Defence Research and Development Canada
Recherche et développement pour la défense Canada

DEFENCE R&D DÉFENSE

The decision centered testing methodology

Marie-Eve Jobidon
Richard Breton
DRDC Valcartier

Defence R&D Canada – Valcartier
Technical Memorandum
DRDC Valcartier TM 2008-155
July 2011

Canada

The decision centered testing methodology

Marie-Eve Jobidon
Richard Breton
DRDC Valcartier

Defence R&D Canada – Valcartier

Technical Memorandum
DRDC Valcartier TM 2008-155
July 2011

Principal Author

Original signed by Marie-Eve Jobidon

Marie-Eve Jobidon

Postdoctoral Fellowship / C2DSS

Approved by

Original signed by Patrick Maupin

Patrick Maupin

Section Head / C2DSS

Approved for release by

Original signed by Christian Carrier

Christian Carrier

Chief Scientist

This report has been produced following the Decision Centered Evaluation Capability project under the Canadian Technology Investment Funding program. The work unit sponsoring the production of this report is 11bu.

Abstract

The objective of a Decision Support System (DSS) is to assist human operators in the execution of the decision-making process by supporting cognitive limitations and reducing the risk of error brought by time constraints and uncertainty. Human and systems must work as a coordinated team to reach mission goals successfully, and therefore the evaluation of a DSS prior to its deployment is crucial. The Decision Centered Testing (DCT) methodology constitutes an innovative approach to evaluation that aims to test the decision-making effectiveness of the human-DSS team, called the Joint Cognitive System (JCS; Hollnagel & Woods, 2005). In particular, DCT focuses on critical decision points (known to be prone to human decision-making errors) rather than on the general outcomes of the decision process. Used as part of an iterative development process, DCT aims to provide insights into the strengths and weaknesses of the JCS prior to the DSS being implemented into an operational setting. Although much work remains to be done for the methodology to reach its full capability, DCT has the potential to make a significant contribution to the DSS evaluation process.

Résumé

L'objectif d'un système d'aide à la décision (SAD) est d'assister un opérateur dans l'exécution du processus décisionnel en supportant ses limites cognitives et en réduisant le risque d'erreur provoquée par les contraintes temporelles et la présence d'incertitude dans la situation. L'être humain et les systèmes technologiques doivent travailler en équipe afin d'atteindre avec succès les objectifs de mission. Ainsi, l'évaluation des SAD avant leur déploiement est crucial. La méthode d'évaluation centrée sur la décision (*Decision Centered Testing* ou DCT) présente une approche innovatrice pour l'évaluation de l'efficacité du processus de prise de décision de l'équipe humain-système (ci-après appelée système cognitive combiné; Hollnagel & Woods, 2005). Particulièrement, DCT se centre sur les points critiques de décision (sujet à l'erreur humaine) plutôt que le résultat final du processus décisionnel. Utilisée dans le contexte de développement itératif, la méthode DCT vise à fournir des indicatifs sur les forces et faiblesses du système cognitif combiné avant son implémentation dans un environnement opérationnel. Bien que beaucoup de travail reste encore à faire afin de permettre à la méthodologie d'atteindre son plein potentiel, elle peut déjà contribuer de façon significative au processus d'évaluation de SAD.

This page intentionally left blank.

Executive summary

The decision centered testing methodology:

Marie-Eve Jobidon; Richard Breton; DRDC Valcartier TM 2008-155; Defence R&D Canada – Valcartier; July 2011.

Introduction: The objective of a Decision Support System (DSS) is to assist human operators in the execution of the decision-making process by supporting cognitive limitations and reducing the risk of error brought by time constraints and uncertainty. Therefore, human and systems must work as a coordinated team to reach mission goals successfully. Because of the monetary costs and the stakes associated with the development, production and deployment of DSS in military C2 environments, the evaluation of a DSS prior to its implementation is crucial.

In order to test the effectiveness of a human decider paired with a DSS, it is necessary to complement current testing practices addressing software validation, human performance and usability. The Decision Centered Testing (DCT) methodology constitutes an innovative approach to evaluation that aims to test the decision-making effectiveness of the human-DSS team, called the Joint Cognitive System (JCS; Hollnagel & Woods, 2005). In particular, DCT focuses on conditions known to be prone to human decision-making errors. The development of this methodology stems from the need to design and build computer-based DSS that effectively support and become part of a cognitive team with their user in a complex, goal-oriented work domain such as military C2.

Results: DCT constitutes a new methodology that focuses on critical decision points rather than on the general outcomes of the decision process in order to assess the decision-making effectiveness of the JCS. The methodology consists of several steps, from describing the work domain to determining the JCS under evaluation to defining cognitive performance of the JCS.

Significance: Used as part of an iterative development process, DCT aims to provide insights into the strengths and weaknesses of the JCS prior to the DSS being implemented into an operational setting, with the objective of improving the JCS decision-making effectiveness. Although much work remains to be done for the methodology to reach its full potential, DCT has the potential to bring a significant contribution to the DSS evaluation process, particularly since it focuses on the human operator-DSS team (the JCS) rather than considering only the computerized decision aid.

Future plans: An initial validation effort has been completed (document in preparation). However, the DCT methodology could greatly benefit from additional validation and conceptual development, and solutions should be explored to make it more affordable time- and resource-wise. Importantly, DCT should to be tested with various decision-making tasks set in C2 environments to ensure that it is suited for the variety of decision-making contexts found in military operations.

Sommaire

The decision centered testing methodology:

Marie-Eve Jobidon; Richard Breton; DRDC Valcartier TM 2008-155; R & D pour la défense Canada – Valcartier; Juillet 2011.

Introduction: L'objectif d'un système d'aide à la décision (SAD) est d'assister un opérateur dans l'exécution du processus décisionnel en supportant ses limites cognitives et en réduisant le risque d'erreur provoquée par les contraintes temporelles et la présence d'incertitude dans la situation. L'être humain et les systèmes technologiques doivent travailler en équipe afin d'atteindre avec succès les objectifs de mission. En raison des coûts importants et des enjeux associés au développement, à la production et au déploiement de SAD dans des environnements militaires de commandement et contrôle (C2), l'évaluation des SAD avant leur déploiement est crucial.

La méthode d'évaluation centrée sur la décision (*Decision Centered Testing* ou DCT) présente une approche innovatrice pour l'évaluation de l'efficacité du processus de prise de décision de l'équipe humain-système (ci-après appelée JCS pour *Joint Cognitive Systems* ou système cognitive combiné; Hollnagel & Woods, 2005). Particulièrement, DCT se centre sur les points critiques de décision (sujet à l'erreur humaine) plutôt que sur le résultat final du processus décisionnel. Le développement de cette méthodologie provient d'un besoin de développer et de construire des SAD qui supportent l'humain au plan cognitif dans l'exécution de son travail dans des environnements complexes et orientés vers les buts, tels que le C2.

Résultats: L'approche DCT constitue une nouvelle méthodologie qui se centre sur des points critiques de prise de décision plutôt que sur les résultats finaux du processus décisionnel afin d'évaluer l'efficacité du JCS. La méthodologie inclut plusieurs étapes, débutant par une description du domaine de travail jusqu'à la définition de la performance cognitive du JCS.

Importance: Utilisée dans un contexte de développement itératif, la méthode de DCT vise à fournir des indicatifs sur les forces et les faiblesses du JCS avant son implémentation dans un environnement opérationnel. Bien que beaucoup de travail reste encore à faire afin de permettre à la méthodologie d'atteindre son plein potentiel, elle peut déjà contribuer de façon significative au processus d'évaluation de SAD en se concentrant sur le JCS plutôt qu'uniquement sur le système ou l'humain.

erspectives: Une première activité de validation a été complétée. Toutefois, la méthodologie DCT devrait grandement bénéficier d'études de validation supplémentaires et d'un travail de développement conceptuel. Également, des solutions devraient être envisagées pour rendre la méthode moins coûteuse en termes d'efforts et de temps requis pour son exécution. De plus, la méthode devrait être testée avec plusieurs types de prise de décision utilisés dans des environnements de C2.

Table of contents

List of figures

Acknowledgements

The authors would like to thank Dr Robert Rousseau for his contribution in this project.

1 Introduction

Today military operations take place in complex and dynamic command and control (C2) environments characterized by time constraints, high workload as well as uncertain and probabilistic intelligence. These environments put growing demands on individuals and teams in operational theatres and may thus jeopardize mission accomplishment due to human cognitive limitations. Therefore, there is a critical need for technological aids to support C2 activities, particularly the decision making process, and new means to improve support systems are constantly sought. Although advances in technology are in themselves of great importance, support systems must be developed with the objective of integrating human cognitive capacities and limitations in order to truly support the dynamic decision making process inherent in military operations. However, Chen and Lee (2003) assert that little research effort has been devoted to the cognitive aspect of decision support so far.

The objective of a Decision Support System (DSS) is to assist human operators in the execution of the decision-making process by supporting cognitive limitations and reducing the risk of error brought by time constraints and uncertainty. Therefore, humans and systems must work as a coordinated team to reach mission goals successfully. Because of the monetary costs and the stakes associated with the development, production and deployment of DSS in military C2 environments, the evaluation of a DSS prior to its implementation is crucial. However, system designers have typically proposed decision support solutions without an explicit knowledge of the decisions and cognitive needs requiring support (Paradis et al. 2002). Figure 1 illustrates the focus of two traditional evaluation approaches, software testing (left side) and usability testing (center). On the one hand, software testing is mainly oriented on the system's technological attributes and whether it matches required specifications in regards to the task to accomplish. On the other hand, usability testing is oriented towards the users' response and performance in regards to the DSS by examining the extent to which users are satisfied and can easily learn and use the DSS (see, e.g., Rousseau et al. 2005).

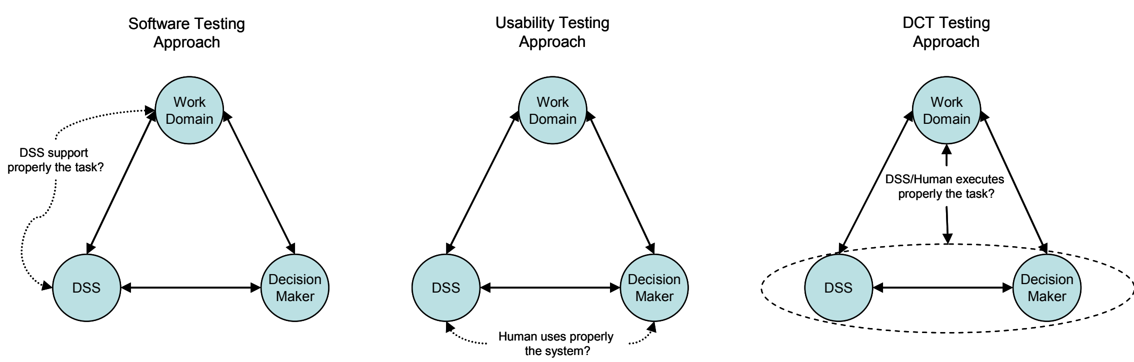

Figure 1: The focus of Decision Centered Testing (DCT) in relation to more traditional evaluation methods, software testing and usability testing.

Decision Centered Testing (DCT) illustrated on the right side of Figure 1 constitutes an innovative approach to evaluation that aims at testing the decision-making effectiveness of the coupling between a human decider and a DSS. It is based on the premises that the decision-making performance is highly influenced by:

- The decision-maker's competencies, skills and expertise;
- The efficiency of the DSS itself;
- The quality of the interaction between these two entities.

While the two types of testing techniques stated in the previous paragraph focus primarily on one entity (either the DSS or the human decider), DCT aims at testing the DSS-Human axis and assessing the combined efficiency of the human and DSS team. Then, the testing objective is not to evaluate if the system matches required specifications or to evaluate the user's satisfaction with regards to the DSS, but rather the evaluation of the human practitioner coupled with a DSS in regards to their combined capacity to efficiently perform the decision-making task. Consequently, the object of interest in the evaluation process is not the system or the human alone, but rather the human-system team, hereafter called the Joint Cognitive System (JCS; Hollnagel and Woods 2005).

In contexts such as military operations where mission accomplishment relies heavily on human-DSS coordination, focusing on decision centered metrics is critical to 'close the loop' of the C2 system design process. It allows iteratively improving the decision-making process by adjusting the decision aid to achieve the desired human-system synergy. Hence, it is crucial to develop metrics that will focus on the net decision-making effectiveness of a JCS.

The notion of *net* indicates that the object of testing is the decision-making jointly achieved by the human(s) and the DSS. *Decision-making* is defined here as the complete set of activities from:

- Collecting data due to alerts related to problems or as a result of casual monitoring through situation assessment;
- Comparing against desired states or goals;
- Planning for remedial courses of action if the current state and the desired state do not match satisfactorily;
- Executing the remedial courses of action;
- Evaluating the feedback resulting from the action against goal achievement.

It should be noted that this definition is significantly broader and more comprehensive than definitions used in the cognitive psychology research domain. Such definitions typically refer to decision making as the process of making a choice among a number of alternatives in order to achieve a goal, with a certain amount of information available on each alternative (see, e.g., Wickens 1992, Baron 2000). In comparison, the definition used here encompasses all phases of the OODA loop by including the processes necessary to collect information on the situation, select and implement a course of action. This is the result of the need, from an engineering perspective, to design and build computer-based DSS that effectively support and become part of a cognitive team with their user in a complex, goal-oriented work domain such as military C2.

In order to address this need for DCT, Defence R&D Canada – Valcartier (DRDC Valcartier) established a Technology Investment Funds (TIF) project titled Decision Centered Evaluation Capability (DCEC). The DCEC project aimed at developing a methodology and an innovative capability for assessing and measuring the net decision-making effectiveness of a JCS for military C2 applications, focusing in particular on conditions known to be prone to human decision-making errors. The overall goal of the DCEC project is to contribute to an improved effectiveness of JCS by providing means to test the quality of the cognitive support the DSS provides to the human decider, and more specifically the combined efficiency of these two entities, before committing a DSS prototype to production and deployment.

The DCEC project consists in a new methodology, called DCT, which focuses on critical decision points rather than on the general outcomes of the decision process in order to assess the decision-making effectiveness of the JCS. Rather than usability, DCT addresses the usefulness and understandability aspects of DSS and JCS evaluation:

- Usefulness: it evaluates the extent to which the JCS is effective in completing the mission and achieving goals;
- Understandability: it evaluates whether the DSS helps the operator to have a good understanding of the situation and to react efficiently to unexpected events.

In comparison to another decision centered approach, Decision Centered Design (DCD) (e.g., Wolf et al. 1991, Hutton et al. 2003), DCT consists of an evaluation methodology and focuses on the effectiveness of the DSS-human decider team (the JCS) rather than the effectiveness of the DSS itself. When put in parallel to the DSS development process described by Breton et al. (2001), it emerges that DCT may contribute to several steps of the process, including the identification of problems and deficiencies, the identification of technological solutions, and testing of technological solutions. Thus, employed in an iterative development effort, DCT is intended to provide insights into the strengths and weaknesses of the JCS prior to the DSS being implemented into an operational system, with the objective of improving the JCS decision-making effectiveness.

As presented before, testing has traditionally been focused on either the technological system or human decider (software testing vs. usability testing, see Figure 1). By particularly focusing on the interaction between the DSS and the human decider, the TIF project addresses the need for a testing approach that joins these two entities forming the JCS. As such, the purpose of the DCT methodology is to provide a framework to assess the decision making effectiveness of a JCS. The intent of DCT is to offer guidelines and principles to JCS testing that span and combine two main domains of knowledge associated with each entity of the JCS – system engineering and human factors – so that the methodology can constitute a useful tool to structure and guide the evaluation of a JCS in both domains.

1.1 Challenges in the realization of the TIF

The realization of this TIF project entailed two main challenges:

1. The project aimed at developing a new affordable and practical evaluation method that significantly departs from traditional assessment approaches. This method is based on a more

complex unit of analysis – the JCS, that is, the human operator and the DSS team. As the focus is on the JCS and not only on the DSS, the development of DCT relied on an understanding of both the human cognition and the DSS development process. Consequently, DCT required the combination of two domains of expertise: human factors and systems engineering.

2. Furthermore, the complexity of developing DCT lied in the fact that it departed from traditional approaches in regards to their focus on the outcome of the decision at the end of a scenario to assess the quality of a DSS. Contrastingly, DCT was meant to allow identifying critical decision-making points in a given C2 work domain and to include metrics to specifically assess the decision-making effectiveness of the JCS for these critical points.

1.2 Content of this report

This report describes the DCT methodology that was developed within the TIF project. The document is organized in the following manner. Section 2 presents the conceptual basis for DCT. Section 3 defines each step of the methodology, as illustrated in Figure 2. Section 4 discusses challenges in applying DCT, and Section 5 presents a brief conclusion and recommendations for future work.

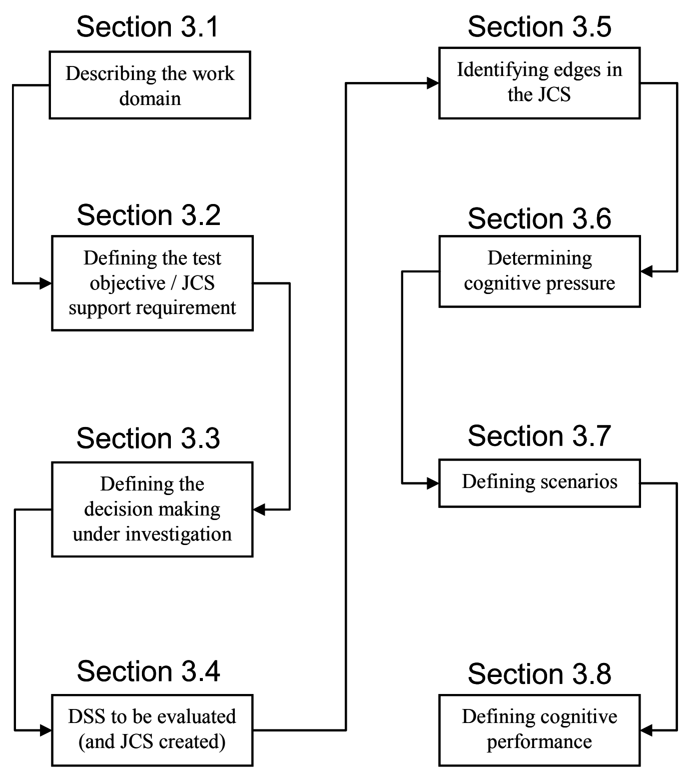

Figure 2: Steps of the DCT methodology and associated sections.

2 Conceptual basis for DCT

The conceptual basis for DCT entails four principal premises:

- The fundamental unit of analysis is the JCS;
- JCS requires unique types of decision support;
- The effectiveness of a JCS depends on the information mapping from the work domain to the end user in terms of human-computer interface;
- Decision-centered evaluations must be based on a functional analysis of the target work domain.

2.1 JCS as the unit of analysis

The *first premise* underlying DCT is that the fundamental unit of analysis is the JCS (Hollnagel and Woods 2005). This notion is derived from the Cognitive Triad, a basic concept in Cognitive Systems Engineering (CSE; Woods and Roth 1988a, 1988b) that represents a JCS as an intra-related network connecting three nodes, the Work Domain, a Human Agent, and an Artifact (a DSS) (see Figure 1, p.1). Therefore, a JCS can be defined as the combination of human decider and support technologies which must act as co-agents to achieve goals in a complex work domain, in a coordinated and effective manner. In the context of DCT, this joint effort of human and DSS to achieve mission objectives within a specific work domain (e.g., above-water warfare) constitutes the net decision-making that is evaluated.

The ultimate implication of this perspective is that the human and DSS must form a team and have a high degree of 'cognitive coupling' between them in order to be effective (i.e., they have to fit at the cognitive level). That is, the human decider must be able to interpret the meaning of the DSS's computer driven images and displays, and the DSS must be effective in enabling the human decider to decode the world and its events. Hence, in complex C2 environments characterized by time constraints and uncertainty, CSE shifts the issue of DSS development and evaluation from overcoming human cognitive limits to supporting adaptability and control (Woods and Hollnagel 2006). Supporting effective decision making for such JCS requires an understanding of the functional constraints and demands of the work domain in addition to an understanding of the perceptual and cognitive capabilities of human deciders involved (Elm et al. 2003).

Figure 3 depicts an adaptation of Hollnagel and Woods (2005) cyclic model of the function of JCS. The cycle emphasizes that current decision making builds on previous decisions and anticipates future decisions, a key aspect of dynamic decision making (Brehmer 1992, Clancy et al. 2003). This anticipation is a fundamental component of efficient military C2. Indeed, military C2 environments are complex and dynamic by nature and the projection of allied and adversary capabilities and actions in the future is a critical dimension for the successful accomplishment of mission objectives. The importance of anticipation has incidentally been acknowledged in models of situation awareness (SA) and decision-making. For instance, the third level of Endsley's model of SA (1988, 2004) consists in the extrapolation of the state of the elements in the environment in the near future. This stage of SA is about understanding how events will unfold, which is critical in achieving operational superiority. Similarly, an integral component of Klein's naturalistic decision-making model – the Recognition-Primed Decision Model – involves story building, that is, context-specific mental simulations of a likely course of action to evaluate how it will fare once implemented in the actual situation (Klein 1989, 1997). While the anticipation aspect of Endsley's and Klein's models focuses on the human node of the triad, in Hollnagel and Wood's cyclic model the extrapolation process is achieved by the JCS. Hence, human deciders and support systems must work as a coordinated team to be successful in reaching mission goals.

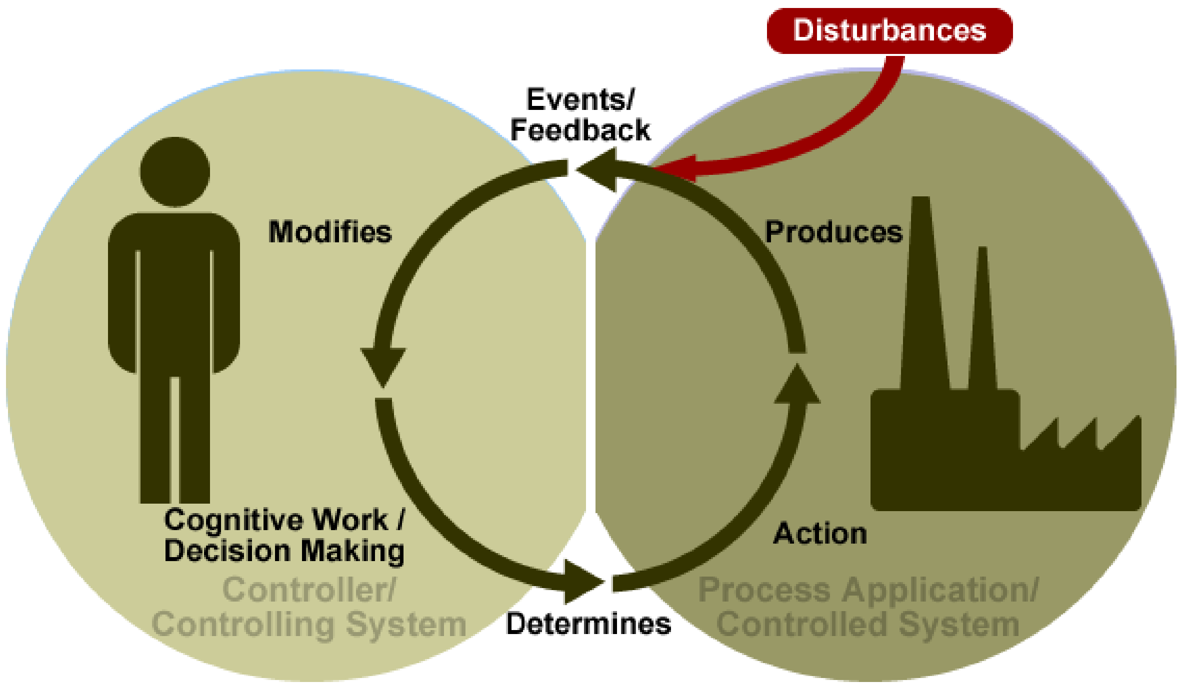

Figure 3: Cyclical model of a Joint Cognitive System (adapted from Hollnagel and Woods 2005).

2.2 JCS require unique types of decision support

The *second premise* at the basis of DCT is that because of the focus on the human decider-DSS team, JCS require unique types of decision support that take the combination of the two entities into account. The core of the JCS perspective is that DSS and human decider must coordinate and work as a team to accomplish cognitive work efficiently. Instances of this cognitive work may be (Rousseau et al. 2005):

- For the DSS to represent events occurring in the world in such a way that the human decider can comprehend them and assess their consequences.
- Determining suitable courses of actions to reach the desired goal.

From JCS development efforts and research in CSE, a series of generic support requirements for JCS have been identified (e.g., Billings and Woods 1994, Christoffersen and Woods 2002, Elm et al. 2005). These requirements comprise the attributes and characteristics that a JCS must possess in order to be able to effectively accomplish cognitive work (e.g., comprehend current and upcoming events, make decisions) in a goal-directed task such as military C2 (Rousseau et al. 2005). As such, these requirements can serve as criteria for evaluating JCS efficiency (Woods and Hollnagel 2006).

These support requirements include:

- Observability – the ability to form insights into a process (either a process in the work domain or in the DSS), based on feedback received. That is, observability notably allows the DSS and human decider to share a common understanding of the situation or problem (e.g., what is the current situation, how is it evolving, what strategies are appropriate to reach mission goals given this situation). Also, it allows each entity of the JCS (DSS and decider) to develop a representation of the other entity's activities (e.g., what aspect or event is each entity currently focusing or working on, what relevant information do they have, what potential strategies do they favor) (Christoffersen and Woods 2002). Observability overcomes the 'keyhole' effect and allows the JCS to get a broader picture of the situation; that is, to see sequences and evolution of events over time, potential future activities and contingencies, and underlying patterns and relationships in a process. For instance, in the context of military operations, observability can allow an officer to formulate timely and appropriate courses of actions with the assistance of a DSS that highlights unexpected changes in the status of adversary capabilities or actions.

- Directability – The ability to direct/redirect resources, activities, and priorities as situations change and escalate. Directability allows the decider to effectively control the processes in response to (or in anticipation of) changes in the environment. This requirement is of particular relevance in military C2. Indeed, it can contribute to gaining a tactical or operational advantage over the adversary by allowing the JCS to be more flexible and to adapt efficiently to complicating factors or unexpected events occurring over the course of a mission.

- Teamwork with agents – Military operations take place in complex and dynamic environments that typically involve the coordination of a number of agents, both humans and systems (e.g., the ops room of a frigate). Teamwork with agents is the ability to coordinate and synchronize activity across agents. This defines the type of coordination (e.g., seeding, reminding, critiquing) between agents. Teamwork with agents allows decider(s) to effectively re-direct agent resources as situations change. This efficient coordination between all the elements forming the JCS is critical for mission accomplishment.

- Directed attention – The ability to re-orient focus in a changing world. This includes issues like tracking others' focus of attention and the ease with which they are interrupted. In complex military environments, it is particularly relevant in regards to operational space monitoring, which depends on the operator's ability to detect change and the capability of a DSS to appropriately support this process. Directed attention allows the human-system team to work in a coordinated manner, contributing to increased situation awareness and decision-making effectiveness.

- Resilience – The ability to anticipate and adapt to surprise and error. This includes issues such as failure-sensitive strategies, exploring outside the current boundaries or priorities, overcoming the brittleness of technological aids, and maintaining peripheral awareness to maintain flexibility. Considering the ever increasing complexity of military environments, resilience constitutes a key requirement to gain and maintain operational superiority.

In the context of DCT, these support requirements define the focus for a particular test of the JCS. For instance:

- Observability – Testing for observability might entail assessing the insights into a situation that an officer is able to gain through the representation provided by the DSS.
- Directability – Testing for directability might involve assessing the extent to which an officer is able to modify the engagement sequence proposed by the DSS in regards to intelligence or events that the DSS may be unaware of.
- Teamwork with agents – The entities (or agents) forming a JCS can potentially have access to different information. Therefore, testing for teamwork with agents could entail assessing the extent to which the entities are capable of critiquing each other based on the respective information they have access to, and of taking such critique into consideration.
- Directed attention – Testing for directed attention might involve assessing the ability of the JCS to shift focus from one problem to another in a multi-threat situation.
- Resilience – Testing for resilience might entail assessing the ability of the JCS to detect a failure of the strategy implemented to achieve mission goals, and to prevent it by taking corrective actions.

2.3 Effectiveness of JCS depends on information mapping

The *third premise* is that the effectiveness of a JCS depends on its human-computer interaction's information mapping from the work domain to the human decider in terms of human-computer interface. The relationship between the visual structure established by a particular DSS display element and the constraints and relationships of the domain itself is fundamental to the effectiveness of the visualization. This premise implies that without an explicit specification of this mapping, it is impossible to determine if the visualizations are supporting deciders' needs as intended, or making the "supported" task more difficult. Compared to the research that has been done on human cognition, very little work has been done to describe the representational nature of visualizations in complex information processing tasks such as C2 environments (Zhang 1996).

Fundamental to this premise is the 'Representation Effect.' This refers to the finding that the way in which a problem is represented influences the cognitive work needed to solve that problem (Norman 1988, 1993, Zhang and Norman 1994). In other words, the design of artefacts in the work domain either improves or hinders our perceptual and cognitive capabilities. Therefore, our comprehension of information displays cannot be achieved in terms of purely visual characteristics, as it must also include an understanding of the representational requirements of the work domain (which Woods 1991a, 1991b, calls the 'Mapping Principle'; see Figure 4). In order to contribute to efficient mission accomplishment, the goal is thus to reveal the critical information requirements and constraints of the decision task through the user interface implemented in the DSS in such a way as to capitalize on the characteristics of human perception and cognition. With the highly complex and dynamic nature of military operations, this can serve to limit the probability of errors and the impact of factors such as uncertainty, time constraints, and workload.

For DCT, the Mapping Principle defines evaluation criteria focused on the degree of correspondence between information requirements and DSS interface instantiation. This defines requirements for assessing what is most typically referred to as situation awareness. Also, it defines requirements for the events and situations (i.e., evaluation scenarios) that need to be reflected in the interface. Finally, since the mapping principle includes information about the automation, it defines new representational requirements resulting from any shift in level of autonomy.

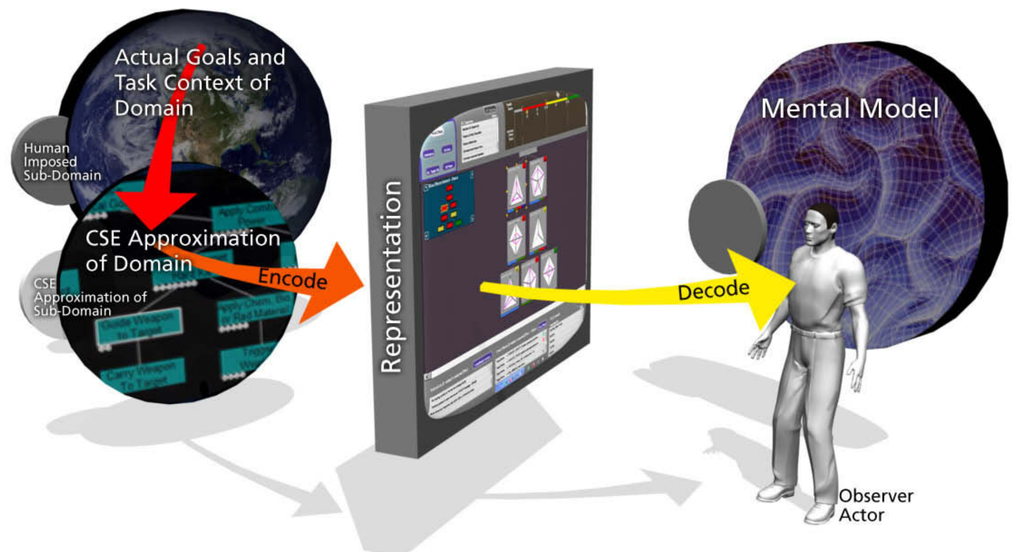

Figure 4: The symbol mapping principle (adapted from Woods 1991a, 1991b).

2.4 DCT is based on a functional analysis of the work domain

The *fourth premise* that constitutes the conceptual basis of DCT is that decision-centered evaluations must be based on a functional analysis of the target work domain (e.g., threat assessment and weapons assignment). Achieving a high level of decision-making effectiveness by the JCS requires that the human and DSS share an effective model of the fundamentals of the work domain. It has been shown that an explicit functional model of the work domain closely parallels experts' mental models faced with complex, uncertain naturalistic decision making conditions (Elm et al. 2003). For the DCEC project, the Applied Cognitive Work Analysis (ACWA; Elm et al. 2003) methodology was adopted to define this functional basis for the evaluations. ACWA is a comprehensive, robust, and proven approach to cognitive analysis that stems from formal, analytic goal-means decomposition methods (Woods and Hollnagel 1987, Rasmussen 1986, Lind 1991, 1993, Vicente 1999). This methodology creates a Functional Abstraction Network (FAN), a function-based goal-means decomposition of the work domain that identifies the goals to be achieved in the domain and the means available for achieving them. In a case study focusing on military C2, Potter et al. (2003) demonstrated that this approach can be valuable in designing decision support tools for complex processes such as those faced by military officers, by modeling the essential underlying concepts and relationships in the work domain that need to be an integral part of the support tools.

3 The DCT methodology

3.1 Describing the work domain

As the starting point to DCT, it is essential to determine a particular context in which the JCS will be evaluated, and that will serve as focus of the evaluation. To start this process, the work domain must be understood. The 'work domain' is the context in which the JCS operates, that is, the set of constraints, boundary conditions, forcing functions, etc. and resulting events that occur. In order for humans to influence the work domain so as to achieve one or more desired goals, they must usually undertake a set of interdependent decisions and actions, often in a timely manner and sequence. These decisions and actions are demands that are required by the work domain without respect for the capabilities, or lack thereof, of the humans that wish to influence the events in the work domain.

For the purposes of DCT, the demands and constraints of the work domain must be understood in order to appropriately specify the context for the evaluation. As established in the fourth premise, this can be accomplished very effectively by a work domain analysis, for instance using ACWA (see, e.g., Potter et al. 2003). The objective of this functional analysis is to model the decision space, that is, to construct a structured representation of the goals to be achieved as well as the processes required and means available to achieve them. While typically this analysis serves as the context for a DSS to be designed, in DCT it is also used to identify potential target areas for the evaluation. For example, particularly complex functional nodes or relationships between functional (goal-means) nodes indicate exceptional demands on the decision-making of the JCS, and potential brittles areas where decision-making effectiveness is susceptible to break down.

This first phase is the basis on which lie the subsequent steps of the DCT methodology. If the work domain analysis lacks specification, it will limit the significance and value of the lessons learned drawn from the results. Therefore, the analysis must be sufficiently detailed to provide an adequate and useful comprehension and representation of the work domain, and allow the DCT methodology to be implemented to its fullest potential.

3.1.1 Example

In military C2, one instance of work domain is Threat Evaluation and Weapons Assignment (TEWA), as found in Operations Rooms of Canadian Navy frigates when engaged in above-water warfare. This work domain requires a set of actions and decisions, such as contact detection and identification, threat assessment, target acquisition and combat power application, in order to achieve mission objectives. TEWA in above-water warfare is characterized by severe time constraints, and failure to make timely and adequate decisions can result in lethal consequences. Figure 5 illustrates an example of a simplified FAN of the TEWA work domain.

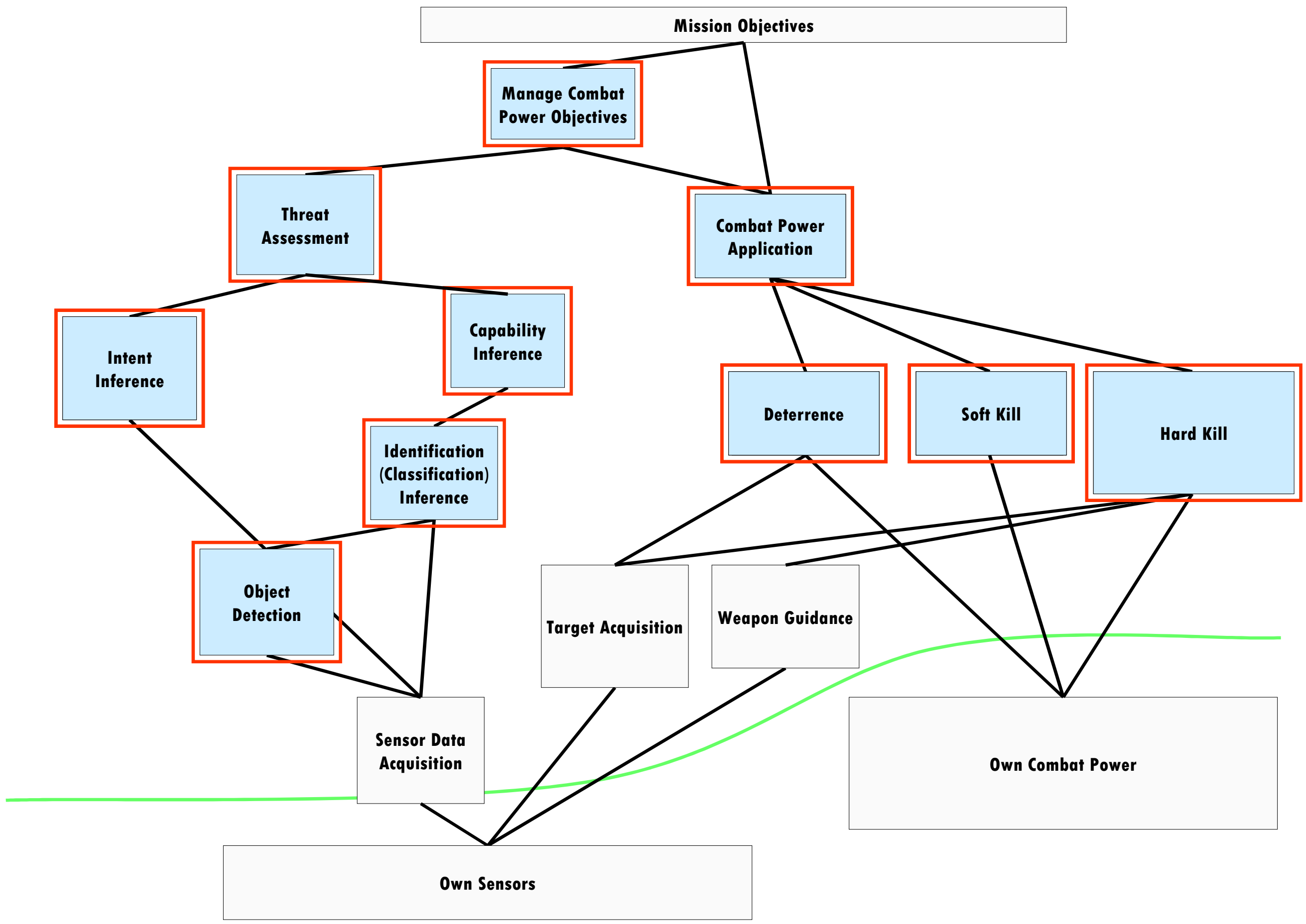

Figure 5: Simplified FAN for the TEWA work domain.

3.2 Defining the test objective / JCS support requirement

Once the particular target area within the work domain has been identified, it is possible to define the particular test objective. Because the JCS support requirements (observability, directability, teamwork with agents, directed attention, resilience) identify the attributes or characteristics of an effective JCS, they represent the focus and aim of the evaluation. Therefore, they constitute the objectives of DCT and provide focus for the detailed test design.

Defining the test objective consists in selecting a JCS support requirement that, within a given work domain, is particularly likely to maximize JCS efficiency. However, the task of selecting which JCS support requirement to choose as the test objective for a particular test is not straightforward. That is, not all of the JCS support requirements are equally relevant or appropriate to a particular decision-making requirement. For example, a given decision that requires aspects of the work domain being made 'observable' in order to assure an effective JCS may not mean that the same decision is dependent on these aspects being 'directable'. The selection of a decision-making focus thus needs to be paired with the selection of the more appropriate JCS support requirement as the test objective. Hence, determining the JCS requirement most suitable as test objective for a given DCT depends on which requirement appears the more critical with regards to the cognitive work to accomplish and the context (work domain) in which this work will be carried out.

Defining the test objectives should include the elaboration of a hypothesis statement. In order to provide as much incentive as possible to designing a test that seriously stresses the cognitive work expressed in the test objective, the test hypothesis should assume that the prototype DSS is incapable of providing adequate support, that is, make the prototype DSS prove that it is capable even when faced with the most demanding world events / circumstances. In other words, the goal is to push the testing envelope as much as possible. By wording the hypothesis in this way, it encourages the designer of the test environment to create an environment that is difficult to properly address.

If the test design does not force the JCS that is being tested to work with the most cognitively demanding instances in the work domain, then the results of the testing cannot be construed to assure that the cognitive team will perform correctly when those instances occur, and the tests will have demonstrated very little. Thus, preliminary testing may be required prior to the completion of the DCT to insure that the cognitive work is sufficiently demanding to tax the JCS.

Hypothesis statements should make specific predictions regarding the DSS being evaluated, particularly in a DCT involving the comparison of multiple DSS. For instance, the hypothesis could specify whether or not one or more of the DSS are expected to show more observability than the others.

3.2.1 Example

For a DCT set in the context of TEWA in above-water warfare, the JCS support requirement of observability can be identified as a test objective. Observability is a key requirement for JCS decision-making effectiveness as it refers to the ability of the JCS to form insights into the events taking place (represented in the DSS) that will help determine the most effective engagement response (Rousseau et al. 2005). Note that other JCS support requirements could have been selected and could be part of other subsequent testing.

The main goal in that work domain is to defend own-ship against missile attacks. The successful completion of this task by the JCS requires, notably, to properly determine the threat level of objects in the operational space and to determine the appropriate and tactically correct engagement sequence of targets. The related hypothesis for this test objective can be formulated in the following manner:

> The DSS that is under test (i.e., the emulator) provides insufficient observability for the test subject to select (for engagement) the tactically correct threat in regions of known decision-making brittleness / difficulty.

3.3 Defining the decision-making under investigation

For the DCT methodology to be applied, it is essential that the specific decision-making to be assessed be clearly identified and defined. Obviously, the decision-making under investigation must be within the scope of support of the DSS. In the context of the DCT methodology, the term ‘cognitive work’ or ‘decision-making’ refers to the set of activities that include:

- data collection due to alerts related to problems, or as a result of casual monitoring;
- state identification;
- comparison against desired states or goals;
- planning for remedial action if the current state and the desired state do not match satisfactorily;
- remedial action execution;
- evaluating the feedback resulting from the action against goal achievement (from Rasmussen 1986).

As mentioned before, this definition encompasses all phases of the OODA loop, by including activities related to data gathering, situation assessment, decision making and implementing COAs.

Defining the decision-making under investigation can be achieved through the application of a variety of cognitive analysis approaches. However, the important issue is that the demands and constraints that make the identified cognitive work difficult be identified as well. In this way, insights into potential manipulations for the test conditions can be derived.

As stressed before, it is critical for DCT that the description and analysis of the work domain be sufficiently detailed in order for subsequent steps, such as the identification of the decision-making under investigation, to be completed properly. For instance, in both the planning and results analysis phases of a DCT test set in above-water warfare (document in preparation), it was apparent that the work domain analysis descriptions of the cognitive work and the supporting information that is necessary to perform it lacked sufficient detail. This finding was linked to most steps in this DCT methodology test. That is, an underspecified description of the cognitive work and its supporting information needs will affect virtually every step of the methodology and limit the results and insights gained from it.

3.3.1 Example

For a DCT set in the military above-water warfare domain, the decision-making under investigation can be "determine the engagement sequence against a set of incoming missiles". The demands for this decision-making include temporal (i.e., time to decide) and volume (i.e., number of incoming missiles) aspects. The goal is to identify the best possible engagement response to avoid being hit by adversary missiles. The cognitive difficulty lies in identifying the threat level of each incoming objects and to engage threatening targets in a tactically correct order within the time constraints dictated by ongoing events.

3.4 The DSS to be evaluated (and the JCS that is created)

In order to fully specify the JCS (and thus conduct a DCT), the support tools to be used by the decision-maker must be specified. Though not necessarily limited to computer based implementations, the decision-making aids considered in this project are computerized DSS. DSS evaluated in DCT can be either support systems already implemented in operational environments or DSS that are in the development stages. Therefore, it is important to define the following characteristics of the DSS:

- the types of representations (i.e., information mapping) provided;
- the degree of automation;
- the nature of the user interaction requirements;
- the scope of responsibility (i.e., the cognitive demands that are supported and those that are not supported).

This description of the DSS provides the basis for the identification of potential edges in the JCS (to be described in the following subsection). The critical issue is that a wide spectrum of JCS is created based on the characteristics of the work domain and the DSS, and these characteristics must be described in order to define the focus for the DCT.

3.5 Identifying edges in the JCS

Given that the focus of the evaluation is on the decision-making of the JCS as a single unit, it is important to define potential weaknesses or brittle points in the JCS where decision-making effectiveness may break down under pressure. Within the DCT methodology, these weaknesses are referred to as 'edges'. An edge is any discontinuity within any of the relationships between the three nodes of the cognitive triad (see Figure 6). This could include:

- Events or changes in the work domain that are not represented well by the DSS (on the Work Domain-DSS axis);
- Automatic control actions taken by the DSS but not communicated to the decider (on the Human – DSS axis);
- Transition points within an algorithm that are not represented to the decider (on the Human – DSS axis);

The decider's perceptual or cognitive capacities limit the comprehension of information coming from the domain (on the Human-Work Domain axis) or provided by the DSS (on the Human – DSS axis).

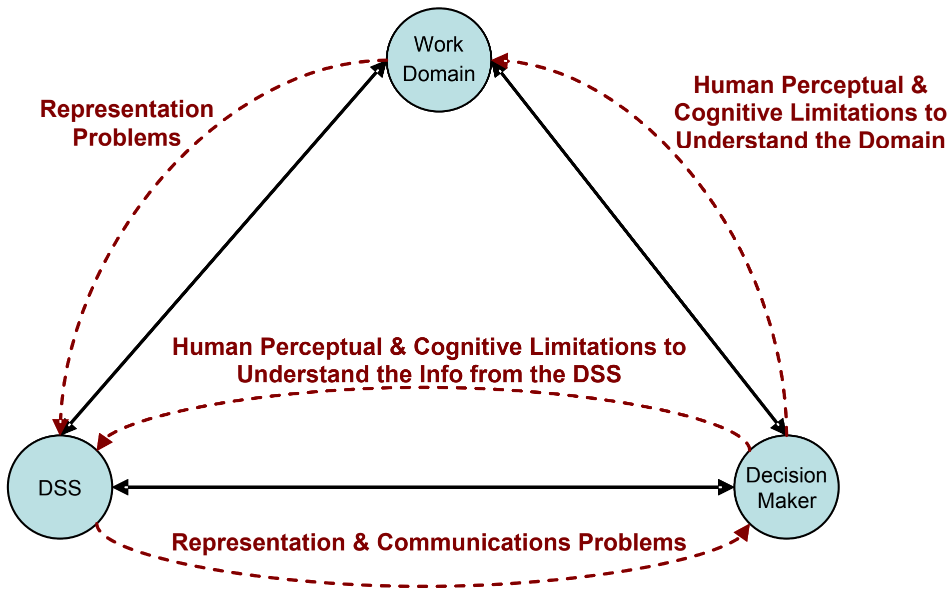

Figure 6: Illustration of the edges in a JCS for the purpose of DCT.

Edges can be used individually or combined, as the test planner decides in regards to test objectives and the work domain in which the test takes place. DCT attempts to test these edges by assessing the decision-making effectiveness of the JCS as it works across an edge. A successful transition across a brittle point is a characterization of a resilient JCS, while decision-making breakdown at an edge is the mark of an ineffective JCS.

The notion of edge addresses the issue of limited coverage inherent to DSS testing and validation. Indeed, no method of evaluation can in itself encompass all aspects of a support system. By focusing on edges, DCT relies on a form of critical testing that deals with specific areas of the work domain where the JCS is susceptible of failing.

It is important to note that an edge is not equivalent to simply overloading the JCS by domain events until a decision-making breakdown occurs (e.g., the number of simultaneous incoming missiles). While the focus of an evaluation is often to identify any improvement in the maximum overall capacity of the JCS, DCT is focused on the latent, brittle edges in the JCS design.

3.5.1 Example

In order to illustrate the distinction between testing an edge for the JCS and overloading the JCS capacity, consider the above-water warfare domain as an example. An edge in this domain might be the combination of heterogeneous bearing/distance/speed of incoming missiles to which the operator is unable to find an effective solution in real time. In comparison, overloading the JCS would simply be presenting a large number of incoming missiles on the same bearing at the same speed and distance that could be effectively engaged without difficulty.

3.6 Determining cognitive pressure to apply to the edge(s)

Once the particular edge has been identified, it is important to determine the most appropriate pressure to apply against it. DCT stresses the JCS through the application of this cognitive pressure on the three basic relationships (defined by the axes) in the cognitive triad in an attempt to cause differences in decision-making effectiveness (see Figure 6).

Within a particular test, cognitive pressure is the variable that is manipulated, that is, varied by the test planner. As mentioned before, cognitive pressure does not simply correspond to an overloading of the JCS. Rather, it is a carefully crafted manipulation of controllable aspects of:

1. The work environment related to JCS support requirements (observability, directability, etc.);

2. Cognitive demands of the work domain (e.g., decision-making, monitoring, scheduling);

3. And/or edges in the JCS (e.g., sudden changes in the work domain not detected correctly by the DSS, information coming from the DSS or work domain to complex to be understood by the decider).

One of the critical aspects of cognitive pressure in DCT is that the manipulation is based on the hypothesized edge and therefore is focused on brittle points where the JCS is susceptible of breaking down (i.e., critical testing).

3.6.1 Example

Here are some examples of potential cognitive pressure in the above-water warfare work domain. Cognitive pressure could be induced by manipulating:

- Time available to determine a tactically correct engagement sequence of targets;
- Number of targets;
- Uncertainty regarding the threat level of targets;
- Uncertainty regarding the identity of contacts (friend, neutral or foe);
- Occurrence of unexpected contacts close to own-ship;

Sudden changes in the properties of contacts (e.g., bearing, time to closest point of approach).

3.7 Defining scenarios

Up to this step in the methodology, the DCT is mostly designed without regard for specific events in the world. However, in order to test the complete sequence involved in the CSE Mapping Principle (as illustrated in Figure 4), the test designer must now develop a specific set of events in the world and set them into a time reference, that is, develop a test 'scenario' that can be used to initiate the DSS's encoding processes.

Defining scenarios to instantiate the desired cognitive pressure to stress the edge under test constitutes the most difficult and also most important contribution of DCT. Scenarios include events related to the work domain and identify required or expected behaviors from the DSS, and/or decider. Three categories of scenarios can be defined:

1. Scenarios can be derived from the specific JCS support requirement being tested (e.g., observability);

2. Scenarios can be derived from the specific decision-making demands of the work domain being tested (e.g., determining an engagement sequence);

3. Scenarios can be derived from the specific 'edge' under investigation (e.g., whether a DSS highlights the most relevant information to achieve mission goals).

For example, in the case of scenarios derived from specific decision-making demands, a DCT focused on the cognitive work of inferring degree of threat of various contacts in a military C2 domain will need to include characteristics of what makes this decision difficult (e.g., missing or inconsistent data) in the scenarios.

Changes to developed scenarios should be made only after careful consideration of their impact on the test design. There is often great temptation to 'add something' to the scenario as a clearer realization of what it includes and what possible consequences will be invoked, particularly during test execution. However, these 'on-the-fly' changes should be avoided, unless they have been carefully thought out as to their impact on the underlying validity of the DCT test design. Among possible consequences, variations to a scenario may invoke complicating factors. For example, consider a situation in the above-water warfare domain in which the observability support requirement is the object of interest of the testing activity. Adding additional airborne objects in a scenario (either threatening or non-threatening) so as to add to the cognitive pressure would potentially compromise the test objective. Indeed, this additional cognitive work could take the form of 'distractions', which would lead to the need for attention refocusing mechanisms. Should that be the case, then the JCS support requirement 'directed attention' would become the focus or test objective, and as such will likely reduce the test's focus on the test objective of 'observability'. As a result, this potential 'dilution' must be accounted for when analyzing the test results and when drawing conclusions. Without careful test design such changes become 'confounds' in the DCT test plan's execution.

Also, the potential effect of training during the test and its impact on test results should be taken into account when developing scenarios. Test participants (people playing the scenario) may exhibit an 'asymptotic' behavior as the trials in a given test progress (e.g., the time taken by the test subject to complete the cognitive work during the first trial will likely be the longest, but will continuously shorten and approach a constant value with succeeding trials as the test subject learns the scenario). This phenomenon is explained by the learning curve of participants. Temptation to counteract this training effect by changing a scenario in a way that the indigenous training from previous trials is no longer relevant during a given trial is hazardous. One must be very cautious in doing so since changes are likely to affect or dilute one or more of the test design constraints / controls, the most likely of which are the chosen test objective and the test hypothesis. Therefore, parameters of a scenario that will be varied in order to counteract the effect of training should be selected so as to ensure that the test objective is not affected.

3.8 Defining cognitive performance

Defining cognitive performance within the DCT methodology focuses on measuring the cognitive work under investigation. These metrics must be indicative of JCS decision-making effectiveness. As noted earlier, the focus of DCT is not on assessing the overall outcome of the decision-making (if the decision is good or not), but rather the effectiveness of the decision-making process in achieving this outcome. Thus, for a given scenario, two different JCSs could result in the same outcome (the COA selected by the JCS A is not better than the COA of the JCS B) but with very different measures of decision-making effectiveness (the decision-making process achieved by the JCS A is much better than the one of JCS B). Therefore, cognitive performance in DCT is assessed primarily through measures of effectiveness (MOEs) rather than measures of performance (MOPs). That is, instead of assessing task accomplishment (MOPs), selected metrics in DCT must be focused on measuring the extent to which the JCS is effective when working across critical decision points in the effort to attain a specific objective (MOEs).

4 Challenges in applying DCT

One of the main objectives of the DCEC project was to develop and validate a test methodology that is intended to provide a major breakthrough in the evaluation of the decision-making effectiveness of a JCS, that is, a combined human agent / DSS team (the JCS). The following subsections present some of the challenges in applying the DCT methodology.

4.1 Challenges in defining the test objective / JCS support requirement

The DCT methodology focuses on five different JCS support requirement: observability, directability, teamwork with agents, directed attention and resilience. So far in the DCEC project, only observability has been tested (document in preparation).

In essence, observability implies that a given DSS must make visible work domain characteristics such that the decider can understand the meaning of this information and formulate the correct response to it. Therefore, observability can be broken down as:

[Visibility → Understanding → Correct Response] = Decoding (DSS interface).

For instance, the requisite data for completing the cognitive work (e.g., decision-making) must be visible. Specifically, the data must reside somewhere in the DSS's displays. This can be tested by a 'by-inspection' test, that is, by looking through the DSS's display set and comparing the data available against a list of the requisite data. Also, the 'correct response' for the cognitive work that is under test is the focus of test variable determination. The correct response may not necessarily be an action; the correct response may be a conclusion that does not require physical action. This realization may help select the test method or suggest that multiple methods should be employed.

It is worthwhile noting that the middle part of observability in the definition mentioned above is 'understanding'. By focusing on a precise and limited piece of cognitive work, the DCT methodology enables a test planner to limit what is meant by 'understanding'. This may, again, suggest selecting or including in the test plan a test method that will collect data that is indicative of what the test subject did or did not understand when faced with the test's cognitive work.

The partition of components provides insight on the test planning with the objective of observability, which can also be extended to the other JCS support requirements. More work focused on the other support requirements is required.

It is important to note that the JCS support requirements are recent insights from research in CSE (e.g., Christoffersen and Woods 2002, Elm et al. 2005) and remain somewhat immature. For instance, it is not yet established exactly the level of independence that each JCS support systems requirement has with the others and there is some concern that some of them are not independent. Research in this area will need to be monitored and results will need to be included as updates to the DCT methodology.

4.2 Challenges in selecting the edge

In the context of DCT, identifying the edge to be investigated constitutes a critical issue in consequently defining the different scenarios to be used, the cognitive pressure, as well as the cognitive performance metrics. However, a clearer definition and understanding of ‘edge’ are needed, notably since this notion is key to the construction of a DCT scenario. An approach to the ‘edge’ definition might be to define classes of edges, for instance “exceed the DSS/technological aid”, “break the algorithm within the technological aid”, “discontinuities in natural phenomenon / processes”, etc. These classes can then act as guides to searching the cognitive work requirements (CWRs) and information requirements (IRRs) from the work domain to locate work domain instances in which such ‘edges’ exist. This may then suggest events to be used in test scenarios to pressure these edges, and aid in the selection of test variables that would identify the edges in the test data.

4.3 Challenges in defining cognitive pressure

Defining the appropriate cognitive pressure to the test can be very difficult. In fact, the concept of “cognitive pressure” needs to be clarified in itself. For instance, it could be useful to define categories or attributes of cognitive pressure (e.g., time pressure, uncertainty). In complex C2 environments, defining cognitive pressure may be particularly challenging. Indeed, the complex nature of military C2 environments can complicate the selection of the appropriate cognitive pressure to apply on the edge.

Therefore, multi-thread environments such as military C2 require a close analysis in the determination of cognitive pressure, as one single factor may not be sufficient to exert pressure on an edge significant enough to get JCS effectiveness to break down. With the intrinsic time constraints of C2 environments, time available can be a relevant factor through which apply cognitive pressure on an edge. As with other potential variables however, it is important to bear in mind that cognitive pressure must be assessed prior to conducting the full process of DCT – e.g., through pilot studies – in order to ensure that it challenges JCS effectiveness.

The dynamic nature of C2 environments raises another issue with regards to cognitive pressure. If factors such as time available are used to define cognitive pressure, one must take into account potential carry-over effects. That is, for one given decision within a scenario, the test planner should assess the actual total time available for all aspects of test subject decision making. Although trying to prevent these overlaps could affect the naturalistic validity of the DCT, they should be taken into account in the determination and analysis of cognitive pressure.

4.4 Challenges in generating scenarios

Defining scenarios based on edges within the JCS and derived from characteristics of the work domain’s functional structure is a tremendous insight provided by DCT. However, it is important to build scenario that are neither too simple nor too difficult. On the one hand, scenarios that are too simple would not foster or stimulate much cognitive work. On the other hand, scenarios that are too complicated could lead the participants to give up on the task. In any case, this point to the need for a tighter coupling between the world events in the scenario and the cognitive work and the associated information requirements that they are intended to stimulate.

4.5 Challenges in analyzing cognitive performance

The critical issue in defining the cognitive performance metrics for these DCT tests is the creation of scenarios around the edges to be investigated. The notion of cognitive performance aims to detect limitations in the JCS decision-making effectiveness, with the idea of improving DSS design and foster task accomplishment. As with previous steps of the methodology, a tight coupling between cognitive performance and cognitive work, edges and scenarios is critical for DCT to yield valuable results.

4.6 Summary of DCT limitations

As this document reports the initial phase in the development of DCT, the methodology is still somewhat immature and some limitations remain. For instance, DCT is based on the use of generic support requirements as criteria for evaluating the effectiveness of a JCS. However, although these requirements are based on case studies, they have yet to receive empirical support. In addition, the operationalization of each of these requirements can be rather challenging and their respective independence has not been clearly established yet. Consequently, the support requirements should be investigated further so that they can be defined for use as evaluation criteria. Also, as mentioned before, another limitation lies in the definition of the edge. Indeed, the current version of DCT lacks clear criteria to determine the edge. In order for a DCT evaluation to be as constructive as possible, it is needed to establish a set of criteria that will not only serve to identify what constitutes an edge, but that will be applicable to pinpoint edges in various work domains.

5 Conclusions and recommendations for future work

DSS are essential to support human deciders in military C2 because of limitations brought by limited cognitive capacities, uncertainty and time constraint. In this context, evaluating the extent to which a DSS contributes to increase decision-making effectiveness is crucial, but remains challenging. Indeed, such an evaluation must not only take the DSS into account, but instead consider the team formed of the DSS and the human decider – the JCS. Therefore, DCT was developed within the TIF project Decision Centered Evaluation Capability (DCEC) to fulfill the need for a decision-centered evaluation method that would assess whether or not a DSS truly augments cognition and increases a JCS decision-making effectiveness. DCT constitutes an innovative methodology that significantly departs from traditional assessment approaches by focusing on the DSS-human operator team as the unit of analysis and by evaluating JCS effectiveness based on critical decision points of the work domain rather than on the global outcome of the decision process.

The expected outcome of the DCEC project was for DRDC to have a capability to test prototypes of DSS developed for military applications in order to increase the quality of the products delivered to the Canadian Forces. Used as part of an iterative development process, DCT aims to provide insights into the strengths and weaknesses of the JCS prior to the DSS being implemented into an operational setting, with the objective of improving the JCS decision-making effectiveness. We assert that DCT shows promises in that regards, as suggested by the initial effort in validating the methodology (document in preparation). However, much work remains to be done for the methodology to reach its full potential, both in terms of validation and conceptual development.

5.1 Follow-on work to further validate the DCT methodology

There are several avenues to advance the development of DCT, and the methodology could greatly benefit from additional validation and testing. As mentioned before, some elements of DCT need either further validation (e.g., JCS support requirements) or more precise definitions and criteria (e.g., the notion of edge). Also, solutions should be explored to modify the actual version of DCT in order to make it more affordable time- and resource-wise.

DCT should be tested with various decision-making tasks set in C2 environments, either military or non-military. An example of such a task is the microworld C3Fire (Grandlund 1998), which simulates forest firefighting on a geospatial map. The basic properties of non-military complex and dynamic tasks make them very much in line with tactical decision-making in C2. Therefore, transfer to military settings can be done without loss of precision. Applying DCT with various types of tasks will allow ensuring that DCT is suited for a variety of decision-making contexts.

Furthermore, efforts should be made to expand DCT to teamwork. The present version is limited to the testing of an individual. However, JCS support requirements do include "teamwork with agents" as a critical requirement for testing the support a DSS provides to action synchronization in teamwork. Developing the capacity to test for that requirement would be a major improvement for DSS testing in military settings, given the prominence of teamwork in military operations.

While the original project relied heavily on a given DSS having been designed with the ACWA methodology, it appears that DCT could be applied to any DSS that has been designed based on some level of decision-making task modeling. The more systematic the modeling, the more powerful the application of DCT will be. Therefore, DCT could be adapted so that other types of analysis can be applied to gather knowledge about the work domain. That is, instead of ACWA, analyses such as the Hierarchical Goal Analysis (Hendy et al. 2002, Chow et al. 2006) could be used. One key concern in selecting a method to analyze the work domain is that it must create an output that can be overlaid with the JCS support requirements (observability, directability, etc.) in order to be able to identify the cognitive demands / tasks / decisions that arise in the domain. Ideally, such output should take the form of a goal-means decomposition of the work domain. For instance, Cognitive Task Analysis (CTA) is based on mental models that deciders / operators develop about their work domain. The resulting output is very specific, as it varies highly from one decider to another, and may not be appropriate for DCT. Therefore, the DCT methodology should be tested on operational prototypes in order to evaluate the extent to which it is able to test complex DSS that have not been designed with ACWA. That test of DCT is crucial in order to adapt the methodology to the requirements of military DSS in C2 systems.

5.2 Conclusion

The DCEC project is unique in that it focuses on testing the DSS-Human axis of the triad. DSS are crucial for an efficient control of complex dynamic situations, which is an important aspect of C2 in military operations. The project illustrates how testing can be part of the design cycle of a DSS. Given its focus, it can readily be applied at the prototype level in a rapid prototyping design approach. Hence, DCT can provide a useful contribution to the development of efficient DSS since testing can be incorporated during the design cycle and not left for the end of the design process when a number of design decisions put constraints in modification requirements to a DSS that could result from late testing.

DCT has the potential to bring a significant contribution to the DSS evaluation process, particularly since it focuses on the human operator-DSS team (the JCS) rather than considering only the computerized decision aid. One should remember that DCT is centered on the brittle parts of the JCS that are especially prone to decision errors, and consequently does not take into account more stable parts that also have an impact on the decision-making effectiveness of the JCS. So in order to carry out a thorough evaluation process, DCT should be combined with other JCS evaluation methods. This will allow maximizing the JCS effectiveness in complex military C2 environments, and reduce the risk of costly errors. Future work will allow refining DCT, so that it can achieve its full potential as an evaluation method.

This is an example of a paragraph. The paragraph does not say anything important, but does provide a visual of what a paragraph looks like. The exact form of the paragraph is specified in the document specifications.

References

[1] Chen, J. Q., & Lee, S. M. (2003). An exploratory cognitive DSS for strategic decision making. *Decision Support Systems, 36*, 147-160.

[2] Paradis, S., Breton, R., Bossé, E., Elm, & Potter, S. (2002, October). A pragmatic cognitive system engineering approach to model dynamic human decision making activities in intelligent and automated systems. Paper presented at NATO RTO-HFM Symposium "The Role of Humans in Intelligent and Automated Systems", Warsaw, Poland.

[3] Rousseau, R., Easter, J., Elm, W., and Potter, S. (2005). Decision centered testing: Evaluating joint human-computer cognitive work. In *Proceedings of the Human Factors and Ergonomics Society 49nd Annual Meeting*, Santa Monica, CA: Human Factors and Ergonomics Society.

[4] Hollnagel, E., & Woods, D. D. (2005). *Joint cognitive systems: Foundations of cognitive systems engineering*. Boca Raton, FL: Taylor & Francis.

[5] Wickens, C. D. (1992). Engineering psychology and human performance (2nd ed.). New York: HarperCollins.

[6] Baron, J. (2000). Thinking and deciding (3rd ed.). New York: Cambridge University Press.

[7] Wolf, S. P., Klein, G. A., & Thordsen, M. L. (1991). Decision-centered design requirements. Proceedings of the IEEE 1991 National Aerospace and Electronics Conference (NAECON 1991).

[8] Hutton, R. J. B., Miller, T. E., & Thordsen, M. L. (2003). Decision-centered design: Leveraging cognitive task analysis in design. In E. Hollnagel (Ed.), Handbook of Cognitive Task Design (pp. 383-416). Mahwah, NJ: Erlbaum.

[9] Breton, R., Paradis, S., & Roy, J. (2001). Command Decision Support Interface (CODSI) for human factors display and concept validation. Proceedings of the Fourth International Conference on Information Fusion (Fusion 2001), Montreal, Canada, August 7-10.

[10] Woods, D. D. & Roth, E. M. (1988a). Cognitive engineering: Human problem solving with tools. *Human Factors*, *30*(4), 415-430.

[11] Woods, D. D., & Roth, E. M. (1988b). Cognitive systems engineering. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 3-43). New York: North-Holland.

[12] Woods, D. D., & Hollnagel, E. (2006). *Joint cognitive systems: Patterns in cognitive systems engineering*. Boca Raton, FL: Taylor & Francis.

[13] Elm, W.C., Potter, S.S., Gualtieri, J.W., Easter, J.R., & Roth, E.M. (2003). Applied cognitive work analysis: A pragmatic methodology for designing revolutionary cognitive affordances. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design* (pp. 357-382). Mahwah, NJ: Erlbaum.

[14] Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica, 81*, 211-241.

[15] Clancy, J. M., Eliott, G. C., Ley, T., Omodei, M. M., Wearing, A. J., McLennan, J., & Thorsteinsson, E. B. (2003). Command style and team performance in dynamic decision-making tasks. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 586-619). New York: Oxford University Press.

[16] Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. *Paper presented at the Human Factors Society 32nd Annual Meeting,* Santa Monica, CA.

[17] Endsley, M. R. (2004). Human factors issues in synthetic vision displays: Government, academic, military, and industry perspectives. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting,* Santa Monica, CA.

[18] Klein, G. (1989). Recognition-primed decisions. In W. B. Rouse (Ed.), *Advances in man-machine systems research,* Vol. 5, pp. 47-92. Greenwich, CT: JAI Press.

[19] Klein, G. (1997). The recognition-primed decision (RPD) model: Looking back, looking forward. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 285-292). Mahwah, NJ: Erlbaum.

[20] Billings, C. E., & Woods, D. D. (1994). Concerns about adaptive automation in aviation systems. In M. Mouloua and R. Parasuraman (Eds.), *Automation and Human Performance: Recent Research and Trends* (pp.264-269). Hillsdale NJ: Erlbaum.

[21] Christoffersen, K., & Woods, D.D. (2002). How to make automated systems team players. In *Advances in Human Performance and Cognitive Engineering Research*, *Vol. 2*, (pp. 1–12). Elsevier Sciences.

[22] Elm, W., Potter, S., Tittle, J., Woods, D., Grossman, J., and Patterson, E. (2005). Finding decision support requirements for effective intelligence analysis tools. In *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting*. Santa Monica, CA: HFES.

[23] Zhang, J. (1996). A representational analysis of relational information displays. *International Journal of Human-Computer Studies*, *45*, 59-74.

[24] Norman, D.A. (1988). *The psychology of everyday things*. New York: Basic Books.

[25] Norman, D.A. (1993). Cognition in the head and in the world: Introduction to a debate on situated cognition. *Cognitive Science*, *17*, 124-138.

[26] Zhang, J. & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, *18*, 87-122.

[27] Woods, D. D. (1991a). Representation aiding: A ten year retrospective. In *Proceedings of the 1991 IEEE International Conference on Systems, Man, and Cybernetics*, Charlottesville, VA.

[28] Woods, D. D. (1991b). The cognitive engineering of problem representations. In G.R.S. Weir & J.L. Alty, (Eds), *Human-Computer Interaction and Complex Systems*, Academic Press, London.

[29] Woods, D. D., & Hollnagel, E. (1987). Mapping cognitive demands in complex problem-solving worlds. *International Journal of Man-Machine Studies, 26*, 257-275.

[30] Rasmussen, J. (1986). Information processing and human-machine interaction: An approach to cognitive engineering. Amsterdam: North-Holland.

[31] Lind, M. (1991). Representations and abstractions for interface design using multilevel flow modeling. In G. R. S. Weir & J. L. Alty (Eds.), *Human-Computer Interaction and Complex Systems*. London: Academic Press.

[32] Lind, M. (1993). Multilevel flow modeling. *AAI93 Workshop on Reasoning about Function*, July 11-15, Washington, DC.

[33] Vicente, K. (1999). *Cognitive work analysis*. Mahwah, NJ: Erlbaum.

[34] Potter, S. S., Gualtieri, J. W., & Elm, W. C. (2003). Case studies: Applied cognitive work analysis in the design of innovative decision support. In E. Hollnagel (Ed.), *Handbook of Cognitive Task Design* (pp. 653-678). Mahwah, NJ: Erlbaum.

[35] Granlund, R. (1998). The C^3Fire microworld. In Y. Waern (Ed.), *Co-operative process management* (pp. 91-101). London: Taylor & Francis.

[36] Hendy, K. H., Beevis, D., Lichacz, F., & Edwards, J. L. (2002). Analyzing the cognitive system from a perceptual control theory point of view. In M. D. McNeese & M. A. Vidulich (Eds.), *Cognitive systems engineering in military aviation environments: Avoiding the cogminutia fragmentosa*. (pp. 201-252). Wright-Patterson Air Force Base, OH: Human Systems Information Analysis Center.

[37] Chow, R., Kobiersky, B., & Crebolder, J. (2006). Applied comparison between hierarchical goal analysis and mission, function and task analysis. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. Santa Monica, CA: HFES.

This page intentionally left blank.

List of acronyms

ACWA	Applied Cognitive Work Analysis
C2	Command and Control
CSE	Cognitive System Engineering
CTA	Cognitive Task Analysis
CWRs	Cognitive Work Requirements
DCEC	Decision Centered Evaluation Capability
DCT	Decision Centered Testing
DND	Department of National Defence
DRDC	Defence Research & Development Canada
DRDKIM	Director Research and Development Knowledge and Information Management
DSS	Decision Support System
FAN	Functional Abstraction Network
IRRs	Information Requirements
JCS	Joint Cognitive System
MOEs	Measures of Effectiveness
MOPs	Measures of Performance
OODA	Observe-Orient-Decide-Act
R&D	Research & Development
SA	Situation Awareness
TEWA	Threat Evaluation Weapons Assignment
TIF	Technology Investment Fund

This page intentionally left blank.

DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

1. ORIGINATOR (The name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g. Centre sponsoring a contractor's report, or tasking agency, are entered in section 8.) Defence R&D Canada – Valcartier 2459 Pie-XI Blvd North Quebec (Quebec) G3J 1X5 Canada	2. SECURITY CLASSIFICATION (Overall security classification of the document including special warning terms if applicable.) UNCLASSIFIED

3. TITLE (The complete document title as indicated on the title page. Its classification should be indicated by the appropriate abbreviation (S, C or U) in parentheses after the title.)

The decision centered testing methodology

4. AUTHORS (last name, followed by initials – ranks, titles, etc. not to be used)

Jobidon, M.E.; Breton, R.

5. DATE OF PUBLICATION (Month and year of publication of document.) July 2011	6a. NO. OF PAGES (Total containing information, including Annexes, Appendices, etc.) 42	6b. NO. OF REFS (Total cited in document.) 37

7. DESCRIPTIVE NOTES (The category of the document, e.g. technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)

Technical Memorandum

8. SPONSORING ACTIVITY (The name of the department project office or laboratory sponsoring the research and development – include address.)

Defence R&D Canada – Valcartier
2459 Pie-XI Blvd North
Quebec (Quebec)
G3J 1X5 Canada

9a. PROJECT OR GRANT NO. (If appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant.) 11bu	9b. CONTRACT NO. (If appropriate, the applicable number under which the document was written.)
10a. ORIGINATOR'S DOCUMENT NUMBER (The official document number by which the document is identified by the originating activity. This number must be unique to this document.) DRDC Valcartier TM 2008-155	10b. OTHER DOCUMENT NO(s). (Any other numbers which may be assigned this document either by the originator or by the sponsor.)

11. DOCUMENT AVAILABILITY (Any limitations on further dissemination of the document, other than those imposed by security classification.)

Unlimited

12. DOCUMENT ANNOUNCEMENT (Any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (11). However, where further distribution (beyond the audience specified in (11) is possible, a wider announcement audience may be selected.))

Unlimited

13. ABSTRACT (A brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual.)

(U)The objective of a Decision Support System (DSS) is to assist human operators in the execution of the decision-making process by supporting cognitive limitations and reducing the risk of error brought by time constraints and uncertainty. Human and systems must work as a coordinated team to reach mission goals successfully, and therefore the evaluation of a DSS prior to its deployment is crucial. The Decision Centered Testing (DCT) methodology constitutes an innovative approach to evaluation that aims to test the decision-making effectiveness of the human-DSS team, called the Joint Cognitive System (JCS; Hollnagel & Woods, 2005). In particular, DCT focuses on critical decision points (known to be prone to human decision-making errors) rather than on the general outcomes of the decision process. Used as part of an iterative development process, DCT aims to provide insights into the strengths and weaknesses of the JCS prior to the DSS being implemented into an operational setting. Although much work remains to be done for the methodology to reach its full capability, DCT has the potential to make a significant contribution to the DSS evaluation process.

(U)L’objectif d’un système d’aide à la décision (SAD) est d’assister un opérateur dans l’exécution du processus décisionnel en supportant ses limites cognitives et en réduisant le risque d’erreur provoquée par les contraintes temporelles et la présence d’incertitude dans la situation. L’être humain et les systèmes technologiques doivent travailler en équipe afin d’atteindre avec succès les objectifs de mission. Ainsi, l’évaluation des SAD avant leur déploiement est crucial. La méthode d’évaluation centrée sur la décision (*Decision Centered Testing* ou DCT) présente une approche innovatrice pour l’évaluation de l’efficacité du processus de prise de décision de l’équipe humain-système (ci-après appelée système cognitive combiné; Hollnagel & Woods, 2005). Particulièrement, DCT se centre sur les points critiques de décision (sujet à l’erreur humaine) plutôt que le résultat final du processus décisionnel. Utilisée dans le contexte de développement itératif, la méthode DCT vise à fournir des indicatifs sur les forces et faiblesses du système cognitif combiné avant son implémentation dans un environnement opérationnel. Bien que beaucoup de travail reste encore à faire afin de permettre à la méthodologie d’atteindre son plein potentiel, elle peut déjà contribuer de façon significative au processus d’évaluation de SAD.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (Technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

Measure of performance; decision-making; design evaluation; decision centered design; human computer interaction; cognitive system engineering

Defence R&D Canada

Canada's Leader in Defence
and National Security
Science and Technology

R & D pour la défense Canada

Chef de file au Canada en matière
de science et de technologie pour
la défense et la sécurité nationale

DEFENCE R&D DÉFENSE

www.drdc-rddc.gc.ca