysis identified a key decision-difficult challenge consistently faced in ASW sonar operators which involved conducting middle frequency active (MFA) searches. Such search work relies on an operator’s ability to discriminate contacts between bottom-down and direct path frequency scans. However, this monitoring entails navigating separate modes across different displays to integrate this sonar information and then decide when and what to assign to an automated tracker. From a cognitive analysis standpoint, switching between modes can rapidly amplify data overload problems (e.g., monitoring increasing numbers of erroneous trackers and frequency bounces), which implies criteria for evaluation. An alternative display (e.g., integrating multiple sonars, or one with observable automated tracking support) can be compared against the existing display, (e.g., using a constant stimuli, forced choice, or other relevant method) to assess performance (e.g., temporal pressure as a function of workload) against this fundamental support challenge.

Such a DCT-like approach implies a unique capability for DSS evaluators—the ability to take insights from JCSE-based work analyses, and then using those insights to develop more formal usability evaluations and test cases. While there are many well-established and empirically validated existing experimental methodologies that can be used for investigation, what is novel about adapting this approach is that it gives a principled engineering approach to identify and define what questions the experiments and evaluations should answer based on the hypothesized cognitive and collaborative work support related to different operator/analyst behaviors. Within the DNDW, we can use DCT as one such methodology, not for developing experiments, but as a sound method for developing hypotheses for evaluation not just of the joint cognitive system but of system components and the integration between different functional components. The DNDW uses pragmatic JCSE approaches to identify the most critical empirical evaluations by taking predictions from cognitive analyses, and then turning those predictions into well-designed experiments or system tests. This is a novel approach for using JCSE to identify and target potential decision-making weak points in a given display or analytic process design with regard to larger joint system functionality, and then employing existing and established methodologies to discover key areas for testing and focused evaluation.



CONTROL TASK ANALYSIS **DLSUM 2.7.1**
Decision Ladder Summary (Expert A)

Rerouting Task (Control Task 2.7)
Related Documents
Abstraction Hierarchies: AH0, AH0.2, AH2.1
Decision Ladders: DL2.7.1, DL2.7.2, DL2.7.3

Step	Description	Type	Ladder Code	Abstraction Level
A	Knowledge of whether multiple aircraft within area of responsibility have intersecting flight paths	Knowledge State	System State	Functional Purpose
B	Determine criticality of a pending convergence	Information Processing Activity	Interpret	Abstract Function
C	Which aircraft flight path(s) must be modified to eliminate potential 'loss of separation' event	Knowledge State	Task	Generalized Function
D	Select specific strategy for accomplishing rerouting of aircraft	Information Processing Activity	Formulate Procedure	Generalized Function
E	Knowledge of desired aircraft rerouting strategy	Knowledge State	Procedure	Physical Function
F	Convey series of flight modifications (plan) to aircraft for execution	Activity	Execute	Physical Form

Figure 4: Decision Ladder Traces

Figure 4 shows two products of a representative task analysis. On the left is a decision ladder that describes the decision making environment within an organization. This decision ladder includes the trace of a particular decision making process that begins at step A and proceeds through steps B, C, D, and E before ending at step F. The table on the right of Figure 2 documents the operations at each step. Successful completion of this process requires success at each step. The process described a use case that must be verified in the DSS. Each step in the process represents a potential test point. A pragmatic approach to testing will focus on the weak points in the decision making process and allocate fewer resources to routine or highly reliable processes.

For example, every connection between an Air Force Satellite Control Network (AFSCN) ground station and a satellite requires many steps (e.g., the connection must be scheduled, the antenna must be aligned, C2 and data communications must be established). Scheduling a connection might be highly reliable and a reduced focus in testing. In contrast, establish connections might be considers a stress point and comprehensively assessed. The focus on system stress points not only limits the cost of assessment but also ensures higher reliability that promotes user confidence.

4. SYSTEM ASSESSMENT FOR OPERATOR CONFIDENCE

Technology acceptance is both a product of user perception of technology suitability and technology trustworthiness [13]. Klein, Moon, and Hoffman [14] explain that individuals use mental models to structure complex facts into meaningful perceptions. A mechanism is need to operationalize the end user's mental model so end user perception directs system development and assessment. A result of focusing on the user's subjective priorities and the points of most stress in a system will promote the perception of both suitability and trustworthiness.

Kaplan and Norton [15] introduced balanced scorecards as a way to decompose strategic goals into operational objects and measure of performance. Scorecards provide a two way model. Strategic planners develop the model top-down to ensure that operational objective and measures of performance align with the organization's mission and strategic goals. Operational managers use the scorecard from the bottom-up to ensure that organizational performance is achieving the objectives that support the organization's mission.

Operators and decision makers that rely on intelligent systems are in a similar position as the users of balanced scorecards. During the development of an intelligent system, users must communicate performance requirements in a top-down manner to developers. During system testing and evaluation, the developers must communicate performance assessment results in a bottom-up manner to end-users. Business process traces in a decision ladder enable scarce resources to be focused on experiments and test that have the highest value. Scorecards provide the means for including user perception of the system.

Dynamic performance assessment scorecards (Figure 5) use the paradigm of balanced scorecards to capture the end-user's mental model for system performance that will satisfy the intelligent systems intended mission. The first four columns describe the end-user's goals, objectives, and measures of performance (MOPs) for the system. Influence is the user's perception of the relative importance of a factor. The right-most column contains the results of the single test and an estimate of the end user's perception of the system's fit based on those results.

Goal:	Reliable detection and reporting of space weather events in a timely manner.		**System Fit:**	0.51
		Perception		
		>= .7 support, trust system		
		>= .5 support, Useful but don't trust		
		< .5 support, Not a useful result		
Objective	**Measure**	**Perception**	**Influence**	**Actual**
Accurate Results			***0.80***	
	Never miss an event		0.90	94.23
		>= 98% Achieves Objective	0.70	0.66
		<= 90% Fails Objective	0.40	0.33
	Few reports that are not events		0.70	53.8
		<= 30% False Alarms Achieves Objective	0.80	0.33
		>= 60% False Alarms Fails Objective	0.50	0.66
Timely Results			***0.40***	***42***
	Minimal delay between observation and report		0.70	11.6
		<= 18 Hour Delay Achieves Objective	0.80	1
		Otherwise Delay Fails Objective	0.20	

Figure 5: Dynamic performance assessment scorecard for space weather event detection system.

Aligning performance assessment reports with the end-user's mental model for performance promotes the perception of technology fit and system trustworthiness that are critical contributors to eventual technology adoption. Scorecards are useful for describing a user's subjective values, but developers benefit from the operationalization of those values during system evaluation. For example, the scorecard in Figure 5 indicates that improving detection rate has a greater influence on user perception of fit than decreasing false alarms. Dynamic performance assessment scorecards enable developers to better understand user requirements and allocate resources appropriately.

5. CONCLUSIONS AND FUTURE WORK

Decision centered testing (DCT) provides a pragmatic approach for focusing system evaluation of the most vulnerable points in a cognitive system. Dynamic decision scorecards integrate users' subjective criteria for suitability and trustworthiness. The key factors that promote adoption and system satisfaction. The DNDW provides a framework for integrating both (DCT and subjective criteria) to maximize user confidence in decision supporting systems. The next step in our work is to integrate DCT with dynamic performance assessment scorecards.

7. ACKNOWLEDGEMENT

This material is based upon work supported by the United States Air Force under Contract No. FA8750-14-C-0011. The author's gratefully acknowledge the support of the by the Air Force Research Laboratory (AFRL) and the Defense Advanced Research Project Agency (DARPA). Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Air Force or DARPA.

8. REFERENCES

[1] J. Gorman, "Performance Assessment, Trust, and Decision Support," in *Infotech@Aerospace 2011*, St. Louis, MO USA, 2011.

[2] J. Rasmussen, A. M. Pejtersen and L. P. Goodstein, Cognitive Systems Engineering, New York, NY USA: Wiley, 1994.

[3] D. L. Hall and J. Llinas, "Multisensor Data Fusion," in *Handbook of Multisensor Data Fusion*, New York, NY USA, CRC Press, 2001, pp. 1-1 to 1-10.

[4] A. N. Steinberg and C. L. Bowman, "Model, Revisions to the JDL Data Fusion," in *Handbook of Multisensor Data Fusion*, New York, NY USA, CRC Press, 2001, pp. 2-1 to 2-19.

[5] C. L. Bownam and A. N. Steinberg, "A Systems Engineering Approach for Implementing Data Fusion Systems," in *Handbook of Multisensor Data Fusion*, New York, NY USA, CRC Press, 2001, pp. 16-1 to 16-39.

[6] A. Sliva, J. Gorman, C. Bowman and M. Voshell, "Dual Node Decision Weels: an Architecture for Interconnected Information Fusion and Decision Making," in *Defense + Security Conference*, Baltimore, MD USA, 2015.

[7] A. Sliva, J. Gorman, M. Voshell, J. Tittle and C. Bowman, "Combining Cognitive Engineering and Information Fusion Architectures to Buils Effective Joint Systems," in *Defense + Security Conference*, Baltimore, MD USA, 2016.

[8] R. R. Hoffman and M. J. McCloskey, "Envisioning Desirments," *IEEE Intelligent Systems*, vol. 28, no. 4, pp. 82-89, 2013.

[9] M. Voshell, S. Guarino, J. Tittle and E. M. Roth, "Supporting Representation Management in Intelligence Analysis through Autormated Decision Aids," in *Proc. Human Factors and Ergonomics Society Annual Meeting*, 2014.

[10] E. Hollnagel and D. Woods, Joint Cognitive Systems Engineering: Foundations of Cognitive Systems Engineering, Boca Raton, FL USA: CRC Press, 2005.

[11] A. Ashoori and C. Burns, "Team Cognitive Work Analysis Structure and Control Tasks," *Journal of Cognitive engineering and Decision Making*, vol. 7, no. 2, pp. 123-140, 2013.

[12] R. Rousseau, J. Easter, W. Elm and S. Potter, "Decision-Centered Testing (DCT): Evaluating Joint Human-Computeer Cognitive Work," in *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, 2005.

[13] J. Benamati, M. A. Fuller, M. A. Serva and J. Baroudi, "Clarifying the integration of Trust and TAM in E-Commerce Environments: Implications for Systems Design and Management," *IEEE Transactions on Engineering Management*, vol. 57, no. 3, pp. 380-393, 2010.

[14] G. Klein, B. Moon and R. R. Hoffman, "Making Sense of Sensemaking 1: Alternative Perspectives," *IEEE Inteligent Systems*, vol. 21, no. 4, pp. 70-73, 2006.

[15] R. S. Kaplan and D. P. Norton, "The Balanced Scorecard: Measures that Drive Performance," in *Harvard Business Review on Measuring Coporate Performance*, Boston, MA USA, Harvard Business School Press, 1992, pp. 123-145.